THE COLOR OF IDEAS: RACE, NETWORKS, AND THE DIFFUSION OF KNOWLEDGE

Marlène Koffi & Roland Pongou & Leonard Wantchekon

July 2025

Abstract

This paper examines racial disparities in the diffusion of ideas using a comprehensive dataset comprising over 330,000 economics publications (1950–2021) alongside hand-collected CV data. We document a persistent citation gap: Papers authored by Black, Hispanic, or Asian scholars receive 5.1% to 9.6% fewer citations than those authored by White scholars. This gap, which remains or widens upon considering author seniority and standard quality metrics, is particularly pronounced for Black authors. It also extends to indirect citations and is associated with more clustered citation networks among non-White scholars. Topic choice does not explain this gap; instead, we uncover strong evidence of in-group citation patterns. Thus, we propose a conceptual framework in which clustered networks shaped by information frictions, perceptions, and preferences drive these disparities. Empirical tests looking at variation in online availability, the timing of cross-group collaborations, and author-ordering conventions support this framework over quality-based explanations. Finally, using natural language techniques, we provide suggestive evidence of intellectual complementarity across racial groups, highlighting the potential costs associated with barriers to idea diffusion.

JEL Classification: J15, A14, I23.

Koffi: Department of Economics, University of Toronto (marlene.koffi@utoronto.ca) and NBER. Pongou: Department of Economics, University of Ottawa (rpongou@uottawa.ca). Wantcheckon: Department of Politics and Princeton School of Public Policy and International Affairs, Princeton University (lwantche@princeton.edu) and NBER. We thank Joshua Angrist, Bocar B. Ba,, Cynthia Bansak, Patrick Bayer, Jakina Debnam Guzman, Desire Kedagni, Marie Christelle Mabeu, Matthew Notowidigdo, Zach Ward, Roman Zarate, participants at SOLE 2025, GLab Toronto, Canadian Labour Economics Forum, the ASSA 2024, the 2024 Africa Meeting of the Econometric Society, the Conference on Race, Racism, and Structural Inequality, the 2024 annual meeting of the Canadian Economics Association, the 2024 annual meeting of Société canadienne de science économique, and the 2024 AERIP/ASHE/NEA Summer Conference, and seminar participants at York University, Duke University, Drexel University, Grand-Valley State University, and Yale gender reading group for insightful feedback. We are grateful to Sandy Darity for sharing data on Black economists affiliated with the National Economic Association and Christopher Snyder and Mark McCabe for sharing their data on online availability with us. We also thank Landon Liu, Ghislain Junior Sidie, and Ashna Zaman for their excellent research assistance.

1 Introduction

Racial inequality in education and labor markets remains a significant issue around the world. In the United States, for instance, as of 2022, only 25% of Hispanics and 34% of Blacks in the labor force held a bachelor's or higher degree, well below the 44% of Whites and 68% of Asians with the same. Moreover, Blacks face an unemployment rate nearly twice that of Whites (6.1% compared to 3.2%), while both Hispanics and Blacks earn significantly less than their White and Asian counterparts.[1] These broad social disparities are also evident in academia despite the field's reputation for requiring a high degree of skill and for entry being selective. In 2022, Black and Latino scholars collectively made up just 13% of the academic workforce, while White scholars made up 72%, and this under-representation was even more pronounced in STEM fields (Kozlowski et al. [2022]). This raises the question of whether racial disparities in academia persist beyond the entry stage.

This paper seeks to answer this question by examining racial disparities in the diffusion and dissemination of ideas. Specifically, we document patterns and dynamics of citations as they relate to racial categories. Analyzing citation flows is important for several reasons: citations are used to quantify research quality and influence, as well as the knowledge transmission (Angrist et al. [2020]). In addition, citations predict important career outcomes, including salary (Hamermesh et al. [1982], Hilmer et al. [2015]), employment at prestigious institutions (Ellison [2013]), grants, and prestigious awards (Fiala and Tutoky [2017], Card et al. [2022], Hager et al. [2023]).

This paper focuses on a single discipline—economics—to allow for a more precise and nuanced understanding of the dynamics and mechanisms behind the propagation of scientific ideas. The field of economics is particularly illustrative when it comes to matters of diversity, as PhDs in economics are among the least ethnically and racially diverse of the STEM fields (Bayer and Rouse [2016]). This dynamic has raised concerns over the persistent underrepresentation of minority groups in the field of economics, as highlighted by panel discussions and articles presented at meetings of the American Economic Association (AEA) and the Econometric Society.[2] However, there is very little large-scale evidence of racial disparities and inequitable practices in the profession.[3]

We present a series of analyses aimed at addressing this gap in the literature. First, we undertake an important data collection effort, combining existing structured data with hand-collected data. More specifically, we use publication records and bibliometric information from over 200 economics journals—encompassing approximately 330,000 publications from 1950 to 2021 and representing 221,249 economists—in addition to manually collecting and extracting information from the CVs of nearly 4,500 scholars. By combining algorithmic procedures with manual verification, we identify the race of each author.[4] Based on this, we infer the racial composition of each team of

[1] https://www.bls.gov/opub/reports/race-and-ethnicity/2022/.

[2] Allied Social Science Association (ASSA) Annual Meeting, January 3, 2020, San Diego, "How Can Economics Solve Its Race Problem?"; Econometric Society 2020 World Congress.

[3] Koffi and Wantchekon [2024] present evidence of the under-representation of Africans in the NBER.

[4] We acknowledge that this method does not account for self-identification, which may differ from external perceptions of race based on visual cues and information about origins.

authors. We then construct a detailed set of control variables by leveraging article characteristics and creating an annual publication profile for each author in the database. Exploiting the full bibliographic references of articles, we refine the analysis and categorize papers by journals of publication, affiliations[5], fields, and methodological approaches. We also build citation flows over time and cumulative citations received up to a given year for each publication and author, providing a robust measure of the impact and perceived quality of a researcher's work. Furthermore, by analyzing the acknowledgments sections of articles and conducting an extensive manual data collection of the CVs of authors publishing in the top five journals over time, resulting in nearly 100,000 observations, we extract information on seminar and conference attendance to construct a measure of authors' active outreach and networking efforts. This comprehensive data collection strategy serves to mitigate concerns about omitted variable bias, particularly in the context of racial analysis, where inequalities do not necessarily reflect bias.

Ultimately, we find evidence of a racial citation gap in economics despite the significant progress made in the racial diversity of authorship over time (in 2021, just over half of papers were authored exclusively by White authors, well below the rate of nearly 95% in the 1950s). The estimated average unconditional citation gap is 13.9 log points. After controlling for observable paper characteristics (e.g., number of authors, field, journal, year of publication, author prominence), papers with at least one non-White author (Black, Hispanic, or Asian) receive 5.1 log points fewer citations than papers authored solely by White individuals. We replicate our analysis by focusing only on papers published in top journals to address concerns that the racial citation gap may be driven by lower-quality publications. Restricting the sample to articles published in a set of 35 high-impact journals (as identified by Card et al. [2020]), 16 high-impact journals (as identified by Koffi [2025]), or the traditional top five journals, we continue to observe a persistent and, in some cases, larger citation gap. For instance, the gap for the top five journals is 9.6 log points, compared to 5.1 log points in the original full sample. We further disaggregate the analysis by non-White racial categories, revealing that Asian researchers experience a substantial citation penalty, estimated at 7.1 to 8.5 log points in the top five and 16 high-impact journals, respectively, while the penalty is 7 and 14.7 log points for Hispanic authors. Yet, the largest citation penalty is experienced by Black researchers, for whom the gap is approximately double that of Asian researchers in both the top five and 16 high-impact journals.

Our analysis also indicates significant heterogeneity in the racial citation gap along several dimensions, including affiliations, methodological approaches, and subfields. The gap is nearly twice as large for papers authored by individuals affiliated with top-tier institutions than it is for those authored by individuals affiliated with mid- or low-tier institutions (12 log points versus 6 log points). It is almost negligible among papers employing empirical methodologies, though it is more pronounced in fields like microeconomics and international economics. The racial citation gap has also narrowed

[5]Prior studies have used indicators such as publication in a top-five journal, affiliation with a highly ranked department, and attendance at a top-ranked PhD program as proxy for *ability* (Carrell et al. [2024]). We include some of these indicators in our analyses, alongside additional measures.

over time, standing at just over 3% in the most recent decade (2010-2021).

While analyzing citation counts is important, it fails to capture other dimensions of a paper's scholarly relevance. Thus, as an alternative approach, we assess research dissemination by incorporating measures of centrality, or a paper's significance within a citation network. We use one class of measures to evaluate a node's direct and indirect links (eigenvector, Katz, PageRank, degree centrality) and another to assess its position within the network (betweenness centrality). Across these five measures, we find that papers authored by non-White scholars score lower, suggesting that they are not only less frequently cited overall but also less frequently cited by highly influential works and less likely to serve as bridges between other papers. We also examine one other measure of centrality, the clustering coefficient, to capture the likelihood of a paper's neighbors citing one another and forming a tightly knit group. Our analysis shows that papers authored by non-White scholars exhibit higher clustering coefficients, indicating that these papers are more often located within cohesive subgroups of the citation network. This dynamic suggests a more limited dissemination of their ideas.

We conduct several robustness checks, beginning with an assessment of potential measurement error in race classification. Using our hand-collected dataset of authors published in the top five journals, we validate the algorithmic procedure, finding statistically similar citation gaps between manually coded and algorithmically inferred race.

We also address the concern that the racial citation gap may stem from differences in field or topic specialization. To account for this, we include detailed field-fixed effects using two- and three-digit JEL codes and apply natural language processing techniques to paper abstracts in order to capture finer topical variations via vector representations. These adjustments facilitate the granular control of the research focus, yet do not give way to a decrease in the estimated citation gap.

Furthermore, the citation penalty remains robust across alternative definitions of team racial composition, controls for text readability, different citation measures, the exclusion of self-citations, and the inclusion of an extensive set of additional covariates.

Although the citation penalty for non-White-authored papers persists across multiple robustness checks, variation in the magnitude of this penalty across different specifications points to the need for a deeper investigation of the underlying mechanisms.

We examine the role of differential author ability in several ways. First, we assess whether publication history reduces the citation gap, as belief-based models would predict that more signals of quality lead to updated priors. We find that publication history only narrows the gap for non-White authors at the extreme end of the publication distribution—those with at least 20 total publications. Second, we use authors' citation histories at the time of publication as proxies for quality. While prior citations are positively associated with subsequent citations, controlling for them does not eliminate the racial citation gap. The gap also persists in a panel specification that examines annual citations for a given paper while controlling for cumulative citations at both the paper and author levels, despite the endogeneity of these controls and their likely downward bias on the estimated effect. Finally, we discuss whether affirmative action—or perceptions thereof—might explain the gap. Yet

the persistence of the citation penalty since the 1950s and its presence among Asian authors make this explanation unlikely. Taken together, although author ability and paper quality are inherently unobservable, our results are not consistent with these factors being the primary drivers of the citation gap observed for papers with at least one Black, Hispanic, or Asian author.

We find that in-group preference plays a significant role in the racial citation gap. Papers with at least one Black, Hispanic, or Asian author receive 23% more citations from non-White scholars but 30% fewer from White scholars. This homophily effect persists across fields and over time and is not explained by geographic proximity, holding for both within- and across-continent citations.

Overall, we document a persistent racial citation gap that is evident in both direct and indirect citations, particularly among elite journals and institutions. This gap is associated with greater network clustering among non-White scholars, suggesting insular citation patterns. It is not explained by differences in research topics or observable measures of paper quality. Rather, we find strong evidence of in-group citation preferences, indicating that racial similarity between the citing and cited authors plays a central role.

These patterns are consistent with the hypothesis that clustered networks—shaped by information frictions, perceptions, and preferences—underlie the citation gap. To formalize this idea, we develop a conceptual framework that models citation behavior in a multiethnic scholarly environment with randomly assigned referees. In this setting, authors are more familiar with co-ethnic work, reducing the cost of citation and enhancing the perceived value of such papers. The model predicts homophilic citation behavior across all groups when citation costs alone drive decisions. However, when referee preferences and selection mechanisms are introduced, homophily becomes less pronounced for minority scholars, and their papers receive fewer citations than comparable work by majority authors. The model also highlights heterogeneity in the gap, showing that it widens upon considering higher-quality publications and more prolific authors. It further offers a potential explanation for the observed differences in racial and gender-based citation patterns that aligns closely with our empirical results.

Building on this framework, we turn to an empirical evaluation of the clustered network hypothesis.

We begin by examining the role of information frictions in explaining the racial citation gap. Using an event-study design, we assess whether increased online availability reduces the gap. While online access eventually raises citations for papers by non-White authors, the gap widens in the short run, suggesting that information frictions alone do not account for the disparity. This interpretation is reinforced by our analysis of author visibility. Using our hand-collected CV database, we control for multiple visibility-related factors, including seminar participation, invitations from top institutions, editorial roles, and affiliations with elite organizations. These factors collectively account for up to 1.5 log points of the gap, though their impact is not statistically significant, with most of the variation linked to unequal access to prestigious associations. Participation in seminars and conferences has a limited independent effect, largely due to its correlation with author prominence, which the model

already captures.

Next, we explore the role of perceptions (potentially intertwined with information frictions) by drawing on contact theory to examine whether collaboration with non-coethnic authors influences citation behavior. Using an event-study framework, we find that authors cite more out-group scholars after their first collaboration with a non-coethnic coauthor. This effect is robust to matching and persists over time, though it attenuates, suggesting that sustained change may require repeated exposure (as in Dahl et al. [2021]).

Finally, we examine the role of preferences and biases by testing whether author visibility interacts with racial identity. Leveraging economics' alphabetical authorship convention, we find that papers with non-White first authors receive fewer citations from all-White teams and more citations from non-White teams. This pattern holds both at the paper level and in author-level specifications, where individuals appear in varied positions on mixed-race teams. These results suggest that visibility amplifies racial dynamics in citation, consistent with the influence of preferences alongside information-based mechanisms.

Together, these findings suggest that the racial citation gap cannot be fully explained by differences in paper quality, but instead reflects information frictions and deeper behavioral dynamics related to network composition, perception, and preferences, ultimately giving rise to clustered citation patterns.

Finally, we offer a discussion of the potential costs to scientific progress posed by racial barriers in idea diffusion. Using natural language processing, we find indicative evidence of complementarity across racial groups, suggesting that greater inclusion could bolster innovation and, in turn, amplify the impact of economic research.

1.1 Related literature

Our paper contributes to the literature on identity-based inequalities in academic research. It is among a handful of studies that systematically examine how the dissemination of academic ideas and contributions varies based on authors' racial characteristics and document potential explanations on both the side of the citing authors and that of the cited authors. Within the field of economics, while the literature on gender inequalities in academic research has expanded significantly in recent years, there is much less information on racial inequalities.

Two papers have examined the racial citation gap, but they are focused primarily on the economics of race and crime. Mason et al. [2022, 2023] find that articles published in top economics journals on topics related to race and crime by Black scholars receive fewer citations than similar articles authored by non-Black scholars. They also reveal that articles authored by Black scholars are more likely to document the presence of racial discrimination. However, their research is undervalued by mainstream economics, despite evidence that Black economists have persistently engaged on economic policy issues related to race and other relevant topics such as poverty, education, and public finance (Francis et al. [2022]). Our scope, analysis, and findings significantly differ from those of these studies.

We include articles on all the topics covered in economics and focus on all non-White groups.

Other works on racial and ethnic inequalities within the economics profession examine the underrepresentation of minority scholars as well as scholars from developing countries. Bayer and Rouse [2016] argue that women and members of underrepresented minority groups remain significantly underrepresented in the field of economics. Bayer et al. [2020] identify factors such as bias, a hostile academic climate, limited access to information, and inadequate mentoring as key contributors to this persistent underrepresentation. Koffi et al. [2024] document racial inequalities in economic publishing, uncovering notable differences in authorship between White and non-White authors, with Black, Hispanic, and Asian economists being significantly underrepresented relative to their proportion of the overall population. Foster et al. [2022] and Ginther et al. [2025] further analyze career trajectories in economics by race and gender. Amarante et al. [2022] highlight the dramatic underrepresentation of researchers from developing countries in the field of development economics. Similarly, Greenspon and Rodrik [2021] show not only that authors from developing countries are rarely represented in top economics journals but also that there has been little improvement in this dynamic over time. In this paper, we document a positive trend in authorship from racially diverse profiles but use citations to show that their work remains insufficiently recognized.

Our paper is also linked to prior research on citation patterns in the field of economics. Hamermesh [2018] reports a significant increase in citations of empirical work in top journals since 1990. Lutmar and Reingewertz [2021] use citations to identify in-group bias in economics, with authors who share journals' institutional affiliation potentially receiving preferential treatment. Other studies have examined the insularity of economics and its influence on other disciplines through cited references (Pieters and Baumgartner [2002], Fourcade et al. [2015], Angrist et al. [2020]. We extend this literature by documenting citation insularity and the formation of racial clusters within citation networks.

Our research complements the growing literature on gender-based inequalities in the economics profession. Using bibliometric data from papers published in top economics journals, Koffi [2021a] finds that female-authored papers receive fewer citations from men and authors in top-tier journals than similarly situated papers authored by men. Koffi [2025] uses machine learning to establish similarities between papers and construct relevant counterfactual citations, finding that papers receive fewer citations if they are authored solely by women than if they are authored solely by men, although most of the uncovered variation in citation was attributable to authors' publication history. Using data on interactions between presenters and their audiences in seminars at leading economics departments and conferences, Dupas et al. [2021] find that women presenters are treated differently than similarly situated men; not only do women receive more questions than men during seminars, but the questions asked of women tend to be more hostile. Other studies showing unfavorable outcomes for women economists point to the higher standards imposed on women (Hengel [2022]), prestige rankings (Zacchia [2021]), career trajectory, tenure, and promotion (Ginther and Kahn [2004], Sarsons [2017], Chen et al. [2022]), and their treatment on online forums (Wu [2020]). Conversely, recent

studies indicate that women have been more likely to be elected as members of the Econometric Society in recent years (Card et al. [2022]) and face review times on par with those faced by men (Card et al. [2020]). In contrast to this wide array of studies, research on racial inequalities within the field of economics remains limited. Notably, while teams including women have been shown to receive a citation advantage on account of higher publication standards (Card et al. [2020]), there is no research on whether teams' inclusion of non-White authors influences citation rates or how this influence, if observed, should be interpreted. Our paper aims to fill this gap in the literature.[6]

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 details the empirical analysis and presents the results. Section 4 discusses the potential mechanisms underlying the findings as well as their implications. Section 5 concludes the paper.

2 Data description

2.1 Data collection

We collected publication records from more than 200 journals using the Web of Science (WoS) database, EconLit, and the Microsoft Academic Graph (MAG) database, all of which are widely recognized as some of the largest repositories of academic research in the field of economics. WoS serves as the core database, EconLit provides information on Journal of Economic Literature (JEL) codes, and MAG assists with name disambiguation.

Collected information was subsequently organized into a new database. The sample includes the (widely recognized) top five general-interest journals in the field—*American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economics Studies*—alongside other high-impact journals. The complete list of journals is provided in Appendix Table C.20.

The resulting database comprises full-length articles, excluding various other types of content, such as proceedings papers, comments, articles shorter than three pages, book reviews, bibliographical items, editorial materials, letters, and corrections.

In total, the database contains 329,338 unique papers. Each paper is accompanied by a range of associated pieces of information, including the title, journal of publication, publication date, authors' names, authors' affiliated institutions, number of citations, DOI, abstract, keywords, JEL codes, and references. The citation data was downloaded between March and May 2021.

[6]Racial disparities in academia have previously been documented in fields other than economics (Kozlowski et al. [2022]). In psychology, fewer than 1% of lead authors and research samples are from Africa [Thalmayer et al., 2021], raising concerns about the validity of scientific evidence in developing contexts [Mughogho et al., 2023]. Similar imbalances have been observed in neuroscience reference lists [Bertolero et al., 2020], postdoctoral evaluations [Eaton et al., 2020], and NIH funding decisions [Ginther et al., 2011, Taffe and Gilpin, 2021]. Notably, citation biases are a global concern, with Chinese authors receiving fewer citations from U.S. researchers [Qiu et al., 2022]. This paper contributes to the literature by identifying the sources of these disparities and detailing the underlying mechanisms.

2.2 Race coding

2.2.1 Manual and algorithmic predictions

The determination of an author's race in this study involved a combination of automated procedures using *ethnicolr* package (Sood and Laohaprapanon [2018]) as well as manual data-collection efforts. *Ethnicolr* leverages U.S. Census and Florida voter registration data to classify race based on first and last names.

The race-determination protocol was as follows. First, the authors' names were retrieved from the bibliographic data. To minimize the risk of errors, we initially made a prediction based on the last name alone before making a prediction using both the last name and the first name (when available). The following conditions were imposed: (1) there must be agreement between the two predictions; and (2) the probability of assigning a given racial label to an individual must be higher than 0.7. We also presented alternative specifications based on other cutoffs and conditions.

As shown in Hofstra et al. [2020], the algorithm is quite accurate when it comes to identifying Whites and Asians (and, to some extent, Hispanics), but it is less accurate for Blacks. In fact, the Black category has the highest error rate due to name assimilation stemming from the heritage of slavery and colonisation. To address this limitation, we supplemented the data with a list of Black economists from the National Economic Association. We manually verified and traced the networks (e.g., co-authors, PhD universities) of these Black economists to identify additional Black economists.

Overall, the database features 221,249 unique authors. Race assignments were achieved for approximately 80.5% of the authors in the dataset, among which 70.2% were identified as Whites, 23.8% were identified as Asians, 5.4% were identified as Hispanics, and the remaining 0.6% were identified as Blacks.

Furthermore, due to teams' skewed racial compositions (more White authors and fewer non-White authors), we adopted a conservative measure of team racial composition. Specifically, the racial composition of each co-author team was categorized into three distinct groups: "All-White," indicating that all co-authors are White; "At least one Black/Hispanic/Asian," indicating that the co-authorship includes at least one non-White author; and "Other races," for cases where authors' race could not be classified into the conventional categories. This latter category typically includes individuals of Arabic or Persian descent or those with mixed perceived ethnicities whose names do not clearly align with a single category. Notably, we also present results using alternative measures of team racial composition.

2.2.2 Measurement errors

Race coding based on names is likely to capture perceived race. Although applying cutoffs may improve the alignment between predicted racial identity and actual race, it cannot fully eliminate uncertainty or the risk of measurement errors.

To address this, we manually collected data from the CVs of a subsample of the authors to verify

the algorithmic predictions (see section 2.3 below for details). We audited the algorithmic race-classification procedure in Appendix Table C.1 and found an overall accuracy rate of 90%. More importantly, the econometric results based on the manually verified sample are largely consistent with those obtained using the algorithmically inferred classifications (these results are presented in the robustness section 3.5.1 of the paper).[7] Therefore, we employ simplified terminology throughout the paper and refer to an author's predicted race as their race.

2.3 Hand-collected CV data

Acknowledging that authors' backgrounds and visibility may impact their citations, we conducted a thorough manual collection of the CVs of authors publishing in top-tier journals. We focused on authors with at least one top-five publication since 1990, a category featuring 5,150 unique authors. We extracted demographic information from these hand-collected CVs, including country of origin, date of birth, marital status, and number of children (at the time of data collection), whenever available. We also gathered data on PhD, MA, and bachelor's degree-granting institutions and the year in which each degree was granted. Whenever possible, the names of authors' PhD advisors were also gathered. Moreover, we collected annual data on the different institutional affiliations of each author since the year in which they completed their PhD. More precisely, data was gathered on their main affiliations, visiting appointments, and memberships in esteemed research organizations, such as the NBER, JPAL, CEPR, BREAD, IZA, and CIFAR. Moreover, we collected annual information on each author's academic rank, editorial activities, and rank at every journal with which they were associated. Furthermore, we compiled details on each author's attendance at seminars and workshops as well as annual data on the fellowships and awards they had received since obtaining their PhD. Finally, we manually verified each author's gender and race. Following each of these authors every year post-PhD-completion resulted in the accumulation of nearly 100,000 observations.

2.4 Other variables

Beyond race, several key constructed variables in this study merit discussion.

To obtain granular information on citation patterns, we collected each published paper's bibliographical references, enabling us to track the citations over time and match the characteristics of the citing paper to those of the cited paper. JEL code information is available in Econlit, although not for all articles. For some articles, text information matched to JEL codes was available. We used the consortium mapping of JEL codes to JEL descriptions from the AEA website to match JEL descriptions, as used in the EconLit database, to the JEL codes. However, the field classification was still missing for about a third of the data. To address this, we used Koffi [2025]'s field classification algorithm to predict the primary field of the remaining articles.[8]

[7] Even when predicted probability matches perceived race, there is still a distinction between perceived race and self-identification that is not captured in our analysis.

[8] The field-classification algorithm is based on text classifications using information from papers' titles, abstracts,

The article database was converted into an author-level database to facilitate the tracking of each researcher's publications. To match the three databases at the author level—conditional on an article-level match—we used authors' last names and positions, employing a disambiguation procedure similar to that in Card et al. [2022].A yearly publication history for each author was then compiled, enabling the creation of variables representing their publication trajectories, including lifetime and date-specific publications. Article-level metrics such as the most prolific author, average publications per author, and total publications across journal tiers were calculated. The first publication date for each author was also tracked to identify the most senior author on a paper.

Author institutions were identified following a procedure similar to Hengel [2022] and Koffi [2025]. Indeed, given that each paper can potentially have multiple authors from different affiliations, we ranked the affiliations based on their total number of publications in high-impact journals.

2.5 Summary statistics

Table I presents the summary statistics for the dataset, which includes a total of 329,231 papers. Of the total, 178,411 papers (54.2%) were authored exclusively by White teams. In contrast, 83,242 papers (25.3%) included at least one Black, Hispanic, or Asian author, with 2,754 papers involving at least one Black scholar, 20,983 papers including at least one Hispanic scholar, and 59,505 papers featuring at least one Asian scholar. Additionally, 67,578 papers (20.5%) were authored by individuals from other racial groups. As shown in Figure I, there has been a steady rise in the proportion of published papers featuring non-White authors over time. This growth becomes particularly apparent from the 1990s onwards. In the set of 35 high-impact journals, their share even reached 50% as of 2021. This growth is evident across all racial groups, with the most pronounced increase evident among Asians (Panel b of Figure I).

Papers authored by all-White teams tend to have fewer authors, averaging 1.7 authors per paper. In contrast, papers with at least one Black, Hispanic, or Asian author average 2.43 authors per paper, while papers from other racial groups average 2.19 authors per paper.

Regarding publication records, all-White author teams had the highest average number of publications across various journal tiers as of 2021. For example, all-White teams averaged 19.97 publications in the top 200 journals, compared to 19.33 for papers with at least one non-White author and 15.62 for papers from other racial groups. A similar statistically significant trend is evident among top-tier journals (Top 35, Top 16, and Top 5), for which all-White teams consistently exhibited higher averages than teams featuring non-White authors.

The dataset also provides valuable insight into the distribution of papers across different academic fields. Papers authored by all-White teams were most concentrated in microeconomics (19.14%) and other empirical fields (26.21%), while teams with at least one non-White author were more concentrated in finance (19.27%) and econometrics (11.84%). Non-White authors tend to produce

and keywords. We employed a similar classifier to identify each paper's methodological style. See Koffi [2025] for further details on these classification algorithms.

higher proportions of empirical papers and lower proportions of theoretical papers than their White counterparts.

Finally, the dataset covers authors' institutional affiliations, ranking them into low-, mid-, and top-tier institutions. All-White teams had the highest proportion of papers affiliated with top-tier institutions (14.3%), surpassing the 13.37% among teams with at least one non-White author and the 11.01% among papers from other racial groups.

3 Empirical strategy and results

3.1 Empirical framework

We test for how citation patterns differ depending on the racial composition of a paper's authors. Assume that paper p was published in year t and cited by articles published in a given subset of journals. Then, our outcome variable, the total number of citations (using the inverse hyperbolic sine function), denoted $C_{p,t}$, received by paper p is given by the following linear model:

$$C_{p,t} = \beta_0 + R_p\beta_R + X_p\beta_X + \gamma_{jt} + \lambda_f + \epsilon_{p,t} \tag{1}$$

In this study, the primary variable of interest, R_p, defines the racial composition of the authors of article p. In the baseline specification, this indicator is defined as follows: it is equal to 0 if the article has only White co-authors (all-White); 1 if the article has at least one co-author who is Black, Hispanic, or Asian; and 2 if all the authors in the article cannot be considered White and the article does not have any Black, Hispanic, or Asian authors (see Section 2.2.1). In Equation 1, the parameter of interest, β_R, captures in particular the difference in the number of citations received by article p when it includes at least one non-White co-author, compared to when it is authored exclusively by an all-White team.[9]

To account for potential heterogeneity between articles, we include a range of control variables represented by the vector X_p. These controls encompass characteristics such as the number of authors, the affiliation of the article, and the academic achievements of the authors. The number of authors can influence both the racial composition of the article and the likelihood of citation, as a larger network may increase citation probability. Including authors' affiliations helps control for differences in the visibility or the perceived quality of the authors. Similarly, considering the academic reputation of the authors allows us to condition the prestige and prominence of the authors. We also include journal times year fixed effects, denoted by γ_{jt}, to control for factors related to the journals and their publication years. Additionally, we incorporate dummy subdomains using the primary field of the paper, represented by λ_f, to account for variations across economic subfields.

[9]The parameter β_R is a vector of two coefficients that capture the effects of papers authored (i) by at least one non-White scholar and (ii) by scholars who are not considered White and do not include any Black, Hispanic, or Asian individual, respectively. Our primary interest, however, is the effect of papers authored by at least one non-White scholar.

3.2 A racial gap in citation

The estimation results of Equation 1 above are presented in Table II. We use various specifications and controls across the columns. Column (1) includes no controls. Column (2) incorporates controls for the number of authors. Column (3) additionally controls for authors' academic performance. Column (4) introduces year-fixed effects, while Column (5) includes affiliation rank fixed effects. Column (6) adds field-fixed effects, and Column (7) controls for all the aforementioned variables, including journal-year fixed effects.

The table shows a significant citation gap between papers exclusively authored by White scholars and those with non-White authors, even after controlling for team size, prior publications (as a signal of productivity), affiliations, primary field, and journal-year of publications. Articles with at least one Black, Hispanic, or Asian author consistently receive 5.1 to 25.6 log points fewer citations than those authored solely by White scholars, with the unconditional gap being around 13.9 log points.

Three variables significantly influence the estimates. First, increasing team size tends to widen the citation gap, increasing the difference by 12 log points. Non-White scholars typically have larger team sizes, which should attract more citations. However, these diverse teams do not receive as many citations as similarly sized all-White teams. In contrast, publication history reduces the gap, explaining 10 log points, while publication years account for at least 9 log points.[10] In the sample, non-White scholars tend to have fewer publications up to 2021, and their publication years are, on average, ten years more recent. Citations typically accumulate over time, contributing to the observed disparity in citations.

In our preferred specification in Column (7), the estimated citation penalty for articles with at least one Black, Hispanic, or Asian author remains, on average, 5.1 log points or approximately 5%. Similarly, the citation penalty for the other races category averages 4.4 log points.

Furthermore, due to the substantial heterogeneity among the journals included in our baseline estimation, it is unclear if the citation gap persists among a subset of high-quality journals. For instance, the citation gap may be driven by lower-quality journals, while higher-quality journals could be immune to this gap. However, Table III shows that the citation penalty for non-White scholars is not confined to lower-ranked journals. We estimate a gap of approximately 9% (9.6 log points) for the top five journals in economics, 8% (8.5 log points) for the set of top 16 journals used in Koffi [2025], and 8% (8.6 log points) for the set of 35 high-impact journals laid out by per Card et al. [2020].

Lastly, to gain deeper insight into the underlying dynamics, we separately categorize the papers based on whether they feature at least one Black, Hispanic, or Asian author, with these categories being mutually exclusive.[11] Figure II shows that the citation penalty persists across all racial cat-

[10] We accounted for the number of publications by the most prolific author at the time of the paper's publication and, as of 2021, the final year of the database's data collection. This dual approach acknowledges that an author's current prominence and their prominence at the time of publication can impact citation frequency. The results show that prominence up to 2021 positively affects citation rates, while prominence at the time of publication negatively impacts them when controlling for prominence up to 2021.

[11] Since Black scholars are the least represented, followed by Hispanic and Asian scholars, we prioritize classification

egories and journal ranks. However, the penalty is more pronounced for Black scholars in the top 35, top 16, and top 5 journals, with coefficients for Black-authored papers of 10%, 17%, and 18%, respectively.

3.3 Heterogenous effects

First, we examine how the racial citation gap varies by authors' affiliation. We categorize each paper based on whether its most prolific author belongs to a top 15 institution ("top-tier"), a top 15–100 institution ("mid-tier"), or an institution ranked below the top 100 ("low-tier"). Figure III shows that the citation penalty for non-White scholars is twice as large in top-tier affiliations compared to mid and lower-tier affiliations. In other words, non-White scholars at top-tier institutions face a stronger citation penalty than their White counterparts, and this penalty is more pronounced than that experienced by non-White scholars at mid- or low-tier universities compared to White scholars at similarly ranked institutions.

We then estimate the model across different subfields and methodological approaches within the field of economics, as shown in Figure IV. The citation gap for papers with at least one non-White author is evident across almost every subfield. The most severe penalties occur in microeconomics and international economics (slightly exceeding -10 log points), followed by econometrics, labor and education, macroeconomics, and finance. Moreover, the presence of at least one non-White author is consistently associated with a citation gap, particularly among predominantly theoretical papers (-0.079; standard error = 0.007) and predominantly structural papers (-0.061; standard error = 0.024). Notably, the citation gap for papers authored by non-White individuals is nearly five times greater among theoretical papers compared to empirical papers (-0.016; standard error = 0.009).

In Appendix Table C.2, we estimate a citation penalty for non-White authors of 5 log points with a single author and 7.2 log points with two authors. For papers with more than three authors, we estimate a 3.2 log-point penalty.

Finally, an important aspect of analyzing heterogeneity is determining whether the observed citation penalty persists in contemporary settings. More specifically, it is important to examine whether the citation gap has narrowed over time alongside the rising frequency of teams featuring at least one non-White author. Figure V shows that, while the negative impact of having non-White authors on article citations persists, the penalty itself has decreased since the 1990s. For articles published before 1989, the citation penalty was between 8 log points and 13 log points; after 1989, the penalty is closer to 5 log points.

accordingly. Teams with a Black author are categorized first, followed by those with a Hispanic author, and then those with an Asian author. For example, a team with both a Black scholar and a Hispanic scholar is classified as having a Black scholar.

3.4 Characterization of the citation network

In the previous section, we measured a paper's ability to disseminate a research idea using the number of citations it has received. However, scholars view some citations as more important than others, and the influence of a paper is not only measured on the basis of the number of citations it has received but also by the number of "influential" papers it has inspired. We analyze the role of authors' races in determining the influence of their scholarly works.

To measure paper importance, we assume that published papers form a directed network in which a paper is linked to another if it cites the latter. In this context, papers are "nodes," and citation links are "edges." In graph theory, node importance is measured using various centrality indices categorized into two broad classes. We use the first class of measures to evaluate a node's direct and indirect links (eigenvector, Katz, PageRank, degree centrality) and the second class to assess its position within the network (betweenness centrality). We consider all six measures to determine a paper's importance.

Degree centrality counts the number of direct citations a paper receives. Betweenness centrality measures a paper's importance based on how often it acts as a bridge along the shortest path between two other papers. A paper with high betweenness centrality often mediates the influence of an earlier article on a later one, providing explanations, extensions, or applications of previous ideas or serving as a literature review summarizing original findings. Eigenvector centrality, on the other hand, assigns scores to papers based on the idea that citations from highly cited papers are more valuable than those from less cited papers. The clustering coefficient quantifies how close the neighbors of a paper are to forming a complete graph, indicating how tightly knit they are. The respective formulas of these indices are presented in Appendix A.2. We estimate the baseline equation using degree centrality, betweenness centrality, eigenvector centrality, Katz centrality, and PageRank centrality as outcome variables.

The results, presented in Table IV (Columns 1 to 5), show that papers with at least one non-White author score significantly lower on all measures compared to papers by all-White teams. Therefore, the citation penalty for papers authored by non-White scholars extends beyond direct citations to encompass various measures of a paper's importance within a citation network. This includes indirect citations and the paper's position within a network of interconnected papers. These findings indicate that the citation disparity is not merely a matter of fewer direct citations, but also reflects a broader trend in how these papers are integrated and disseminated within the scholarly community.

Column (6) of Table IV presents the clustering coefficient. We observe that papers by non-White authors tend to have higher clustering coefficients, meaning that papers citing non-White authors are likely to cite one another, giving way to the formation of a tightly interconnected group. This suggests that these papers are more likely to be situated within a cohesive subgroup of the citation network, leading to the limited dissemination of their ideas.

3.5 Robustness

In this section, we show that our results are robust to a wide range of alternative specifications.

3.5.1 Measurement errors on race

Algorithmic methods for determining an individual's race rely on name-based classifiers, which may not accurately capture racial identity due to the diversity of names within racial groups, the effects of cultural mixing, and the complexities of self-identification. Therefore, they are inherently prone to measurement error and may introduce bias and noise into the analysis. This measurement error is particularly problematic in studies of racial disparities, as misclassification can obscure the true extent of racial gaps or, conversely, exaggerate differences.

To validate the results of the algorithmic classification, we conducted a manual verification of the authors in our dataset. Given the large size of the database, which includes over 220,000 unique authors, we limited the manual verification process to authors who published after 1990 and have at least one publication in a top 5 journal. We manually verified the races of the 5,149 authors who meet these criteria by reviewing their personal and institutional webpages and searching for photos and information about their countries of origin.[12]

The findings from the manual verification, presented in Appendix Table C.3, confirm the presence of racial citation bias, demonstrating that non-White authors are cited less frequently than White authors. Algorithmic predictions estimate a bias approximately 2 log points higher than the results of the manual verification, but this difference is not statistically significant. In our most conservative specification, the bias is estimated at 7.5 log points with a standard deviation of 2.9 log points.

3.5.2 Finer field controls

We conduct robustness checks by adding more granular field controls, as presented in Table V. Column (1) shows the baseline specification with all controls, restricting the sample to papers with available JEL codes. Similar to the baseline, we find a racial citation gap of approximately -5 log points for papers with at least one Black, Hispanic, or Asian author compared to papers with all-White authors. Additionally, there is a penalty of 3.5 log points for papers whose authors are neither classified as Black, Hispanic, Asian, nor White. Columns (2) to (4) alternatively control for the 1-digit, 2-digit, and 3-digit JEL codes, yielding citation gaps of 7.0, 7.0, and 6.5 log points, respectively.

To enhance the depth of our analysis, we leverage natural language processing methods to examine the abstracts of the papers in Column (5). By transforming each abstract into vectorized representations, we capture their semantic structure and content, enabling a more nuanced exploration of the fields and patterns within the dataset. More specifically, we use Sentence-BERT (Reimers [2019]) to generate an embedding representation (a 768-dimensional vector) of the papers.[13] As shown in

[12] Although this method reduces the potential effect of measurement error, it is still distinct from self-identification.

[13] BERT (Bidirectional Encoder Representations from Transformers) is a large language model introduced by Devlin [2018]. BERT revolutionized natural language processing through the introduction of a bidirectional transformer

Koffi [2025], embedding models offer a more nuanced and detailed representation of a text's topical content compared to the use of JEL codes.

Then, the citation gap is 6.9 log points in Column (5). The gap persists when controlling for methodological approach in Column (6) and when combining it with JEL codes in Column (7).[14] In Column (8), which controls for three-digit JEL codes, methodological approach, and embedding representation, the gap persists at 6.4 log points.

Our findings confirm that the negative citation premium for papers authored by non-White scholars persists, even with more granular field controls. It may initially have seemed plausible that researchers tend to specialize in certain fields influenced by characteristics such as race or ethnicity (Table I). Therefore, if non-White researchers are overrepresented in fields with lower citation rates or focus on less prominent topics, they might be expected to receive fewer citations. However, our analysis does not support this hypothesis. In fact, as shown in Table V, accounting for an extremely detailed research and topic categorization tends in the full sample, if anything, to increase the average gap by 20%-40%.[15]

3.5.3 Other robustness checks: alternative definitions of race, readability, author prominence, and self-citation

We conduct several robustness checks to gauge the sensitivity of our results. Appendix Tables C.4 and C.5 present estimates using alternative definitions of race. The citation penalty for papers with a non-White author ranges from 3.8 to 6.3 log points, indicating that our results are not driven by the definition of race.

Appendix Table C.6 addresses concerns that papers by non-White authors may be harder to read, especially for non-native English speakers. The results show that controlling for textual readability does not meaningfully impact the estimated citation gap.

Given recent concerns over the inverse hyperbolic sine transformation [Bellemare and Wichman, 2020, Chen and Roth, 2024], Appendix Table C.7 shows that our results are robust against alternative functional forms of the citation variable.

To account for potential differences in author prominence, Appendix Table C.8 incorporates alternative measures of authors' publication records.

Finally, Appendix Table C.9 demonstrates that the exclusion of self-citations does not meaningfully affect the qualitative pattern of the results.

model, which allows for a deeper contextual understanding of language. Sentence-BERT is an extension of BERT geared toward enhancing sentence-level understanding. See Appendix A.3.2.

[14] Our analysis reveals that empirical papers are cited more frequently than theoretical ones, in line with the findings of Angrist et al. [2020].

[15] In Section 4.2, a heterogeneity analysis examining the race of authors in citing papers shows that adding detailed field controls produces effects of varying magnitude across racial groups.

3.5.4 Selection on unobservables

A potential concern is that unobserved factors correlated with race may explain the estimated citation penalty for non-White author teams. As shown in Table II, the change in the coefficient after adding controls suggests that such unobservables may contribute to the observed gap. Therefore, we estimate the Oster [2019] bound for our main specifications in Table II. The identified set uses parameters $\delta = 1$ and $R^2_{\max} = \min(1, 2R^2_c - R^2_u)$, where R^2_c represents the R^2 resulting from regressing the dependent variable on observable factors, and R^2_u signifies the R^2 from regressing the dependent variable exclusively on the race variable. The adjusted coefficient is given by $\beta = \beta_{\text{controlled}} - \delta \cdot [\beta_{\text{uncontrolled}} - \beta_{\text{controlled}}] \cdot \frac{R^2_{\max} - R^2_c}{R^2_c - R^2_u}$.

Appendix Table C.10, Columns (1) and (2), shows that the identified set for the coefficient on papers with non-White authors is $[-0.051, 0.026]$ under the extreme assumption that $R^2_{\max} = 1$. When $R^2_{\max} < 0.78$, the upper bound falls below -0.01 log points. Similarly, assuming $R^2_{\max} = 1$, the identified set for the coefficient on papers with at least one Black, Hispanic, or Asian author likely includes zero for the top 35 and top 16 journal subsamples—except for the Top 5, where the set remains bounded away from zero at $[-0.12, -0.096]$.

Overall, the extent to which unobservables can fully explain the citation gap depends on the assumed value of $R^2_{\max}$. Since even bias is unobservable, caution is warranted in interpreting these bounds. Given this ambiguity, the next section provides a detailed exploration of potential mechanisms, including those related to unobserved factors.

4 Possible sources of the racial citation gap

4.1 Differential quality and impact

4.1.1 Noisy ability signal and information updating

If agents have lower priors when it comes to non-White scholars' abilities and anticipate that papers by non-White scholars are, on average, of a lower quality, then traditional belief-based bias models based on mean quality will imply that they will update these priors upon receiving signals of non-White scholars' abilities.

We investigate this dynamic by analyzing how additional publications affect the citation penalty experienced by non-White researchers. Figure VI illustrates the estimation results. Overall, the mitigating effect of additional publications on the citation penalty is non-monotonic. Strikingly, among authors with only one publication in a top-five journal, the citation penalty for non-White scholars is higher than that among authors with no prior publications. This suggests that an additional publication, even in a top journal, tends to benefit all-White teams more than teams that feature at least one Black, Hispanic, or Asian member.

Moreover, within a dataset encompassing over 200 journals, non-White researchers see a reduction in the citation penalty when they have at least 20 prior publications in total, at least 11 publications

across the 35 high-impact journals, or at least six publications across the 16 high-impact journals or the top 5 journals. Authors with publications in all 35 high-impact journals or all 16 high-impact journals need at least 11 prior publications in the set of 35 high-impact journals or six publications in the top 16 high-impact or top 5 journals to see a noticeable reduction in the citation gap. In the case of the top 5 journals, at least six additional prior publications are required to significantly reduce the citation penalty.

Therefore, while additional publications appear to have an impact, their significance in reducing the gap clearly depends on the type of journal, even when controlling for the journal of the paper under consideration. More importantly, this effect does not become evident until a critical mass of publication records is reached.[16]

Typically, factors like affiliation ranking, journal of publication, and author prominence are considered reliable indicators of quality. However, our study finds that considering these factors only reduces the citation gap between White and non-White scholars at the extreme end of the distribution of publication records.

4.1.2 Past citations as a proxy for quality and potential impact

To substantiate the argument that the quality differential or impact is likely to play a minor role in explaining the citation gap, we use the author's citation history at the time of publication as a proxy for author quality and, in turn, potential article quality. More specifically, for each article, we calculate the average number of citations accumulated by the authors prior to its publication.

Table VI presents the results of the regression analysis, incorporating a non-linear form of the team's past citations into the baseline equation from section 3.1. The variable "At least one Black-/Hispanic/Asian" consistently shows a significant negative coefficient across all journal groups, indicating a citation penalty for papers authored by these groups. However, this penalty is reduced by 1 to 3 percentage points compared to the baseline. Specifically, in the more than 200 journals category, the penalty is 4.3 log points, while it is 6.8 log points in the top 35 journals, 6 log points in the top 16 journals, and 8.3 log points in the top 5 journals. Similarly, the "Other races" category also displays negative and significant coefficients across all journal groups. We also observe that past citations are strong and positive predictors of the citations for current articles.

In Appendix Table C.11, we further refine our analysis by disaggregating the outcome variable to consider the annual citation count of each article. Specifically, for each year, we calculate the number of citations an article receives by extracting references from our database. In addition to the control variables previously considered, we introduce several new controls: cumulative previous citations (citations accumulated by the article up to, but not including, time t); the average number

[16] A similar analysis examining the citation gap as a function of author seniority—measured by the number of years between the first publication of the more senior coauthor and the publication year—reveals that the gap closes only at very high levels of seniority, typically after at least 30 years (or 20 years for authors publishing in Top 5 or Top 16 journals). This finding is consistent with the notion that a substantial accumulation of publications over time is required for the citation gap to narrow (Appendix Figure D.1).

of publications in the top 5, top 16, and top 35 journals; as well as total publications by the authors at time t. We also control for the average cumulative previous citations of the authors up to time t, excluding time t and the current article citations. This specification aims to account for both the quality of the authors and the perceived quality of the paper prior to receiving citations at time t. The trend remains consistent: teams with at least one Black, Hispanic, or Asian member consistently show a citation penalty compared to teams composed entirely of White authors. The estimates are relatively lower because it is an annual reduction.

Overall, even after controlling for all of these factors, the tables illustrate a persistent disadvantage for articles authored by non-White researchers when it comes to citations. This disadvantage remains significant even when accounting for citation history, which is endogenous and likely to introduce downward bias.

4.1.3 Editorial-based quality differential

If citations serve as a proxy for quality (albeit an imperfect one), then a lower citation count may be indicative of a lower-quality paper. Tables III and C.10 reveal that the differences are more pronounced among publications in top economics journals (or among publications by authors affiliated with top-ranked institutions, and the differences are persistent even conditioning on the authors' publication records). This contradicts the argument that papers from non-White scholars are of a lower quality, at least using traditional proxies of quality. However, consider a publisher that accepts a paper if its expected quality meets a certain threshold $\bar{q}$. If a policy similar to affirmative action, which has been criticized for potentially lowering the threshold for certain racial groups (Arcidiacono et al. [2023]), is implicitly applied in academic publication, or if it is perceived that such a policy exists, then the threshold (or the perceived threshold) for non-White authors, $\bar{q}_{nonWhite}$, would be lower than the threshold for White authors, $\bar{q}_{White}$, suggesting that papers authored exclusively by White scholars would be (or would be perceived to be) of a higher quality.

While there is no direct evidence of such phenomena in academic publishing, the emphasis on diversity and inclusion initiatives may influence perceptions and beliefs regarding the quality of papers authored by non-White scholars. This could inadvertently have the same effect that would be seen if such a policy were to actually be implemented, resulting in a lower perceived quality threshold for papers by non-White authors ($\bar{q}_{nonWhite} < \bar{q}_{White}$).

However, two points suggest that this scenario is unlikely. First, our data extends back to the 1950s, a period in which affirmative action policies were less common. During this time, non-White scholars who successfully entered the publication world would likely have been at least as competent as their White counterparts ($\bar{q}_{nonWhite} \geq \bar{q}_{White}$). Second, if anything, criticisms of differential quality-based admission policies have typically focused on Blacks and Latinx, with some evidence, as previously mentioned, indicating that Asians may face discrimination under such policies. Therefore, we would not expect papers with Asian authors to be perceived as of lower quality.

Overall, it seems unlikely that the editorial-based quality differential favoring non-White authors

drives the observed results.

4.2 Citation patterns and racial similarity

We investigate the extent to which racial similarities between authors and citing scholars influence citation patterns. The results, detailed in Table VII, distinguish between citations from non-White and White scholars, revealing a pronounced homophilic pattern. Specifically, the analysis shows that papers authored by at least one Black, Hispanic, or Asian individual receive a citation premium of approximately 23% from non-White scholars compared to papers authored exclusively by White scholars. Conversely, this same group (papers authored by at least one Black, Hispanic, or Asian individual) experiences a citation penalty of around 30% from White scholars. These findings indicate a significant homophily effect, where citation practices are more likely to be influenced by the racial composition of the author team.

Figure VII looks further at this homophilic effect, depicting citation patterns over time and across different fields based on the race of the authors behind citing papers. The figure shows a consistent citation premium for papers authored by non-White scholars when cited by non-White scholars; this premium is evident across all time periods but features a notable increase in the 1990s. In contrast, citation patterns from White scholars consistently showcase a penalty for papers authored by non-White scholars.

A comparison of Figure V with Panel a) of Figure VII reveals that the decline in the overall citation gap since the 1990s is accompanied by a general increase in the propensity of non-White authors to cite other non-White authors as well as an increase in that of White authors to cite other White authors. This pattern suggests that the observed reduction in the citation gap does not necessarily reflect a broad shift toward more inclusive citation practices across racial groups; rather, it reflects intensified intra-group citation behavior.

This pattern persists when analyzing citation patterns by field and methodological approach. Even among empirical papers, where the citation penalty for teams with at least one Black, Hispanic, or Asian member (relative to all-White teams) is the lowest, the trend persists: Non-White scholars are more likely than White scholars to cite papers authored by non-White scholars. The magnitudes are also quite substantial, with penalties consistently exceeding 25 log points and premiums surpassing 15 log points. We also specifically examine the subfields with the highest proportions of non-White scholars, as shown in Appendix Figure D.2, and find that the homophilic pattern persists.

Race may also be associated with spatial and geographic differences. In other words, what may appear as a citation preference for individuals of the same race could instead be attributable to geographic factors that influence access to articles. To test this hypothesis, we analyze authors' institutions, focusing on a sample of papers that have at least one author based in the United States (U.S.) and are cited by other U.S.-based authors. The choice of the U.S. here is significant given the country's central role in the economics field. The results, presented in Appendix Tables C.12 and C.13, indicate that the findings remain virtually unchanged, suggesting that homophily is not likely

driven by geographic differences.

Appendix Figure D.3 demonstrates that the homophily pattern is robust but somehow sensitive to the set of control variables, although the qualitative trends and signs remain consistent. While we cannot fully rule out unobserved quality as a contributor to the racial citation gap, the observed homophily as well as the results from Section 4.1 suggest that perceived quality may vary across racial groups, with individuals potentially assigning higher value to work from their own group. Ultimately, determining which perception is accurate is inherently unobservable.

4.3 Summary of findings

Overall, our findings may be summarized as follows: (1) There is a persistent racial citation gap that is evident in both direct and indirect citations; (2) this gap is associated with higher clustering coefficients among non-White scholars, indicating smaller and more insular citation networks; (3) the gap is more pronounced among elite journals (top 5) and top-tier institutions, and it is not mitigated by publication records except at extreme levels of productivity; (4) the gap persists even when accounting for potential measurable differences in paper quality; and (5) the gap is marked by strong homophilic patterns in citation behavior.[17]

Motivated by these results, we propose a conceptual framework that assumes equal absolute quality across racial groups and highlights the role played by clustered networks as a key contributor to the citation gap. We demonstrate how such network structures—through various mechanisms, such as information frictions, perceptions, and preferences—can give rise to the observed homophilic patterns and disparities in citation outcomes. We then provide empirical tests informed by this framework.

4.4 A clustered network hypothesis

4.4.1 A simple conceptual framework

We develop a simple model of clustered networks to explain the racial citation gap. We characterize those clustered networks using parameters for preferences, perceptions, and the cost of information.

Concretely, consider the following setup: There are two ethnic groups of authors, G_1 (minority) and G_2 (majority), with group sizes $n_1 = |G_1|$ and $n_2 = |G_2|$, where $n_1 < n_2$. Let author i cite s_1 papers authored by members of G_1 and s_2 papers authored by members of G_2.

Scholars are assumed to be more familiar with work produced within their own racial or professional networks, shaped by shared nationality, regional conferences, and other social connections. Additionally, ideas that resonate more closely with a referee's experiences are likely to be more valued. So, it may be easier to assess and appreciate the value of papers authored by scholars from

[17] Since prior studies use variables such as publication in a top-five journal, affiliation with a highly ranked department, and attendance at a top-ranked PhD program as proxies for ability (Carrell et al. [2024]), our findings suggest that the racial citation gap we document cannot be attributed to differences in author quality.

the same group. We model the perceived value of a paper for a referee from the minority group as $\alpha[a\ln(s_1)+b\ln(s_2)]$, and for a referee from the majority group as $\alpha[b\ln(s_1)+a\ln(s_2)]$, where $a, b > 0$, $a+b=1$ and $a \geq b$ (relative recognition). Furthermore, the choice of s_1 and s_2 may be influenced by anticipated referee preferences or the likelihood of being evaluated by referees from a particular group. The resulting average perceived quality of a paper is:

$$v(s_1, s_2) = \frac{n_1}{n_1+n_2}\alpha[aln(s_1)+bln(s_2)] + \frac{n_2}{n_1+n_2}\alpha[bln(s_1)+aln(s_2)]$$

Writing a paper entails nontrivial costs. For instance, accurate citation involves effort related to literature review, staying current with the field, and engaging in professional networks. Let $c(s_1, s_2)$ denote the total cost incurred by author i. Assuming that authors are more familiar with work by co-ethnic scholars implies a lower marginal cost of citing co-ethnic papers relative to non-co-ethnic ones. We further assume that authors invest more effort when targeting higher-quality journals, indexed by q, so that writing costs increase with journal quality. Conversely, higher-ability authors (indexed by r) incur lower costs, all else equal. The total cost function is given by: $c(s_1, s_2) = h(q, r)g(s_1, s_2)$ where h and g are two arbitrary functions, where h is increasing in q and decreasing in r and g is increasing in both s_1 and s_2.

The author i's maximization problem therefore, consists of the following:

$$\max_{(s_1, s_2)} \quad u(s_1, s_2) = v(s_1, s_2) - c(s_1, s_2)$$

The details and resolution of this problem are provided in Appendix B, considering different functional forms. The analysis produces two key results:

Proposition B.2. (Racial citation gap) *When the citation costs are symmetric, $E(s_1^* - s_2^*) < 0$, that is, papers authored by majority scholars are, on average, cited more frequently than papers authors by minority scholars.*

Proposition B.3. (Racial homophily gap) *Let H_i denote the homophily level of group i. The homophily level is greater for majority authors than for minority authors (that is, $\frac{H_2}{H_1} > 1$) if and only if the cost of citing the work of a non-co-ethnic scholar relative to a co-ethnic scholar (defined as $\frac{c_{i'}^i}{c_i^i}$) is sufficiently larger for majority authors compared to minority authors; that is, if and only if $\frac{c_1^2}{c_2^2} > K\frac{c_2^1}{c_1^1}$ where $K = (\frac{k_1 n_2}{k_2 n_1})^2 \geq 1$.*

Proposition B.2 establishes the existence of a racial citation gap: Under reasonable assumptions regarding citation costs and group sizes, papers by majority-group authors receive more citations than comparable papers by minority-group authors. This gap depends on the degree to which minority-group authors prefer citing their co-ethnic peers and whether that preference is sufficient to counterbalance the analogous bias among majority-group authors.

Proposition B.3 examines the homophily present in the racial citation gap and derives a necessary and sufficient condition under which majority-group scholars display stronger homophily than

minority scholars: They exhibit greater homophily if and only if their cost of citing non-co-ethnic authors—relative to co-ethnics—is sufficiently higher than that of minority-group scholars. This threshold is characterized by a factor K, which depends on both referee preferences and group size ratios. Notably, even if majority scholars face higher relative citation costs, minority scholars may still exhibit stronger homophily when K is sufficiently large. The factor K reaches its minimum ($K = 1$) under referee neutrality ($a = b$), and its maximum ($K = (n_2/n_1)^2$) when referees are maximally biased ($a = 1$, $b = 0$). This highlights a tradeoff between citation costs and the likelihood of being evaluated by a co-ethnic referee, particularly under neutral evaluation.

Our analysis may help explain why White authors tend to exhibit stronger citation homophily than non-White authors, whereas in other studies, women show greater homophily than men [Koffi, 2025]. We replicate those findings in our context in Appendix Table C.14. The "cost" of citing across identity lines may be lower for gender than for race. This may reflect greater familiarity and interaction across genders—both in society and within the profession—compared to across racial groups. For instance, most households and departments are gender-mixed, while racial diversity remains more limited.[18] This is also highlighted when looking at collaborative patterns in our sample: 75% of papers with women are mixed-gender, while only 50% of papers with non-White authors are mixed-race.

Furthermore, the model suggests that, given the ratio of estimates capturing citation homophily, $\left|\frac{\beta_{\text{nonWhite by nonWhite}}}{\beta_{\text{nonWhite by White}}}\right|$, ranges from 0.6 to 0.95, preference-based factors may play a substantial role in shaping citation behavior.

Building on this conceptual framework, we next evaluate the clustered network hypothesis. Specifically, the following subsections empirically assess the roles of information costs, perception, and preferences as potential explanations for the observed racial citation gap.

4.4.2 Online availability and racial disparities in citations: Testing the information cost hypothesis

A substantial body of works has shown that open access enhances the use of scientific knowledge (Furman and Stern [2011], Williams [2013], Murray et al. [2016], Staudt [2020], Bryan and Ozcan [2021], Biasi and Moser [2021]). This supports the hypothesis that open access reduces the cost of accessing scientific knowledge and promotes its dissemination. In economics, for instance, McCabe and Snyder [2015] found that online availability can increase citations by up to 10%.

Building on our conceptual framework, if the marginal cost of citing a non-co-ethnic author is higher than that of citing a co-ethnic author, then a shock that lowers information frictions (e.g., online availability) could reduce barriers to accessing work by non-majority scholars and narrow citation disparities. However, such a shock could also plausibly amplify citation inequalities initially, especially if dominant-group authors benefit disproportionately from increased visibility. To investi-

[18] A report from the Brookings Institution (Wessel et al. [2020]) indicates that as of 2020, 19% of economics faculty in academia identify as Black, Hispanic, Asian, or other minority groups, while 27% are women.

gate which mechanism is more consistent with the data, we leverage data from McCabe and Snyder [2015] that includes a panel of 71 economics journals. For each journal volume published before 2005, the dataset identifies the year in which the volume first became available online through major aggregators (e.g., JSTOR, EBSCO, ProQuest, Ingenta, Gale, OCLC, DigiZeitschriften) or journal websites.

We employ an event-study framework to examine how online availability affects the citation gap between papers authored by at least one Black, Hispanic, or Asian scholar and those authored exclusively by White scholars. Specifically, we estimate the following model:

$$C_{p,T,j,t} = \lambda_0 + \lambda_1 R_p + \sum_{\tau=-m}^{m} \lambda_{2,t} \mathbb{1}\{OA_{j,t} = \tau\} + \sum_{\tau=-m}^{m} \lambda_{3,t} R_p * \mathbb{1}\{OA_{j,t} = \tau\} + X'_{p,t}\pi + \alpha_p + \delta_t + \sigma_{j,T} + \varepsilon_{p,c,t} \tag{2}$$

$C_{p,T,j,t}$ denotes the number of citations received at time t by paper p, which was published in journal j in year T. The variable R_p captures the racial composition of the authors of paper p. The vector $X_{p,t}$ represents a set of time-varying, paper-level controls, including the authors' publication history and the age of the paper (modeled using its first, second, third, and fourth powers). The terms $\mathbb{1}\{OA_{j,t} = \tau\}$ are indicator variables denoting the number of years relative to the online availability of volume from year T in journal j. The parameter m defines the event window under consideration. The coefficient at $\tau = -1$ is normalized to zero and serves as the reference period. Accordingly, the parameters of interest, $\lambda_{3,\tau}$, capture changes in the racial citation gap at each event time τ relative to the year immediately preceding online availability. The model includes a comprehensive set of fixed effects: paper fixed effects, publication year-by-journal fixed effects, and citation-year fixed effects.

Identification relies on the assumption that there are no unobserved shocks, correlated with online availability, that differentially affect citation patterns for White and non-White authors, conditional on the included controls.

Figure VIII presents the event-study estimates. We observe no significant pre-trends in the citation gap during the seven years prior to online availability, providing support for the parallel trends assumption. At the time of online availability (year 0), the citation gap widens sharply, followed by a sustained narrowing beginning around year 2. From year 2 onward, the coefficients are positive and statistically significant, indicating a persistent increase in citations for papers authored by non-White scholars relative to White scholars. Specifically, citation differences rise from negative values in year 1 to reach approximately 0.04 on the asinh-transformed scale by years 4 through 7. These results suggest that online availability eventually benefits non-White scholars, but that the short-run effects may initially favor the majority group: White authors.

Appendix Table C.17 presents aggregate estimates and robustness checks using alternative control sets. The findings are consistent across specifications, including when we shift from paper fixed effects and time fixed effects only to the addition of journal times and year of publication fixed effects, or add author citation history.

Interestingly, the same pattern—an initial increase in the citation gap followed by a decline—is observed when examining citations from both White and non-White scholars separately. This weakens the interpretation that information frictions alone explain the observed dynamics. If access costs were the primary constraint, we would expect non-White scholars to increasingly cite majority-group work after online access becomes available. Instead, in periods 0 and 1, we observe a sharp rise in citations to White authors. But in later periods, non-White scholars begin to cite other non-White scholars at a growing rate, suggesting the importance of in-group preferences. For White scholars, the delayed increase in citations to non-White authors (post-year 2) points to longer-term reductions in informational frictions. In contrast, the earlier changes favoring other White scholars suggest a stronger role for in-group preferences and less so for visibility dynamics, given that the results are virtually unchanged when adding authors' citation history.

On net, the citation gap increases by approximately 4–4.3 log points in years 0–1 and decreases by 2.9 log points after year 2. The overall net effect over the seven-year post-treatment period is a reduction of about 1 log point—nearly half the racial citation gap observed in this subsample.

In summary, while information frictions contribute to the observed citation gap, they do not fully account for the differential patterns in citation behavior following online availability. Behavioral mechanisms—such as group-based preferences—likely play an important role. Appendix Figure D.4 confirms the robustness of these findings using the estimator developed by Sun and Abraham [2021].

4.4.3 Visibility and citation behavior: Reassessing the role of information costs

Seminar and conference attendance, visiting appointments, and affiliation to prestigious research institutions can significantly affect citation rates by increasing research visibility and dissemination, ultimately reducing the information cost. Such events give scholars a platform from which to present their work to a broader audience, fostering direct engagement with peers who may later cite their research. Moreover, Colussi [2018] and Carrell et al. [2024] show that social ties significantly influence publication outcomes in economics, underscoring the role of informal networks in shaping academic success. However, non-White researchers may still encounter challenges in gaining visibility for their work, as proposed by Doleac et al. [2021].

To test this hypothesis, we first extracted data on seminar and conference invitations from the database. We then created variables relating to the total count of seminars and conferences, attendance at prestigious conferences (e.g., NBER), and invitations to top-ranked institutions. Since this information is only available for papers referencing potential funding, the results may be subject to selection bias. Therefore, we essentially compare the coefficients' magnitude by analyzing the base coefficient without controls for seminars and conferences in this subset of the database.

The results, presented in Table VIII, show that variables related to seminars and conferences account for, at most, 0.006 log points. The coefficients are insignificant, likely due to the small sample size.

Recognizing the non-comprehensiveness of seminars at the paper level and the fact that citations

may be influenced by an author's visibility at seminars and conferences (the author's broader academic pipeline), we conducted an extensive manual data-collection process, gathering detailed CV information on their participation in such events from the authors wherever possible.[19] The authors considered here all have at least one publication in a top 5 journal after 1990, as with our manual data-collection process for racial information. Based on this data, we created variables to capture the opportunities for the authors' work to be disseminated. The results show that these variables explain only a marginal portion of the coefficient—approximately 0.002 log points (Table C.15, Columns (1)–(4)).

Additionally, we leverage the full scope of our hand-collected database to enhance the model by incorporating variables that are likely to influence author visibility (and, in turn, reduce the cost of information): editorial roles, visiting positions, memberships in prestigious organizations (e.g., NBER, JPAL, CEPR, BREAD, IZA, CIFAR), and PhDs from top-tier institutions. Combined with the seminar variables, this comprehensive model accounts for 1.5 log points of the variation in the racial citation gap, but the difference is not statistically significant, and most of the variation is attributable to disparities in memberships within prestigious associations.[20]

Overall, differences in seminar participation, conference attendance, editorial positions, and PhD-granting institutions do not statistically reduce the racial citation gap. Only membership in prestigious associations—potentially reflecting access to new professional networks—shifts the coefficient, though the effect is also not statistically significant. While authors' outreach efforts may influence citation outcomes, these factors appear insufficient to account for the observed gap.

4.4.4 Cross-group collaboration and citation behavior: A shift in perceptions?

To further test the hypothesis of clustered networks, we rely on the literature on contact theory. This literature explores how interactions between members of different groups can influence the biases and perceptions the majority group holds toward a minority group. The theory suggests that intergroup contact aids in diminishing stereotypes and promoting stronger mutual understandings, and it has been tested in a diverse range of field experiments (Boisjoly et al. [2006], Carrell et al. [2019], Dahl et al. [2021] among others).

In this study, we draw on a concept analogous to contact theory in the context of academic publishing to examine how collaboration may help reduce barriers across groups. Specifically, we investigate the impact of initiating a collaboration with an out-group member (a non-coethnic author). Our empirical strategy employs an event study framework in which the "event" is defined as an

[19] In many papers, we encounter the phrase, "We thank numerous seminar and conference participants…" without explicitly listing the names of the seminars and conferences.

[20] Overall, seminars and conferences only marginally explain the citation gap. However, this should not be taken to mean that seminar and conference attendance is not a key factor behind citation patterns. In fact, the positive signs of the relevant variables indicate that participation in these events does, in fact, increase citations. Furthermore, univariate regressions (Appendix Table C.16) reveal that seminar and conference variables explain nearly 8.3 log points of the racial citation gap. However, participation in these events is also indicative of article quality and author prominence, which the other control variables already account for, leaving this additional variable a marginal role despite its inherent relevance.

author's first coauthorship with a non-coethnic collaborator. We then track all subsequent publications involving only coethnic coauthors, truncating the sample at the point of the next non-coethnic collaboration.

Because we observe the publication date rather than the precise start of the collaboration, we conservatively assume that collaboration begins three years prior to the publication date. As we demonstrate later, however, our results are robust to removing this timing assumptions.[21]

Our outcome variable, denoted by $ShareNonCoethnicReference_{i,t}$, is the share of bibliographic references of papers published by author i in publication year t in which at least one cited author is from the out-group. The corresponding estimation equation is presented below.

$$ShareNonCoethnicReference_{i,t} = \alpha_0 + \sum_{\tau=-m}^{m} \alpha_{2,t} \mathbb{1}\{NonCoethnicCollab_{i,t} = \tau\} + X_{i,t}^{'}\pi + \gamma_i + \delta_t + \varepsilon_{i,t} \tag{3}$$

where $\mathbb{1}\{NonCoethnicCollab_{i,t} = \tau\}$ is equal to 1 if we are τ years of publication relative to a non-coethnic collaboration. The coefficient at $\tau = -1$ is normalized to zero, serving as the reference period. Accordingly, all estimated coefficients can be interpreted as the effect of non-coethnic collaboration at each event time relative to the year of publication immediately preceding the collaboration. We control for a range of potential confounders (X_{it}), including average team size in publication year t, and granular controls and fixed effects for the number of publications at time t and up to year t, as well as author fixed effects (γ_i), publication year fixed effects (δ_t).

Unlike field experiments aimed at testing contact theory, the interpretation of our setting presents several challenges due to the endogenous nature of collaboration. For instance, an author may choose to collaborate with an out-group member as part of a deliberate effort to alter their citation behavior, or such a collaboration may itself stem from a prior shift in behaviors or attitudes. Conversely, collaboration with a non-co-ethnic co-author may also trigger a behavioral change by improving perceptions or reducing informational frictions. Importantly, none of these mechanisms invalidates the broader implications of a clustered citation network; in other words, the citation gap between co-ethnic and non-co-ethnic scholars is not fully attributable to differences in paper quality alone. However, these alternative pathways must be considered carefully to derive a causal interpretation of non-co-ethnic collaboration. In particular, for collaboration with a non-co-ethnic co-author to have a causal effect on subsequent citation behavior, it must be the case that: (i) authors do not alter their citation behavior in anticipation of the collaboration; and (ii) conditional on covariates, the assumption of parallel trends holds (i.e., in the absence of collaboration, the trajectory of citation outcomes would have been similar across treated and untreated authors).

[21] Hadavand et al. [2024] report that the average time from submission to acceptance in four Top Five journals and Review of Economics and Statistics was 24.72 months in 2012–2013, while the average time from submission to publication was 33.15 months. In contrast, non–Top Five journals tend to exhibit shorter median submission-to-acceptance durations: for example, 12 months for the Journal of Development Economics, 15 for the Journal of Econometrics, 19 for the Journal of International Economics, 18 for the Journal of Labor Economics, and 12 months for both the Journal of Monetary Economics and the Journal of Public Economics.

Our empirical results are consistent with these identifying assumptions and suggest that the observed changes in citation behavior may plausibly reflect the effects of collaboration.

Figure IX presents the estimation results. Prior to collaborating with a non-coethnic author, the reference listing patterns of authors who eventually engage in such collaborations are similar to those of authors who never do. However, following the initial collaboration, we observe an increase in the share of bibliographic references to non-coethnic authors. This effect persists above the pre-collaboration trend, though it tends to attenuate over time.

Appendix Table C.18 quantifies this effect, showing that the average share of references to non-coethnic authors increases by 0.016 after treatment (Column 1), which corresponds to 10% of the pre-collaboration median value of 0.16. When we examine heterogeneity by author ethnicity, we find that White authors appear to be the primary drivers of this shift (Column 2), although the pre-trend is not insignificant. Nonetheless, the coefficient for non-White authors, while not statistically significant, is not negligible.

One potential concern is that the observed effect may simply reflect a shift in the relevant literature cited as a result of the collaboration, reinforcing the information-friction mechanism. To address this, we show in Column (4) that the effect persists after controlling for whether subsequent papers share the same JEL codes as the initial collaborative paper—and at least 50% of the follow-up papers fall outside the original JEL classification. Similarly, Column (5) demonstrates that the effect is robust against controlling for the evolving share of non-White scholars within the relevant field over time (the sample size changes because the JEL code is not available for all the papers).

To further support our interpretation, we implement a matching procedure to ensure comparability between treated and untreated authors prior to collaboration. Specifically, we use propensity score matching based on pre-treatment outcomes (i.e., the level and share of references to non-coethnic authors cited in the references) and the year of first publication of the author. Panel A of Appendix Figure D.5 presents post-matching results that mirror the main trend observed in the unmatched sample.

Finally, Panel B of Appendix Figure D.5 confirms that our results are robust to accounting for treatment heterogeneity in the staggered difference-in-differences design (Sun and Abraham [2021]).

Taken together, the persistence of the effect across different specifications, its attenuation over time, and its robustness against matching suggest that collaboration may influence beliefs and perceptions, as in contact theory. However, in line with findings from prior studies (e.g., Dahl et al. [2021]), sustaining long-term changes may require consistent exposure or ongoing collaboration.

4.4.5 The impact of author order: A window into explicit preferences and citation bias

Here, we leverage the convention in the field of economics whereby authors are typically listed alphabetically. This practice carries the implicit assumption that author order is not dictated by contribution and, thus, should not be correlated with paper quality. However, while alphabetical ordering neutralizes contribution-based inferences, the position of an author in the list may still in-

fluence a paper's visibility, particularly when it comes to citation practices for which readers often only remember or refer to the first-listed author.[22] If race—or the perception of race—influences citation behavior, this visibility may interact with racial identity to amplify or mitigate citation patterns.

Appendix Table C.19 presents the results. We focus on a subsample of papers featuring at least one non-White and one White coauthor. Panel A reports estimates at the paper level, while Panel B reports estimates at the author level.

In Panel A, Column (1) shows no statistically significant difference in overall citations between mixed-race papers with a non-White first-listed author and those with a White first-listed author, reinforcing the validity of alphabetical ordering as orthogonal to paper quality. However, once we disaggregate citations by the racial composition of the citing team (Columns (2)–(7)), a clearer pattern emerges. Mixed-race papers with a non-White first-listed author receive more citations from teams featuring non-White authors but fewer citations from all-White teams. More precisely, papers with non-White first-listed authors are cited 14.8 log points less by all-White citing teams than those with White first-listed authors (Column (5))—a penalty that is statistically significantly larger than that observed when the non-White co-author is the second- or third-listed author. These patterns suggest that all-White citing teams respond more negatively to the prominence of a non-White author when they occupy the first-author position (the "et al." effect) and that non-White citing teams are similarly responsive to a non-White author's position.

Panel B presents a complementary analysis at the author level. For each author, we restrict the sample to instances in which they appear in a mixed-race paper and include author-fixed effects. This design enables us to estimate how the same author's citation count varies based on their position in the author list. Once again, Column (1) provides no evidence of first-author status affecting citations in general, conditional on author identity. However, Columns (2)–(7) reveal the same asymmetry as Panel A: When a non-White author is listed first, they receive significantly more citations from non-White citing teams (approximately 6 log points more in Column (2)) but fewer citations from all-White citing teams—a penalty of 14 log points (Column 3) compared to instances in which that same author is listed in a position other than first. In other words, a non-White scholar may receive more citations from White scholars when a White coauthor is the first-listed author, even if their own contribution remains unchanged.

The consistency of these findings across specifications and levels of analysis indicates that author visibility interacts with racial identity in shaping citation outcomes. The results support the view that race, independently of paper quality, plays a meaningful role in determining how and by whom scholarly work is disseminated. This suggests that citation behavior in economics reflects not only informational factors or academic merit, but also social preferences and positional salience along

[22]Prior work has highlighted the relevance of alphabetical ordering to citation frequency (see Huang [2015], Feenberg et al. [2017], Weber [2018]). More recently, Brogaard et al. [2023] used the alphabetical-ordering norm as an identification strategy to examine how an author's pre-existing prominence can shape their citation outcomes. Building on these insights, we test whether author position interacts with racial identity to influence how papers are cited—specifically in contexts where citing authors may or may not share the racial background of those being cited.

racial lines.

4.5 Discussion: On the potential losses of racial diffusion of ideas

In this section, we discuss potential losses in the diffusion of ideas and innovation that could stem from racial barriers. The idea that life experiences shape research perspectives has long been discussed in sociology and psychology. To examine this, we analyze research topics using three-digit JEL codes. Within each JEL code, we cluster papers into groups of relatively small average size (200 observations in the baseline) based on their embedding representations. For each cluster, we treat racial categories as subgroups and analyze the semantic differences across these groups. This aims to analyze more than research topics but the specific angle they pursue within a homogenous cluster of research questions.

First, we analyze the linguistic proximity of each racial category to others. To do this, we employ a term frequency-inverse document frequency (TF-IDF) metric, grouping articles by racial category. We refrain from using advanced similarity techniques relying on embeddings because all representations address the same research questions; our goal is to capture variation in terminology within these questions. Figure D.6 in the Appendix shows the distribution of similarity, using the all-White author category as the reference group. We observe substantial variability in similarity across groups. Specifically, the distribution for teams including Black scholars is left-skewed, with similarity values often dropping below 0.2. Similar patterns emerge when comparing teams with Hispanic or Asian researchers, where similarity is not concentrated near 1. This suggests notable differences in terminology—and potentially in the approach to research questions—across racial groups.

Panel (d) of Appendix Figure D.6 presents an additional finding: the degree of variability in approaches to research questions with respect to the share of non-White researchers follows an inverted U-shape. In other words, as the pool of researchers working on a given question becomes more heterogeneous, so does the variety in topical approaches to a given research question.

To better illustrate these findings, we remove the most commonly used words (those defining the core research questions) within each cluster under a three-digit JEL code and analyze the remaining terms across racial categories to capture the specific angles addressed by each racial group. We provide examples from the database in Appendix Figures D.7 and D.8. One can see, for example, in Panel A, based on JEL code J15 (Economics of Minorities, Race, Indigenous Peoples, and Immigrants; Nonworkplace Discrimination), that the papers focus on research related to child education. Here, all-White teams emphasize school environments and the broader racial gap, while teams with non-White scholars concentrate on class size and peer effects. For another example, one can look at the study of Chinese exports in Panel F; White scholars often examine competition's effects on firms, while non-White scholars take a broader approach, addressing the impacts of technological processes, productivity, and innovation.

Overall, our analysis reveals that scholars from different racial backgrounds bring distinct perspectives to research questions, resulting in varied angles of investigation and emphasis. In line with

Page [2008], our findings indicate that by failing to build on the unique ideas contributed by scholars of various racial backgrounds, we risk overlooking valuable insights that could enhance economic theory, public policy, and social progress.

5 Conclusion

Our analysis reveals significant differences in the way research produced by White and non-White scholars is perceived and cited in economics. Papers with at least one non-White author receive 5.1% to 9.6% fewer citations. While the citation penalty is relatively consistent across non-White authors in the full sample of papers, it tends to be larger for Black authors in papers published in top journals. These findings are robust even after controlling for a wide range of paper characteristics, such as the number of authors, field, journal, and year of publication, as well as prestige and quality-related factors like author affiliation, academic prominence, disaggregated fields, and prior citations (Carrell et al. [2024]). The findings are also robust to controlling for additional variables such as seminar and conference attendance, attendance at prestigious conferences, seminar invitations from top-ranked institutions, visiting appointments in top-tier institutions, editorial roles, and membership in prestigious organizations. Over time, the racial citation gap has narrowed, reflecting an increase in non-White co-authorship and a shift towards empirical economic research over the past three decades (Angrist and Pischke [2010]).

The analysis of centrality to assess paper influence in citation networks reveals that papers authored by non-White scholars not only receive fewer direct citations, but are also less likely to act as a bridge along the shortest citation path connecting two other papers and are less likely to be cited by other highly cited papers. The analysis also implies that papers that cite the works of non-White economists are generally not highly cited, which denies them any indirect influence.

To explain the racial citation gap in economics, we explore various factors. Notably, we find insufficient evidence to suggest that differences in article or author quality fully account for the gap. However, our analysis reveals significant in-group citations, with non-White authors more likely to cite other non-White authors, and White authors more likely to cite other White authors. In a theoretical framework, we demonstrate that this homophily can arise from clustered networks characterized by factors such as the cost of citation—whether due to a genuine lack of awareness or bias—as well as perception. Finally, we show that reducing information costs has decreased the citation gap by up to 50%, although preference and bias seem to be prominent features of the gap. Overall, we may be missing valuable opportunities for the advancement of economic research by not building upon the distinct ideas contributed by each racial group. This raises the importance of connecting those distinct networks.

References

Amarante, V., Burger, R., Chelwa, G., Cockburn, J., Kassouf, A., McKay, A., and Zurbrigg, J. (2022). Underrepresentation of developing country researchers in development research. *Applied economics letters*, 29(17):1659–1664.

Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2020). Inside job or deep impact? extramural citations and the influence of economic scholarship. *Journal of Economic Literature*, 58(1):3–52.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, 24(2):3–30.

Arcidiacono, P., Kinsler, J., and Ransom, T. (2023). What the students for fair admissions cases reveal about racial preferences. *Journal of Political Economy Microeconomics*, 1(4):615–668.

Bayer, A., Hoover, G. A., and Washington, E. (2020). How you can work to increase the presence and improve the experience of black, latinx, and native american people in the economics profession. *Journal of Economic Perspectives*, 34(3):193–219.

Bayer, A. and Rouse, C. E. (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4):221–242.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61.

Bertolero, M. A., Dworkin, J. D., David, S. U., Lloreda, C. L., Srivastava, P., Stiso, J., Zhou, D., Dzirasa, K., Fair, D. A., Kaczkurkin, A. N., et al. (2020). Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. *BioRxiv*, pages 2020–10.

Biasi, B. and Moser, P. (2021). Effects of copyrights on science: Evidence from the wwii book republication program. *American Economic Journal: Microeconomics*, 13(4):218–260.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or antipathy? the impact of diversity. *American Economic Review*, 96(5):1890–1905.

Brogaard, J., Engelberg, J. E., Eswar, S. K., and Van Wesep, E. D. (2023). On the causal effect of fame on citations. *Management Science*.

Bryan, K. A. and Ozcan, Y. (2021). The impact of open access mandates on invention. *Review of Economics and Statistics*, 103(5):954–967.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2020). Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics*, 135(1):269–327.

Card, D., DellaVigna, S., Funk, P., and Iriberri, N. (2022). Gender differences in peer recognition by economists. *Econometrica*, 90(5):1937–1971.

Carrell, S., Figlio, D., and Lusher, L. (2024). Clubs and networks in economics reviewing. *Journal of Political Economy*, 132(9):2999–3024.

Carrell, S. E., Hoekstra, M., and West, J. E. (2019). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy*, 11(4):159–182.

Chen, J., Liu, Q., and Kim, M. (2022). Gender gap in tenure and promotion: Evidence from the economics ph. d. class of 2008. *Southern Economic Journal*, 88(4):1277–1312.

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Colussi, T. (2018). Social ties in academia: A friend is a treasure. *Review of Economics and Statistics*, 100(1):45–50.

Dahl, G. B., Kotsadam, A., and Rooth, D.-O. (2021). Does integration change gender attitudes?

the effect of randomly assigning women to traditionally male teams. *The Quarterly journal of economics*, 136(2):987–1030.

Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Doleac, J. L., Hengel, E., and Pancotti, E. (2021). Diversity in economics seminars: who gives invited talks? In *AEA Papers and Proceedings*, volume 111, pages 55–59. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Dupas, P., Modestino, A. S., Niederle, M., Wolfers, J., et al. (2021). Gender and the dynamics of economics seminars. Technical report, National Bureau of Economic Research.

Eaton, A. A., Saunders, J. F., Jacobson, R. K., and West, K. (2020). How gender and race stereotypes impact the advancement of scholars in stem: Professors' biased evaluations of physics and biology post-doctoral candidates. *Sex Roles*, 82:127–141.

Ellison, G. (2013). How does the market use citation data? the hirsch index in economics. *American Economic Journal: Applied Economics*, 5(3):63–90.

Feenberg, D., Ganguli, I., Gaule, P., and Gruber, J. (2017). It's good to be first: Order bias in reading and citing nber working papers. *Review of Economics and Statistics*, 99(1):32–39.

Fiala, D. and Tutoky, G. (2017). Pagerank-based prediction of award-winning researchers and the impact of citations. *Journal of Informetrics*, 11(4):1044–1068.

Fong, E. A. and Wilhite, A. W. (2017). Authorship and citation manipulation in academic research. *PloS one*, 12(12):e0187394.

Foster, L., McEntarfer, E., and Sandler, D. H. (2022). *Diversity and labor market outcomes in the economics profession*. US Census Bureau, Center for Economic Studies.

Fourcade, M., Ollion, E., and Algan, Y. (2015). The superiority of economists. *Journal of Economic Perspectives*, 29(1):89–114.

Francis, D. V., Hardy, B. L., and Jones, D. (2022). Black economists on race and policy: Contributions to education, poverty and mobility, and public finance. *Journal of Economic Literature*, 60(2):454–493.

Furman, J. L. and Stern, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, 101(5):1933–1963.

Ginther, D. K. and Kahn, S. (2004). Women in economics: moving up or falling off the academic career ladder? *Journal of Economic perspectives*, 18(3):193–214.

Ginther, D. K., Kahn, S., and Milakhina, D. (2025). Same as it ever was: Gender, race, and ethnicity differences in promotion for academic economists. In *AEA Papers and Proceedings*, volume 115, pages 195–201. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., and Kington, R. (2011). Race, ethnicity, and nih research awards. *Science*, 333(6045):1015–1019.

Greenspon, J. and Rodrik, D. (2021). A note on the global distribution of authorship in economics journals. Technical report, National Bureau of Economic Research.

Hadavand, A., Hamermesh, D. S., and Wilson, W. W. (2024). Publishing economics: How slow? why slow? is slow productive? how to fix slow? *Journal of Economic Literature*, 62(1):269–293.

Hager, S., Schwarz, C., and Waldinger, F. (2023). Measuring science: Performance metrics and the allocation of talent. *Available at SSRN 4472441*.

Hamermesh, D. S. (2018). Citations in economics: Measurement, uses, and impacts. *Journal of Economic Literature*, 56(1):115–156.

Hamermesh, D. S., Johnson, G. E., and Weisbrod, B. A. (1982). Scholarship, citations and salaries:

Economic rewards in economics. *Southern economic journal*, pages 472–481.

Hengel, E. (2022). Publishing while female: Are women held to higher standards? evidence from peer review. *The Economic Journal*, 132(648):2951–2991.

Hilmer, M. J., Ransom, M. R., and Hilmer, C. E. (2015). Fame and the fortune of academic economists: How the market rewards influential research in economics. *Southern Economic Journal*, 82(2):430–452.

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., and McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, 117(17):9284–9291.

Huang, W. (2015). Do abcs get more citations than xyzs? *Economic Inquiry*, 53(1):773–789.

Koffi, M. (2021a). Gendered citations at top economic journals. *AEA Papers and Proceedings*, 111:60–64.

Koffi, M. (2021b). Innovative ideas and gender inequality. Technical report, Working Paper Series.

Koffi, M. (2025). Innovative ideas and gender (in) equality. *American Economic Review*, 115(7):2207–2236.

Koffi, M., Pongou, R., and Wantchekon, L. (2024). Racial inequality and publication in economics. *AEA Papers and Proceedings*, 114:300–304.

Koffi, M. and Wantchekon, L. (2024). Racial justice from within? diversity and inclusion in economics. *Monograph of Econometric society*.

Kozlowski, D., Larivière, V., Sugimoto, C. R., and Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences*, 119(2):e2113067119.

Lutmar, C. and Reingewertz, Y. (2021). Academic in-group bias in the top five economics journals. *Scientometrics*, 126(12):9543–9556.

Mason, P., Myers, S. L., Simms, M., Lai, Y., and Liu, X. (2023). Research note: Citation bias and the economics of race and crime literature. *The Review of Black Political Economy*, page 00346446231212717.

Mason, P. L., Myers Jr, S. L., and Simms, M. (2022). Racial isolation and marginalization of economic research on race and crime. *Journal of Economic Literature*, 60(2):494–526.

McCabe, M. J. and Snyder, C. M. (2015). Does online availability increase citations? theory and evidence from a panel of economics and business journals. *Review of Economics and Statistics*, 97(1):144–165.

Mughogho, W., Adhiambo, J., and Forscher, P. S. (2023). African researchers must be full participants in behavioural science research. *Nature Human Behaviour*, 7(3):297–299.

Murray, F., Aghion, P., Dewatripont, M., Kolev, J., and Stern, S. (2016). Of mice and academics: Examining the effect of openness on innovation. *American Economic Journal: Economic Policy*, 8(1):212–252.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Page, S. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition*. Princeton University Press.

Pieters, R. and Baumgartner, H. (2002). Who talks to whom? intra-and interdisciplinary communication of economics journals. *Journal of Economic Literature*, 40(2):483–509.

Qiu, S., Steinwender, C., and Azoulay, P. (2022). Who stands on the shoulders of chinese (scientific) giants? evidence from chemistry. Technical report, National Bureau of Economic Research.

Reimers, N. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint*

arXiv:1908.10084.

Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review*, 107(5):141–145.

Sood, G. and Laohaprapanon, S. (2018). Predicting race and ethnicity from the sequence of characters in a name. *arXiv preprint arXiv:1805.02109*.

Staudt, J. (2020). Mandating access: assessing the nih's public access policy. *Economic policy*, 35(102):269–304.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, 225(2):175–199.

Taffe, M. A. and Gilpin, N. W. (2021). Racial inequity in grant funding from the us national institutes of health. *Elife*, 10:e65697.

Thalmayer, A. G., Toscanelli, C., and Arnett, J. J. (2021). The neglected 95% revisited: Is american psychology becoming less american? *American Psychologist*, 76(1):116.

Weber, M. (2018). The effects of listing authors in alphabetical order: a review of the empirical evidence. *Research evaluation*, 27(3):238–245.

Wessel, D., Barcena, L., and Salwati, N. (2020). Gender and racial diversity of federal government economists: 2020 data. *Report from the BROOKINGS INSTITUTION*.

Wilhite, A. W. and Fong, E. A. (2012). Coercive citation in academic publishing. *Science*, 335(6068):542–543.

Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy*, 121(1):1–27.

Wu, A. H. (2020). Gender bias among professionals: an identity-based interpretation. *Review of Economics and Statistics*, 102(5):867–880.

Zacchia, G. (2021). What does it take to be top women economists? an analysis using rankings in repec. *Review of Political Economy*, 33(2):170–193.

Tables

Table I: Summary table

	All White	At least one black/hispanic/asian	Other races	Total
N	178,411	83,242	67,578	329,231
Number of authors	1.700	2.433	2.192	1.986
Publications as of 2021 (top 200)	19.968	19.333	15.617	18.914
Publications as of 2021 (top 35)	7.368	6.442	4.549	6.555
Publications as of 2021 (top 16)	3.670	3.346	2.173	3.281
Publications as of 2021 (top 5)	1.624	0.995	0.815	1.299
Journal rank (%)				
200 journals	70.041	74.352	77.684	72.700
Top 35	16.030	14.661	12.955	15.053
Top 16	7.310	7.881	5.713	7.126
Top 5	6.619	3.107	3.648	5.121
Fields (%)				
Econometrics	10.685	11.835	9.478	10.734
Microeconomics	19.135	15.025	17.527	17.742
Macroeconomics	8.139	6.688	7.648	7.663
International Economics	6.355	7.205	6.832	6.673
Finance	13.250	19.267	15.389	15.245
Labour & Education	13.191	10.095	11.986	12.143
Other (empirical) fields	26.213	27.043	28.629	26.924
Miscellaneous	3.031	2.841	2.511	2.875
Style (%)				
Mostly theoretical	65.291	55.934	60.986	61.987
Mostly empirical	29.286	35.188	33.495	31.677
Mostly structural	5.423	8.877	5.519	6.336
Affliation categories (%)				
Low Tier	63.595	63.482	68.441	64.561
Mid Tier	22.107	23.147	20.550	22.050
Top Tier	14.298	13.371	11.010	13.388

This table shows the summary statistics for the set of papers in the database.

Table II: Citations and race: baseline estimation

	Outcome variable: Citation (asinh)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
At least one Black/Hispanic/Asian	-0.139***	-0.256***	-0.150***	-0.059***	-0.058***	-0.071***	-0.051***
	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.005)
Other races	-0.166***	-0.244***	-0.132***	-0.076***	-0.059***	-0.062***	-0.044***
	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)
Number of authors		0.159***	0.146***	0.228***	0.222***	0.191***	0.096***
		(0.005)	(0.005)	(0.008)	(0.008)	(0.007)	(0.004)
Prior publications at the time of publication			-0.378***	-0.141***	-0.165***	-0.150***	-0.120***
			(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Publications as of 2021			0.551***	0.266***	0.249***	0.250***	0.234***
			(0.003)	(0.004)	(0.004)	(0.003)	(0.003)
Observations	329231	329231	329231	329231	329231	329231	329231
Adj. R-sqr	0.002	0.011	0.091	0.262	0.282	0.303	0.526
Year fixed effects				Y	Y	Y	
Affiliation fixed effects					Y	Y	Y
Field fixed effects						Y	Y
Journal-Year fixed effects							Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table III: Citation and race: effect across journal types

	Outcome variable: Citation (asinh)			
	More than 200 journals	Top 35 journals (Card et al. [2020])	Top 16 journals (Koffi [2021b])	Top 5 journals
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.051***	-0.086***	-0.085***	-0.096***
	(0.005)	(0.010)	(0.015)	(0.028)
Other races	-0.044***	-0.050***	-0.055***	-0.072**
	(0.006)	(0.012)	(0.018)	(0.031)
Number of authors	0.096***	0.093***	0.124***	0.108***
	(0.004)	(0.006)	(0.009)	(0.015)
Prior publications at the time of publication	-0.120***	-0.123***	-0.167***	-0.211***
	(0.003)	(0.006)	(0.009)	(0.015)
Publications as of 2021	0.234***	0.291***	0.348***	0.415***
	(0.003)	(0.006)	(0.009)	(0.015)
Observations	329231	89881	40322	16860
Adj. R-sqr	0.526	0.455	0.485	0.413
Affiliation fixed effects	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Column (1) reports the results of Table II- Column (7). Column (2) restricts the sample to 35 high-impact journals identified by Card et al. [2020]. Column (3) limits the sample to 16 high-impact journals according to Koffi [2021b]. Column (4) restricts the sample to the top five economics journals: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table IV: Centrality measures

	Outcome variable: Standardized centrality measures					
	In-degree	Betweeness	Eigenvector	Katz	PageRank	Clustering
	(1)	(2)	(3)	(4)	(5)	(6)
At least one Black/Hispanic/Asian	-0.027***	-0.020***	-0.006***	-0.016***	-0.024***	0.041***
	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)
Other races	-0.025***	-0.013***	-0.005*	-0.013***	-0.020***	0.014***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)
Observations	329231	329231	329231	329231	329231	314706
Adj. R-sqr	0.174	0.111	0.034	0.084	0.128	0.105
Publications and team size controls	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y

This table shows the relationship between the centrality measures and the racial composition of the paper's authors, adding granular controls for the fields and style of methodology. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table V: Robustness: Adding granular field controls

	Outcome variable: Citation (asinh)							
	Baseline Papers with JEL codes	1-digit JEL code	2-digit JEL code	3-digit JEL code	768 Embeddings	Style	3-digit JEL code + Style	3-digit JEL code + Style+ Embeddings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	-0.049***	-0.070***	-0.070***	-0.065***	-0.069***	-0.059***	-0.063***	-0.064***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Other races	-0.035***	-0.041***	-0.040***	-0.038***	-0.048***	-0.043***	-0.036***	-0.037***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.005)	(0.006)	(0.007)	(0.006)
Probability of theoretical content						-0.414***	-0.261***	0.028
						(0.019)	(0.026)	(0.051)
Probability of empirical content						0.046**	0.201***	0.054
						(0.019)	(0.026)	(0.059)
Observations	208045	208045	208045	208045	317649	317649	207118	207118
R-sqr	0.538	0.540	0.545	0.553	0.566	0.536	0.559	0.591
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, adding granular controls for the fields and style of methodology. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The sample size in this table differs from that in Table II due to variations in data availability: not all papers have their JEL codes available (Columns 1 to 4), not all papers have their abstracts available (Columns 5 and 6), and not all papers have both JEL codes and abstracts (Columns 7 and 8). Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VI: Citation and race: control for past citations

	Outcome variable: Citation (asinh)			
	More than 200 journals	Top 35 journals (Card et al. [2020])	Top 16 journals (Koffi [2021b])	Top 5 journals
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.043***	-0.068***	-0.060***	-0.083***
	(0.006)	(0.010)	(0.015)	(0.028)
Other races	-0.037***	-0.039***	-0.039**	-0.058*
	(0.006)	(0.012)	(0.018)	(0.031)
Past citations (1)	0.125***	0.098***	0.104***	0.067
	(0.010)	(0.021)	(0.034)	(0.052)
Past citations (1-3.6)	0.197***	0.170***	0.151***	0.124***
	(0.010)	(0.020)	(0.031)	(0.048)
Past citations (3.6-9)	0.265***	0.240***	0.228***	0.166***
	(0.010)	(0.020)	(0.029)	(0.046)
Past citations (9-20)	0.348***	0.348***	0.341***	0.291***
	(0.012)	(0.021)	(0.030)	(0.047)
Past citations (20-47)	0.412***	0.427***	0.435***	0.313***
	(0.013)	(0.023)	(0.032)	(0.050)
Past citations (47-128)	0.508***	0.521***	0.566***	0.449***
	(0.015)	(0.025)	(0.034)	(0.054)
Past citations (>128)	0.655***	0.659***	0.705***	0.503***
	(0.017)	(0.030)	(0.041)	(0.063)
Constant	2.606***	2.797***	2.897***	3.513***
	(0.042)	(0.084)	(0.101)	(0.159)
Observations	329231	89881	40322	16860
Adj. R-sqr	0.529	0.458	0.489	0.416
Affiliation fixed effects	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors while adding controls for authors' past citations. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Column (1) reports the results of Table II- Column (7). Column (2) restricts the sample to 35 high-impact journals identified by Card et al. [2020]. Column (3) limits the sample to 16 high-impact journals according to Koffi [2021b]. Column (4) restricts the sample to the top five economics journals: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. Standard errors are reported in parentheses. The "past citations" variable is a categorical variable based on the decile of the average team citations distribution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VII: Citations and the race of the authors of the citing paper

	Outcome variable: Citation (asinh)							
	> 200 journals		Top 35 journals		Top 16 journals		Top 5 journals	
	From non-White	From White	From non-White	From White	From non-White	From White	From non-White	From White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	0.232***	-0.303***	0.211***	-0.338***	0.212***	-0.323***	0.215***	-0.279***
	(0.004)	(0.004)	(0.010)	(0.010)	(0.015)	(0.015)	(0.027)	(0.028)
Other races	0.012***	-0.180***	0.008	-0.185***	0.014	-0.165***	0.029	-0.144***
	(0.004)	(0.005)	(0.010)	(0.011)	(0.017)	(0.017)	(0.028)	(0.029)
Observations	329231	329231	89881	89881	40322	40322	16860	16860
R-sqr	0.419	0.460	0.420	0.440	0.427	0.413	0.428	0.367

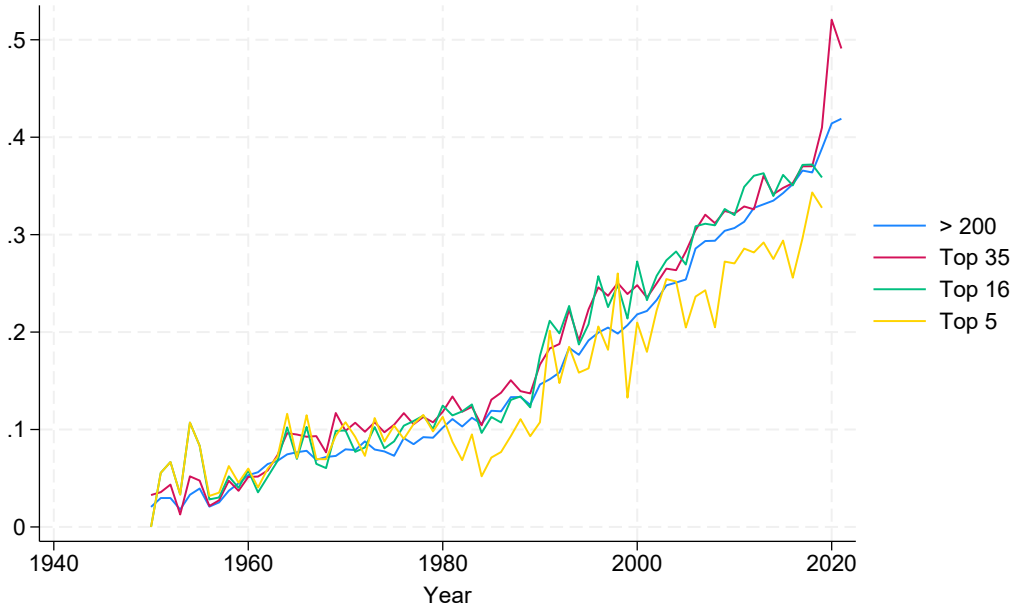
This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, emphasizing the race of the authors of the citing paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Columns (1), (3), (5), and (7) restrict the citation count to citations from only teams including non-White scholars. Columns (2), (4), (6), and (8) restrict the citation count to citations from only teams, including all-White scholars. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table VIII: Effect of seminars and conference attendance

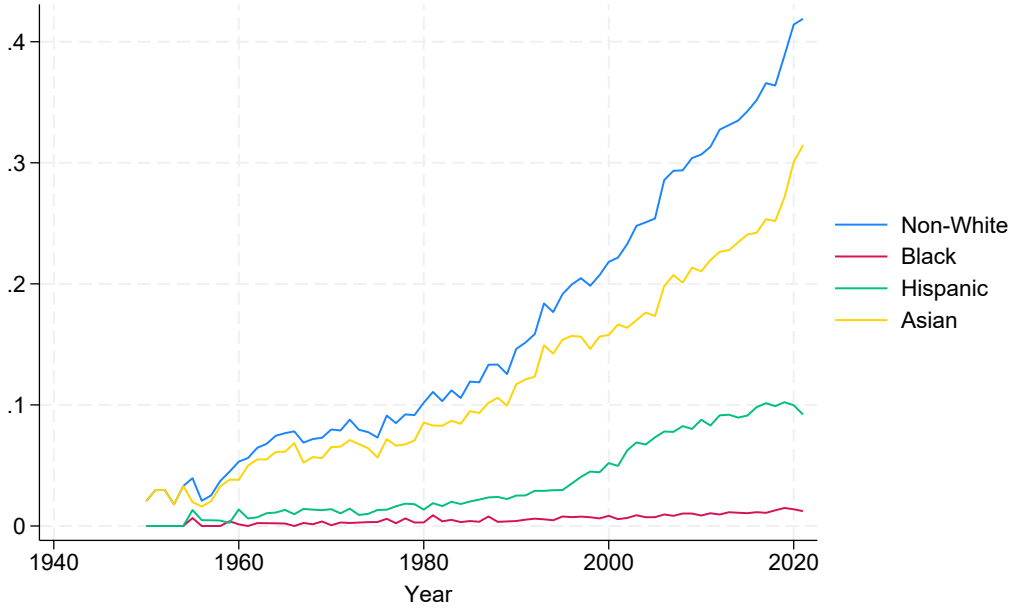
	Outcome variable: Citation (asinh)							
	> 200 journals		Top 35 journals		Top 16 journals		Top 5 journals	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	-0.022	-0.020	-0.087***	-0.083***	-0.063*	-0.057	-0.100	-0.094
	(0.015)	(0.015)	(0.025)	(0.025)	(0.038)	(0.037)	(0.067)	(0.067)
Other races	-0.037**	-0.037**	-0.026	-0.027	0.017	0.013	0.009	0.008
	(0.017)	(0.017)	(0.028)	(0.027)	(0.044)	(0.043)	(0.072)	(0.071)
Seminars/conferences count		0.010***		0.011***		0.009***		-0.002
		(0.002)		(0.002)		(0.003)		(0.005)
Prominent conferences		0.015		0.006		0.046		-0.010
		(0.034)		(0.039)		(0.052)		(0.094)
Seminars in top-tier affiliations		0.024***		0.015		0.003		0.034*
		(0.009)		(0.010)		(0.013)		(0.019)
Observations	22352	22352	8170	8170	3538	3538	1192	1192
Adj. R-sqr	0.648	0.649	0.635	0.637	0.629	0.631	0.656	0.656
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, emphasizing the effect of seminars and conference attendance. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Uneven columns represent the regression model without control variables. Even columns represent the regression model with the full set of control variables. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures



(a) Across journals



(b) Within >200 journals

Figure I:
Share of publications with at least Black, Hispanic, or Asian over time

The figure presents evolution over time of the share of papers with at least one Black, Hispanic, or Asian for different sets of journals (whole sample with more than 200 journals, a set of 35 high-impact journals as per Card et al. [2020], for a set of top 16 journals used in Koffi [2021b], and the traditional top 5 high impact journals.

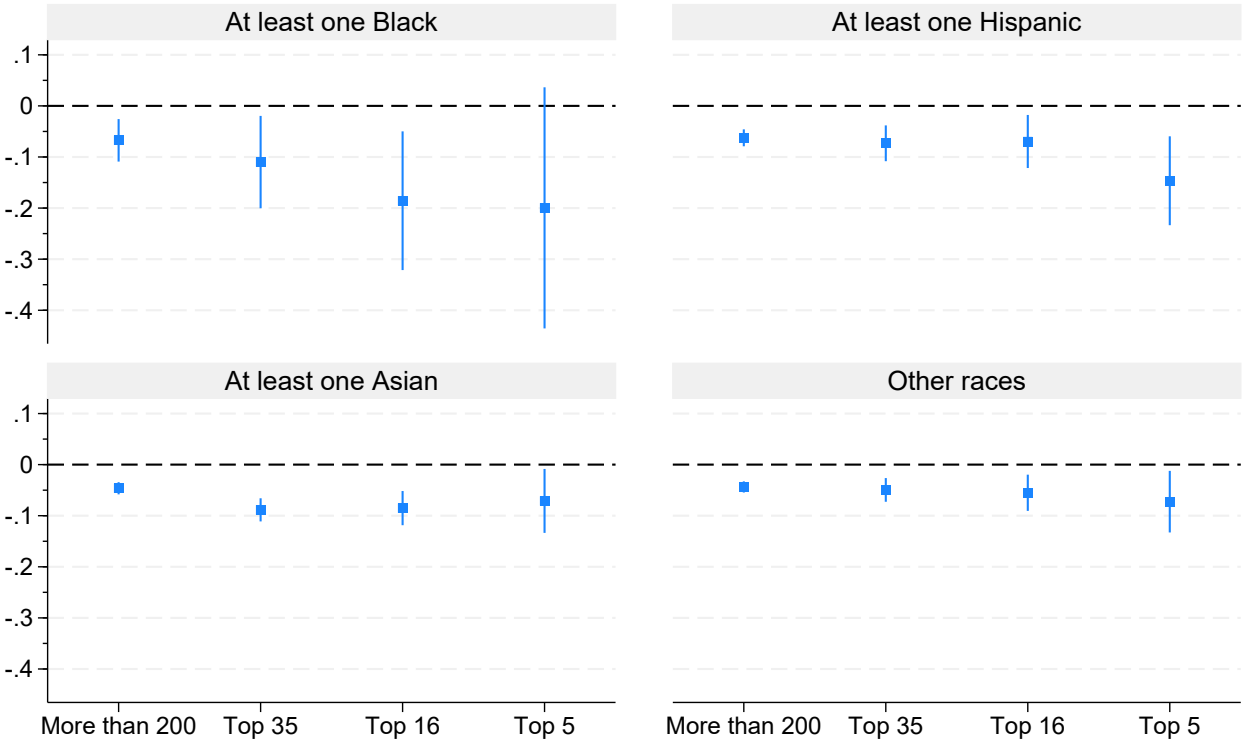


Figure II:
Citation and race

The figure presents the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The figure shows the estimation results by specific racial groups and journal rank.

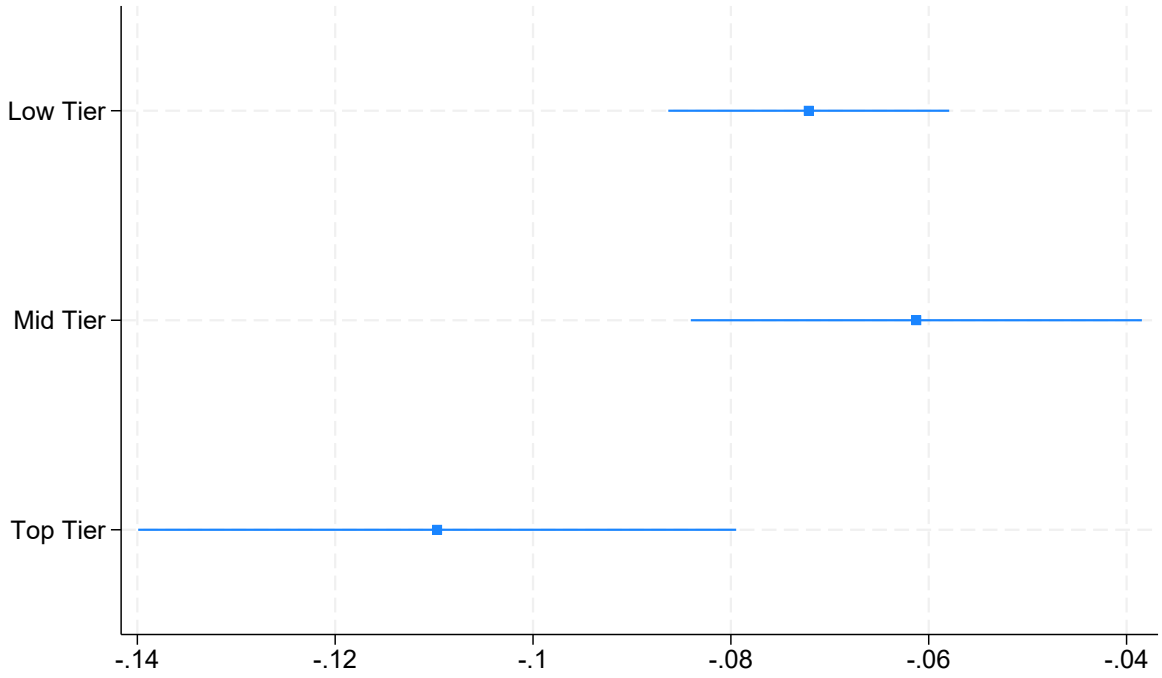


Figure III:
Citation and race: heterogeneity by affiliations

The figure presents the relationship between citation and race, focusing on the heterogeneous results depending on the authors' affiliations rankings.

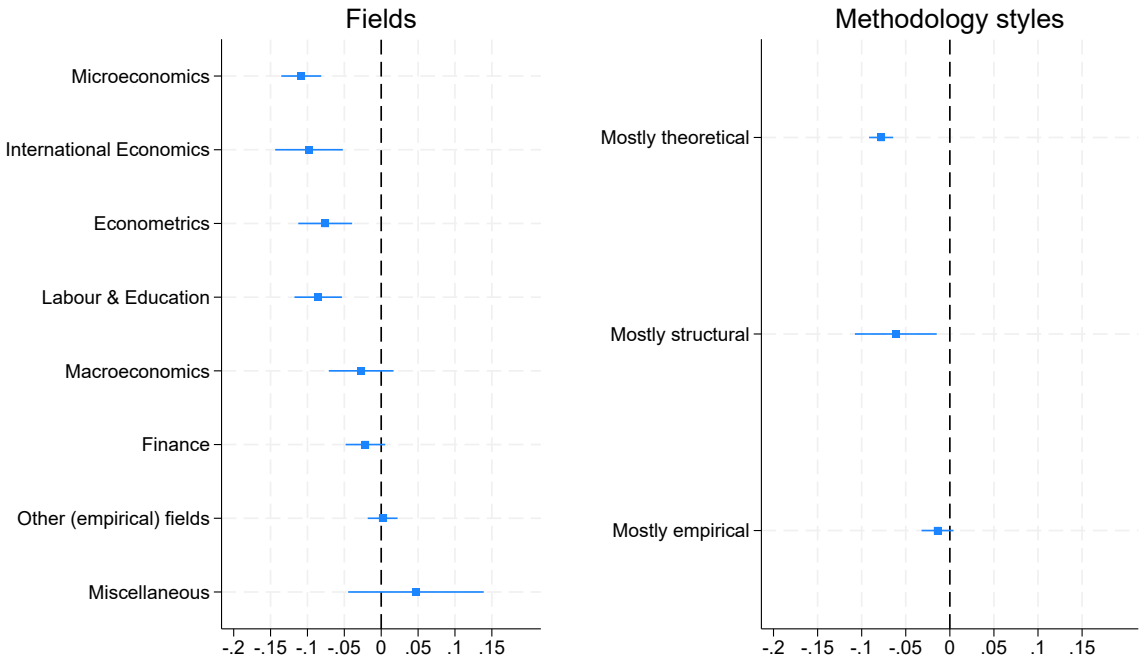


Figure IV:
Citation and race: heterogeneity by fields and methodology styles

The figure presents the relationship between citation and race, focusing on the heterogeneous results depending on the primary field and the methodology style of the paper.

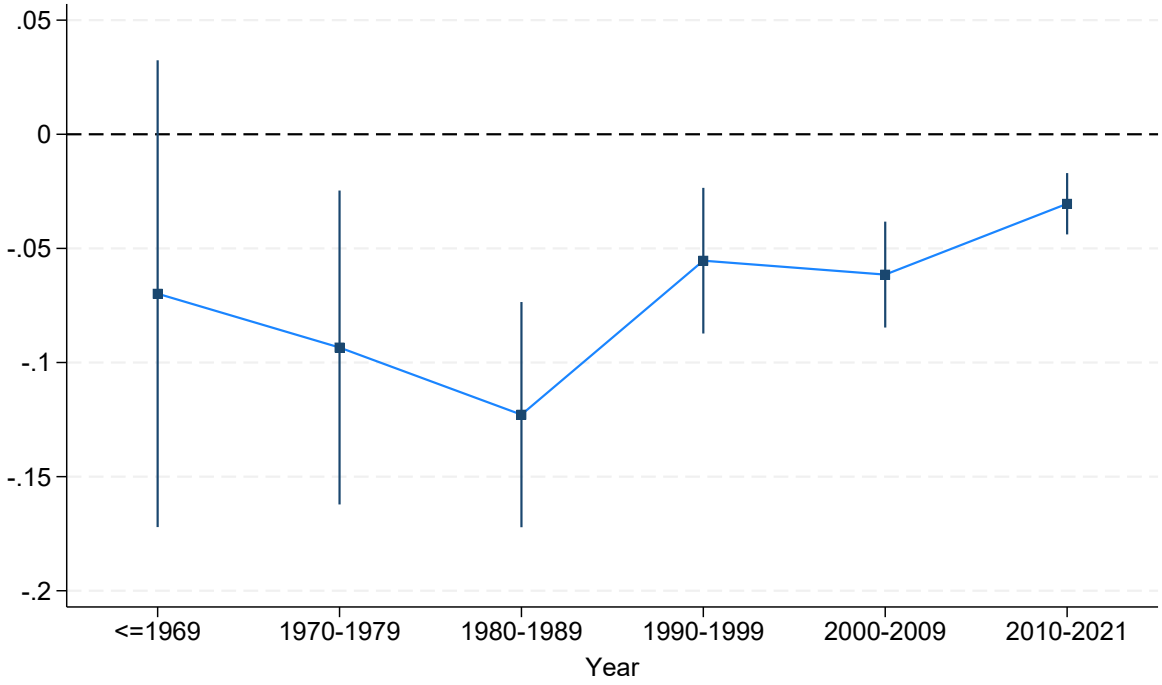
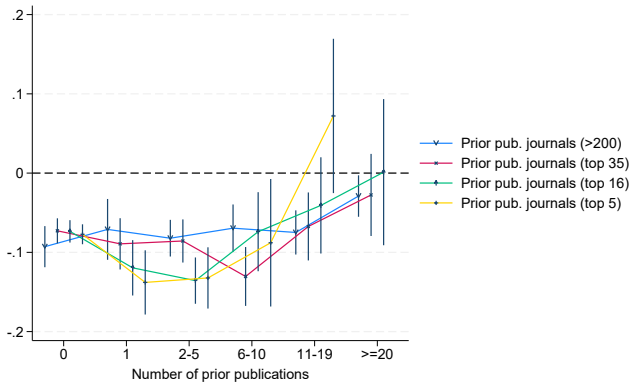
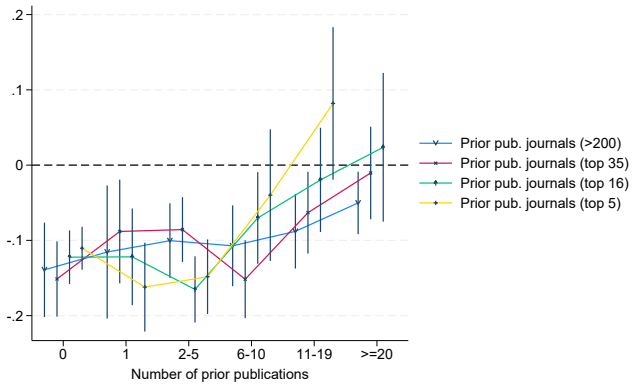


Figure V:
Citation and race: Evolution over time

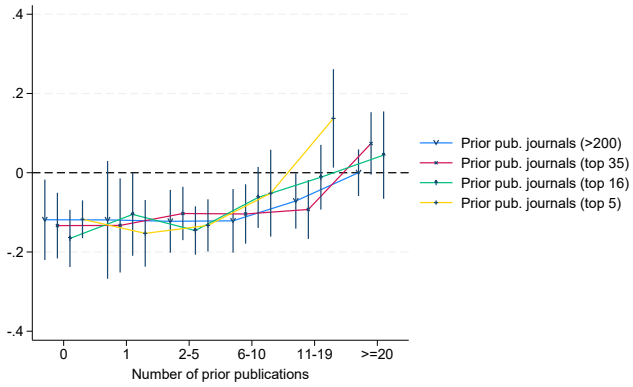
The figure presents the relationship between citation and race, focusing on the evolution over time. The regression coefficient on "At least one Black/Hispanic/Asian" is estimated for each period shown on the x-axis.



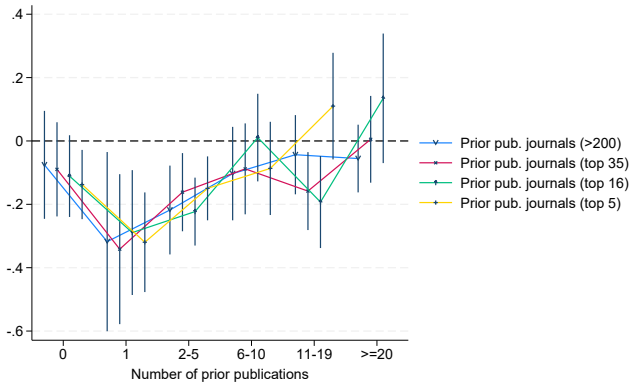
(a) Gap in more than 200 journals



(b) Gap in top 35 journals



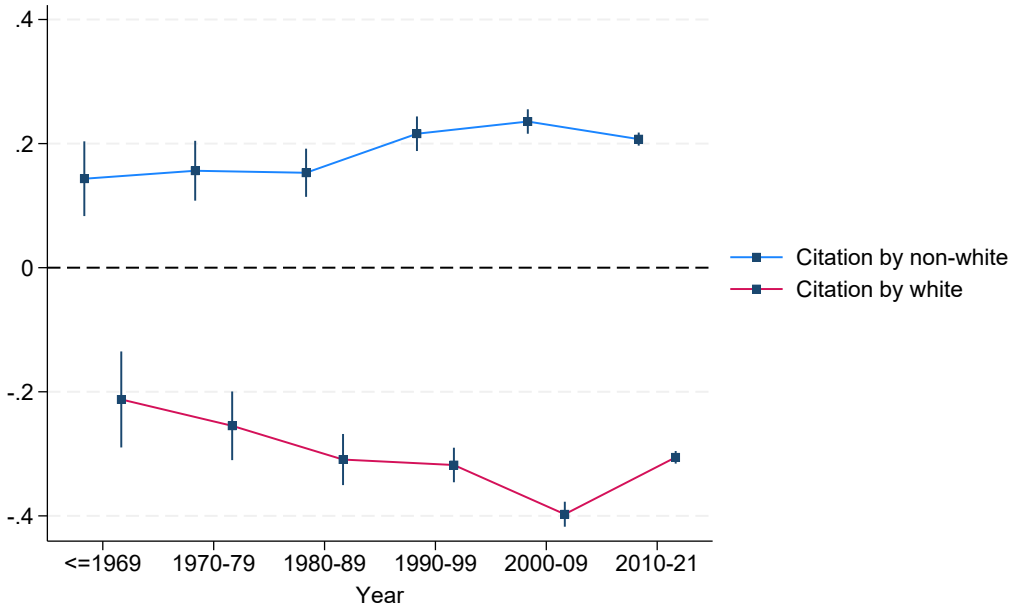
(c) Gap in top 16 journals



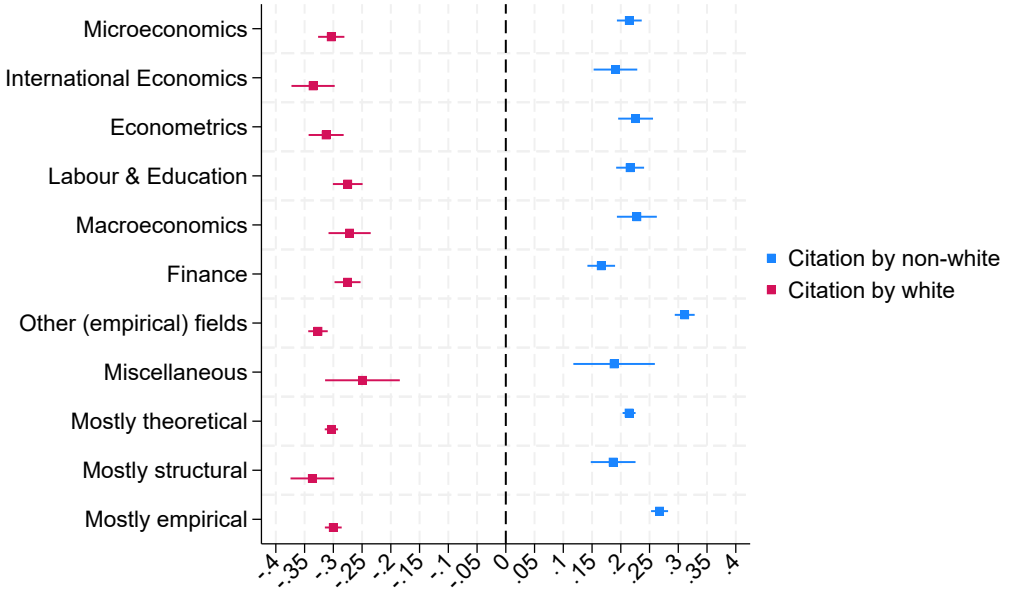
(d) Gap in top 5 journals

Figure VI:
Citation, race and prior publications

The figure presents the relationship between the citation (inverse hyperbolic sine—asinh), the outcome variable, and the racial composition of the paper's authors, the independent variable of interest, focusing on the effect of an additional publication. The number of prior publications is measured at the time of publication of the focal paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The plots display the estimate for teams with at least one Black/Hispanic/Asian author by comparison to a team with all-White authors.



(a) Non-White/All-White citation penalty over premium over time



(b) Non-White/All-White citation penalty across fields and methodology

Figure VII:
Citation and race of the authors in the citing paper

The figure presents the relationship between citation and race, focusing on the racial composition of the authors of the citing paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The plots display the estimate for teams with at least one Black/Hispanic/Asian author by comparison to a team with all-White authors.

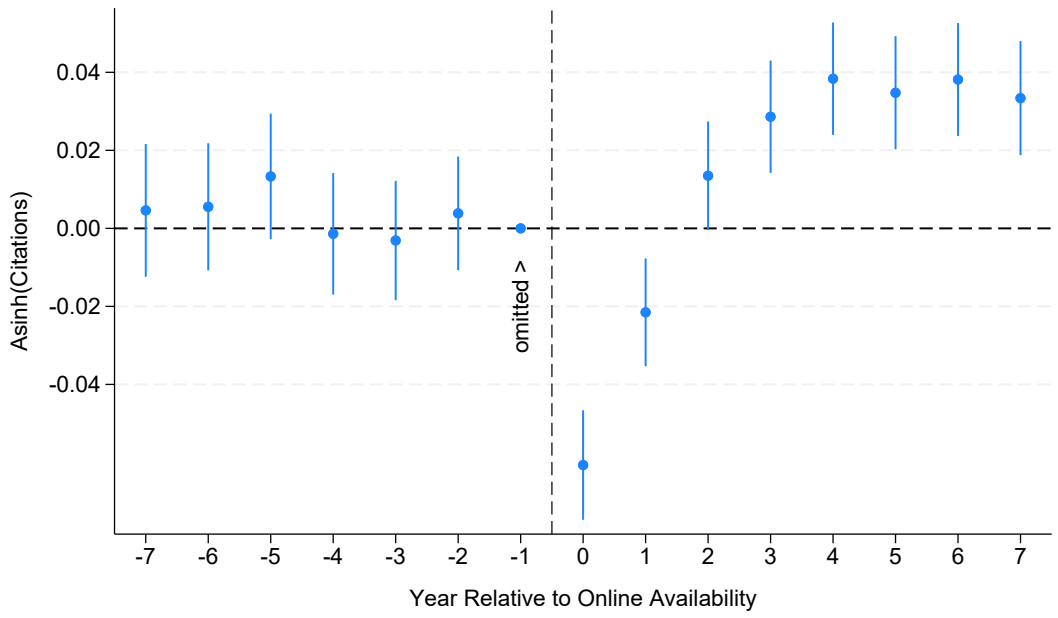


Figure VIII:
Effect of online availability on the racial citation gap

The figure presents the event-study coefficients from estimating Equation 2, which assesses the impact of online availability on the racial citation gap. The coefficient at event time $\tau = -1$, the citation year immediately preceding online availability, is normalized to zero and serves as the reference period. The specification controls for the author's time-varying publication history, includes flexible controls for paper age, modeled using age, age squared, age cubed, and age to the fourth power, as well as paper fixed effects, citation year fixed effects, and journal times publication year fixed effects. Standard errors clustered at the paper level.

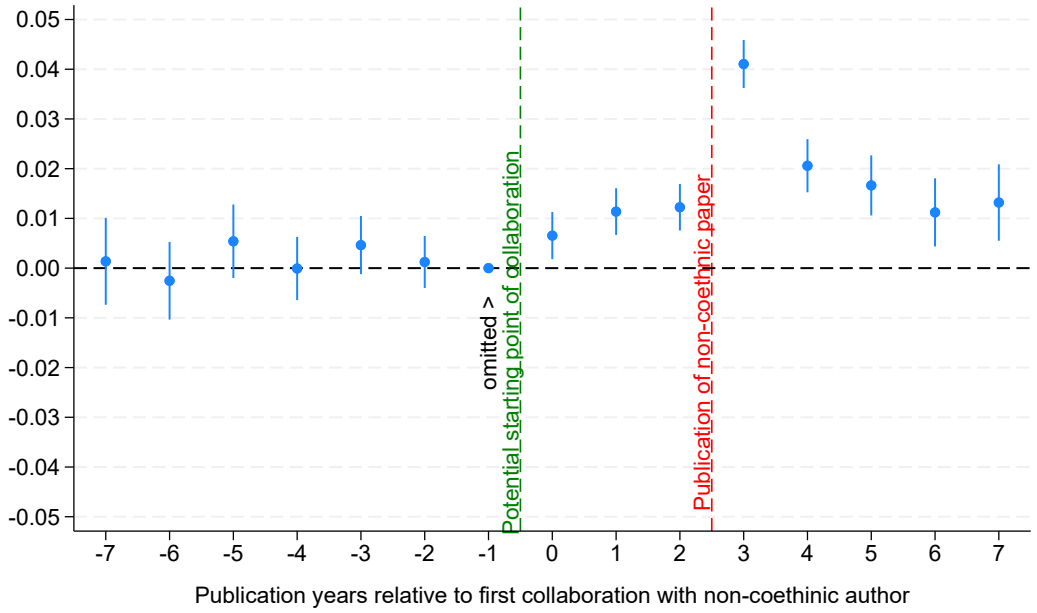


Figure IX:
Effect of collaboration with a non-coethnic author

The figure presents the event study showing the impact of the collaboration with a non-coethnic author on the share of listed bibliographical references to papers with authors of the out-group. To estimate this effect, we consider an author-level estimation and analyze all subsequent publications with only coethnic authors following the first collaboration with a non-coethnic author. We truncate the sample to the next non-coethnic collaboration.

Appendix For Online Publication
The Color of Ideas: Race, Networks, and the Diffusion of Knowledge

Marlène Koffi & Roland Pongou & Leonard Wantchekon

A Expanding on key concepts

A.1 Name disambiguation

A sequential process then parsed each name into first, middle, and last names, with thorough checks at each step to minimize errors.

When complete first, middle, and last names are available, they are considered to belong to the same author. Subsequently, if both first and last names are present, along with an initial middle name, and only one candidate is identified, these names are merged. Following this, cases in which names comprise only the initials of the first and middle names are examined. Finally, cases in which middle names are absent are addressed. In such instances, the frequency of the first and last name in the database is analyzed, considering the relative proportion of the first name in the database and utilizing surname classifications from the US census data. For instance, a name like "Smith," which is very popular, is excluded from the process, and disambiguation was manually checked whenever possible.

Chinese names were systematically excluded during this process, and the disambiguation of those names occurred primarily through MAG disambiguation. While disambiguation procedures are inherently imperfect, this multi-staged process enables the creation of a comprehensive and meaningful publication history for the authors.

A.2 Centrality measures

We compute measures of centrality to assess the importance of papers. Centrality measures account for different notions of importance in a (citation) network; for example, we might only be interested in counting the number of citations received by a paper, or we might also be interested in how citing papers are cited by other papers. We consider six well-known notions of centrality measures: degree centrality, betweenness centrality, eigenvector centrality, Katz centrality, Pagerank centrality, and clustering coefficient.

To compute the values of these centrality measures, we assume that all papers form a network of citations, where a paper i is linked to another paper j if i has received a citation from j. Clearly, this is a directed network as a paper j may cite i without the converse being true. Assume there are n published papers. The citation network among these papers can be represented by an adjacent matrix $A=(a_{ij})$, where for any nodes i and j, the matrix entry or cell $a_{ij}=1$ if i has received a citation from j and 0 if not. Below, we formally define the six centrality measures we use.

- The degree centrality of a paper i, denoted by χ_i, is the total number of citations that i has received from other papers: $\chi_i = \sum_{j=1}^{n} a_{ij}$.

- The betweenness centrality of a paper i, denoted by b_i, is the fraction of the shortest paths between two other papers passing through i: $b_i = \sum_{j,k} \frac{\sigma^i_{jk}}{\sigma_{jk}}$, where σ_{jk} is the total number of shortest paths from agents j to k (with i different from both j and k), and σ^i_{jk} is the total number of shortest paths that pass through paper i.
- The eigenvector centrality is a measure of prestige that reflects the notion that being cited by other highly cited papers is important. The eigenvector centrality of a paper i, denoted by ν_i, measures the extent to which i is cited by other highly cited papers. It is given by the formula $\nu_i = \frac{1}{e} \sum_{j=1}^{n} a_{ij}\nu_j$, where e is the largest eigenvalue and ν (a column vector with n entries) the corresponding eigenvector of the adjacency matrix A: $A\nu = e\nu$.
- The Katz centrality, similar to the eigenvector centrality, values citations from papers that are also cited; however, having distant neighbors penalize a node's score by some factor α. The Katz centrality of a paper i, denoted by K_i, is defined as $K_i = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)_{ji}$, where $(A^k)_{ji}$ is the number of k degree connections between papers i and j and α is an attenuation factor (strictly between 0 and 1) chosen to be smaller than the inverse of the largest eigenvalue of the adjacency matrix A. The fact α^k is a decreasing function of k reflects the fact that the Katz centrality weights a link to a more distant paper less.
- The Pagerank centrality is also similar to the eigenvector centrality. The PageRank centrality of a paper i, denoted by PR_i, is recursively defined as: $PR_i = \alpha \sum_j a_{ji} \frac{PR_j}{\chi_j} + \frac{1-\alpha}{n}$, where χ_j is the number of citations received by j, and α is a dampening factor that is positive and strictly lower than 1. item The clustering coefficient of a paper i, denoted by C_i, is a measure of the extent to which papers that cite paper i tend to cite each other, leading to a tightly knit group of papers. Let $N_i = \{j : a_{ij} = 1\}$ be the set of papers that cite paper i, and let $|N_i|$ be the cardinality of this set. The clustering coefficient of paper i is given by: $C_i = \frac{|\{a_{jk}:a_{jk}=1, j,k \in N_i\}|}{|N_i||N_i-1|}$.

To process our data from a panel data into a network structure, we first create a table of directed links so that every row observation is some article j citing some article i, and then convert this to a *directed* network object using the python library NetworkX; this is then converted to a Networkit object. Because our data is in a node-to-node panel format, a first transformation to a NetworkX graph object is necessary prior to the conversion to a Networkit directed graph. Then, using the generated directed network object, we calculate the above centrality measures to be accurate at the 6 significant figure level using built-in functionality with the NetworkX library. In particular, we were concerned with processing times for loading the dataset, shortest path calculations, PageRank calculations as they most relate to our work, and built-in capabilities to multi-thread the process.

A.3 Natural language processing models

This appendix describes the natural language processing (NLP) methods used to do the paper vector representation as well as assess thematic similarity and complementarity across papers authored by scholars from different racial groups.

A.3.1 TF-IDF models

The Term Frequency-Inverse Document Frequency (TF-IDF) score is computed as the product of two components: term frequency (TF), which measures how frequently a term appears in a document,

and inverse document frequency (IDF), which quantifies the term's specificity by penalizing terms that appear in many documents. Mathematically, for a term t in document d within a corpus D, the TF-IDF score is given by:

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \cdot \text{IDF}(t, D), \quad \text{where} \quad \text{IDF}(t, D) = \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right)$$

Each paper is represented by a vector of TF-IDF scores, computed using n-grams (composition of words) extracted from preprocessed (cleaned) abstracts (remove stopwords, punctuation, and numerical characters, normalize the text).

A.3.2 Contextual embedding models

To better capture semantic meaning beyond word frequencies, we employ contextual embedding models based on transformer architectures. Specifically, we use pre-trained models such as Sentence-BERT (Sentence-Bidirectional Encoder Representations from Transformers, Reimers [2019]), which generates dense vector representations for each abstract by encoding both syntax and context. These models are pre-trained on large corpora using tasks such as masked language modeling or next-token prediction. SBERT modifies the BERT architecture by processing input sentences through a pre-trained BERT model, which consists of multiple transformer layers with self-attention mechanisms.

B A simple model: Homophily and the racial citation gap

B.1 Preliminaries

In this section, we solve the author's maximization problem stated in section 4.4.1. The model has clear implications for the racial citation gap and shows how this outcome is driven by homophily in citation patterns. The model also sheds light on some of the heterogeneous effects of race on citation found in the empirical analysis.

The author's maximization problem stated in section 4.4.1 is:

$$\max_{(s_1, s_2)} \quad u(s_1, s_2) = v(s_1, s_2) - c(s_1, s_2)$$

where:

$$v(s_1, s_2) = \frac{n_1}{n_1 + n_2}\alpha[aln(s_1) + bln(s_2)] + \frac{n_2}{n_1 + n_2}\alpha[bln(s_1) + aln(s_2)]$$

We recall that the total cost function is defined as $c(s_1, s_2) = h(q, r)g(s_1, s_2)$, where $g(s_1, s_2)$ is an increasing function of the number of citations to papers authored by scholars from the minority group (s_1) and those from the majority group (s_2), and $h(q, r) > 0$ is a function that is increasing in q and decreasing in r. As with producer theory's input cost, we assume that the function $g(s_1, s_2)$ is a linear function of the input factors s_1 and s_2: $g(s_1, s_2) = c_1 s_1 + c_2 s_2$. This implies that the total cost function $c(s_1, s_2) = h(q, r)g(s_1, s_2)$ is linear as well. Fixing $h(q, r)$, the parameter c_1 (resp. c_2) is therefore the marginal cost of exploiting and citing a paper authored by a scholar from the minority group (resp. majority group). The function $h(q, r)$ measures how hard it is to publish in a journal of quality q, given an author's ability r. Then, the multiplicative expression in the cost function is a reasonable assumption, as an author aiming for a top-tier journal will typically invest more time in carefully reading and citing a paper to ensure accuracy, compared to when targeting a lower-ranked journal. This is because a more precise citation enhances the quality of the paper. Additionally, a higher-ability author will exert less effort in utilizing a paper for citation than a lower-ability author. Therefore, the marginal cost of writing or utilizing a paper increases when the target journal is of higher quality, and it decreases for authors with greater ability.[23]

In determining s_1 and s_2, an important underlying hypothesis is also that missing an important referee can be costly, either because referees are typically key figures in the relevant literature, and overlooking a significant paper could diminish the perceived contribution, or due to potential editorial bias (Wilhite and Fong [2012], Fong and Wilhite [2017]).

While we solve the maximization problem stated above without accounting for author identity, when analyzing how citation behavior differs by identity, we assume that $c_1 = c_1^1$ and $c_2 = c_2^1$ for an author from the **minority group**, and $c_1 = c_1^2$ and $c_2 = c_2^2$ for an author from the **majority group**, where $c_i^{i'}$ represents the marginal cost for authors in group i' to cite authors from group i, with $i, i' = 1, 2$.

The problem stated above is equivalent to:

$$\max_{(s_1, s_2)} \quad \alpha\tfrac{n_1 a + n_2 b}{n} ln(s_1) + \alpha\tfrac{n_1 b + n_2 a}{n} ln(s_2) - h(q, r)(c_1 s_1 + c_2 s_2)$$

For simplicity, we assume that the numbers of cited papers from minority and majority authors, s_1 and s_2, are positive reals. Let $k_1 = n_1 a + n_2 b$ and $k_2 = n_1 b + n_2 a$. The fact that $a \geq b$ means that it could be easier for a referee from the minority (resp. majority) group to recognize the value of citations to papers authored by minority (resp. majority) scholars.

[23] A natural candidate for the function $h(q, r)$ is $\frac{q}{r}$, though other functions could also be considered.

The first-order conditions (F.O.Cs) for interior solutions are:

$$\partial s_1 : \frac{\alpha k_1}{n s_1} = h(q,r)c_1 \tag{4}$$

$$\partial s_2 : \frac{\alpha k_2}{n s_2} = h(q,r)c_2 \tag{5}$$

From Equation 4 and Equation 5, we deduce that:

$$s_2^* = \frac{k_2 c_1}{k_1 c_2} s_1^* \tag{6}$$

It follows that

$$s_1^* = \frac{\alpha k_1}{h(q,r) n c_1} \tag{7}$$

$$s_2^* = \frac{\alpha k_2}{h(q,r) n c_2} \tag{8}$$

$$s_1^* - s_2^* = \left[\frac{k_1 c_2 - k_2 c_1}{k_1 c_2}\right] s_1^* = \frac{\alpha(k_1 c_2 - k_2 c_1)}{h(q,r) n c_1 c_2} \tag{9}$$

These equations will allow us to study the racial citation and homophily gaps in the next sections.

B.2 The racial citation gap

In this section, we analyze the racial citation gap. From Equations 7 and 8, we have the following:

$$s_1^* - s_2^* = \left[\frac{k_1 c_2 - k_2 c_1}{k_1 c_2}\right] s_1^* = \frac{\alpha(k_1 c_2 - k_2 c_1)}{h(q,r) n c_1 c_2} \tag{10}$$

Recall that $c_1 = c_1^1$ and $c_2 = c_2^1$ for an author from the minority group; similarly, $c_1 = c_1^2$ and $c_2 = c_2^2$ for an author from the majority group. We maintain the assumption that the marginal cost of citing a paper written by a coethnic author is lower than that of citing a paper by a non-coethnic author: $c_i^i < c_{i'}^i$ for $i \neq i'$.

For a **minority author**, Equation 10, which gives the difference in the number of citations to papers authored by minority and majority scholars, becomes:

$$s_1^{1*} - s_2^{1*} = \frac{\alpha(k_1 c_2^1 - k_2 c_1^1)}{h(q,r) n c_1^1 c_2^1} \tag{11}$$

This quantity could be positive, meaning that the number of citations to papers authored by minority scholars exceeds the number of citations to papers authored by scholars from the majority group, or negative, depending on the relative size of the minority group and the relative cost of citing a coethnic author versus a non-coethnic author.

For a **majority author**, Equation 10 becomes:

$$s_1^{2*} - s_2^{2*} = \frac{\alpha(k_1 c_2^2 - k_2 c_1^2)}{h(q,r) n c_1^2 c_2^2} \tag{12}$$

This quantity is clearly negative because it can be shown that the numerator is negative.

Are papers authored by minority authors less frequently cited than papers of similar quality authored by majority authors? If one assumes that the pool of referees is identical to the pool of authors, which is a reasonable assumption in a peer-review system, then the **average** difference in the number of citations to papers authored by minority scholars and papers authored by majority scholars is:

$$E(s_1^* - s_2^*) = \frac{n_1}{n}[s_1^{1*} - s_2^{1*}] + \frac{n_2}{n}[s_1^{2*} - s_2^{2*}] \tag{13}$$

Equivalently,

$$E(s_1^* - s_2^*) = \frac{n_1}{n}[\frac{\alpha(k_1 c_2^1 - k_2 c_1^1)}{h(q,r)nc_1^1 c_2^1}] + \frac{n_2}{n}[\frac{\alpha(k_1 c_2^2 - k_2 c_1^2)}{h(q,r)nc_1^2 c_2^2}] \tag{14}$$

This quantity could be positive or negative. In fact, Equation 14 is equivalent to:

$$E(s_1^* - s_2^*) = \frac{\alpha}{h(q,r)n^2}[k_1(\frac{n_1 c_2^1}{c_1^1} + \frac{n_2 c_2^2}{c_2^1}) - k_2(\frac{n_1 c_1^1}{c_2^1} + \frac{n_2 c_1^2}{c_2^2})] \tag{15}$$

Remark that although $k_1 < k_2$, it is possible that the first term in the brackets could, for some values of the citation costs and groups' sizes, be larger than the second term, leading to papers authored by minority scholars being more frequently cited than similar papers from majority scholars. However, this would plausibly happen only under rare circumstances because the sign is driven by how much minority authors cite coethnic authors more than non-coethnic authors and whether this differential citation outweighs how much majority authors cite coethnic authors more than non-coethnic authors. Given the larger size of the majority group, it is obvious from Equation 14, that the sign of the average citation gap is plausibly driven by the second term of the equation, which is negative.

What happens to the racial citation gap when the citation costs are **symmetric**? We assume that a minority author and a majority author are symmetric in terms of their costs, so that the predictions of the model are not driven by differential citation costs between minority and majority authors. Therefore, let $c_1^1 = c_2^2 = \kappa$ and $c_2^1 = c_1^2 = \gamma$. We know by assumption that κ is smaller than γ ($\kappa < \gamma$) because of the aforementioned assumption that the cost of citing a paper written by a coethnic author is smaller than the cost of citing a paper written by an author from a different ethnic group.

The following proposition states that the racial citation gap favors majority authors when the citation costs are symmetric.

Proposition B.2. (Racial citation gap) *When the citation costs are symmetric, $E(s_1^* - s_2^*) < 0$, that is, papers authored by majority scholars are, on average, cited more frequently than papers authors by minority scholars.*

Proof. When the citation costs are symmetric, the racial citation gap in a paper written by a **minority author** is:

$$s_1^{1*} - s_2^{1*} = \frac{\alpha(k_1\gamma - k_2\kappa)}{h(q,r)n\kappa\gamma} \tag{16}$$

Similarly, racial ciation gap in a paper written by a majority scholar is:

$$s_1^{2*} - s_2^{2*} = \frac{\alpha(k_1\kappa - k_2\gamma)}{h(q,r)n\kappa\gamma} \tag{17}$$

The **average** difference in the number of citations to papers authored by minority scholars and papers authored by majority scholars is therefore:

$$E(s_1^* - s_2^*) = \frac{n_1}{n}[\frac{\alpha(k_1\gamma - k_2\kappa)}{h(q,r)n\kappa\gamma}] + \frac{n_2}{n}[\frac{\alpha(k_1\kappa - k_2\gamma)}{h(q,r)n\kappa\gamma}] = \frac{\alpha(n_1^2 - n_2^2)(a\gamma - b\kappa)}{h(q,r)n^2\kappa\gamma} \tag{18}$$

This quantity is clearly negative since $n_1^2 - n_2^2 \leq 0$ (because $n_2 \geq n_1$) and $a\gamma - b\kappa > 0$ (because $a \geq b$ and $\gamma > \kappa$). This means that works authored by minority scholars receive fewer citations, on average, than works authored by majority scholars. This prediction is consistent with our empirical findings.

B.3 The racial homophily gap

In this section, we will analyze how homophily in citation behavior varies between majority and minority authors. Homophily is the tendency to cite coethnic authors. We define the homophily level of an author of group i, $i = 1, 2$, as:

$$H_i = \frac{s_i^{i*}/s_{i'}^{i*}}{n_i/n_{i'}} \tag{19}$$

where $i = 1$ if $i' = 2$ and $i = 2$ if $i' = 1$. This is the relative number of citations given to works authored by coethnic scholars adjusted by their own relative group size.

It follows from Equations 7 and 8 that:

$$H_1 = \frac{k_1 n_2}{k_2 n_1}\frac{c_2^1}{c_1^1} \tag{20}$$

and

$$H_2 = \frac{k_2 n_1}{k_1 n_2}\frac{c_1^2}{c_2^2} \tag{21}$$

The relative difference in homophily level between majority and minority authors is therefore:

$$\frac{H_2}{H_1} = (\frac{k_2 n_1}{k_1 n_2})^2 \frac{c_1^1}{c_2^1}\frac{c_1^2}{c_2^2} \tag{22}$$

The following proposition gives a necessary and sufficient condition for the level of homophily to be greater among majority authors than minority authors.

Proposition B.3. (Racial homophily gap) *The homophily level is greater for majority authors than for minority authors (that is, $\frac{H_2}{H_1} > 1$) if and only if the cost of citing the work of a non-coethnic scholar relative to a coethnic scholar (defined as $\frac{c_{i'}^i}{c_i^i}$) is sufficiently larger for majority authors compared to minority authors; that is, if and only if $\frac{c_1^2}{c_2^2} > K\frac{c_2^1}{c_1^1}$ where $K = (\frac{k_1 n_2}{k_2 n_1})^2 \geq 1$.*

Proof. To prove this proposition, note that $\frac{H_2}{H_1} = (\frac{k_2 n_1}{k_1 n_2})^2 \frac{c_1^1}{c_2^1}\frac{c_1^2}{c_2^2} > 1$ is equivalent to:

$$\frac{c_1^2}{c_2^2} > (\frac{k_1 n_2}{k_2 n_1})^2 \frac{c_2^1}{c_1^1} \tag{23}$$

We need to prove that $K = (\frac{k_1 n_2}{k_2 n_1})^2 \geq 1$. For this purpose, recall that $k_1 = n_1 a + n_2 b$ and $k_2 = n_1 b + n_2 a$. It follows that $k_1 n_2 - k_2 n_1 = (n_1 a + n_2 b) n_2 - (n_1 b + n_2 a) n_1 = b n_2^2 - b n_1^2$; but $b n_2^2 - b n_1^2 = b(n_2 - n_1)(n_2 + n_1) \geq 0$ because $b \geq 0$ and $n_2 > n_1$ by assumption. Therefore, $k_1 n_2 - k_2 n_1 \geq 0$, which is equivalent to $k_1 n_2 \geq k_2 n_1$ or $\frac{k_1 n_2}{k_2 n_1} \geq 1$, which in turn implies $K = (\frac{k_1 n_2}{k_2 n_1})^2 \geq 1$. It follows that $\frac{H_2}{H_1} > 1$ if and only if $\frac{c_1^2}{c_2^2}$ is sufficiently larger than $\frac{c_2^1}{c_1^1}$.[24]

This result has important implications for understanding how citation homophily differs between majority and minority scholars in the economics profession. Suppose, for example, that $K = 1.5$. Consider two scenarios in which the relative cost of citing a non-coethnic versus a coethnic is higher for majority authors than for minority authors. In the first scenario, let $\frac{c_1^2}{c_2^2} = 10$ and $\frac{c_2^1}{c_1^1} = 5$; in the second, let $\frac{c_1^2}{c_2^2} = 7$ and $\frac{c_2^1}{c_1^1} = 5$. Despite the fact that the relative citation cost is higher for majority authors in both cases, the implications for the homophily gap differ. In the first scenario, the condition $\frac{c_1^2}{c_2^2} > K \frac{c_2^1}{c_1^1}$ holds, implying that majority authors will exhibit greater citation homophily than minority authors. In the second scenario, by contrast, $\frac{c_1^2}{c_2^2} < K \frac{c_2^1}{c_1^1}$, implying that citation homophily will be greater among minority authors.

The illustration above may also help explain why we observe greater citation homophily among White authors compared to non-White authors in this paper, whereas Koffi [2021b] finds higher homophily among women than men. Non-White scholars and women each comprise roughly 20 percent of the author population, yet the relative cost faced by a man in citing a woman versus another man may be smaller than the cost faced by a White scholar in citing a non-White scholar versus another White. In other words, the relative citation costs—captured by $\frac{c_1^2}{c_2^2}$ and $\frac{c_2^1}{c_1^1}$—may differ more starkly between racial groups than between genders. One possible explanation is that cross-gender interactions are more common and socially integrated than cross-racial ones. This assumption is plausible at the societal level: while most households and academic departments include both men and women, significantly fewer include members from different racial backgrounds.

B.4 Other results

B.4.1 Bounding the effects of referee preference/bias

In this section, we bound the effect of referee preference for diversity, analyze how referee bias affects author homophily in citation behavior, and determine whether strategic citation (Angrist et al. [2020]) could arise in response to such bias. For simplicity in the exposition, we assume that the citation costs are symmetric. In the absence of referee bias, citations to works authored by coethnic scholars are given the same weight as citations to works authored by non-coethnic scholars; that is, $a = b$ in the objective function. However, in the presence of extreme bias, a referee only finds works authored by coethnic scholars to be informative; in this case, $a = 1$ and $b = 0$.

In the **absence of referee bias** ($a = b$), we have $k_1 = k_2$, and it follows from Equations 16 and 17 that:

$$s_1^{1*} - s_2^{1*} = -(s_1^{2*} - s_2^{2*}) = \frac{\alpha(\gamma - \kappa)}{2h(q,r)\kappa\gamma} \tag{24}$$

[24] Remark that because $a \geq b$, the constant $K = (\frac{k_1 n_2}{k_2 n_1})^2$ attains its maximum value at $(\frac{n_2}{n_1})^2$ (this is when $a = 1$ and $b = 0$) and its minimum value at 1 (this is when $a = b$).

This quantity implies that when referee bias is minimal, homophilic behavior in citation does not vary by author identity; minority and majority authors display the same level of homophily.

The **average** racial citation gap in this case is:

$$E(s_1^* - s_2^*) = \frac{n_1}{n}[\frac{\alpha(\gamma - \kappa)}{2h(q,r)\kappa\gamma}] + \frac{n_2}{n}[\frac{\alpha(\kappa - \gamma)}{2h(q,r)\kappa\gamma}] = \frac{n_1 - n_2}{n}[\frac{\alpha(\gamma - \kappa)}{2h(q,r)\kappa\gamma}] \tag{25}$$

This is negative, suggesting that works authored by minority scholars receive fewer citations than works authored by majority scholars. We can see from Equation 25 that this is driven by the differential size of the minority and majority groups ($n_1 - n_2$) and the differential cost of citing coethnics relative to non-coethnics ($\gamma - \kappa$). In fact, as the difference in the size of the two groups decreases (i.e., $n_1 - n_2$ tends to zero), the racial citation gap tends to zero as well. Similarly as the differential citation cost decreases, the citation gap decreases as well.

In the **presence of extreme referee bias** ($a = 1$ and $b = 0$), we can show from Equation 16 that the racial citation gap in a paper authored by a **minority author** is:

$$s_1^{1*} - s_2^{1*} = \frac{\alpha(n_1\gamma - n_2\kappa)}{h(q,r)n\kappa\gamma} \tag{26}$$

As for the general case, this quantity can be positive or negative depending on the relative size of the minority group. In particular, if the size of the minority group is sufficiently small relative to the majority group, minority authors will tend to cite works authored by scholars from the majority group more than the works of coethnic authors.

From Equation 17, the racial citation gap in a paper of a **majority author** is:

$$s_1^{2*} - s_2^{2*} = \frac{\alpha(n_1\kappa - n_2\gamma)}{h(q,r)n\kappa\gamma} \tag{27}$$

As for the general case, this is always negative, meaning that scholars from the majority group tend to cite works authored by coethnic scholars more .

The **average** racial citation gap is:

$$E(s_1^* - s_2^*) = \frac{n_1}{n}[\frac{\alpha(n_1\gamma - n_2\kappa)}{h(q,r)n\kappa\gamma}] + \frac{n_2}{n}[\frac{\alpha(n_1\kappa - n_2\gamma)}{h(q,r)n\kappa\gamma}] = \frac{\alpha[(n_1^2 - n_2^2)]}{h(q,r)n^2\kappa\gamma} \tag{28}$$

This is negative because the numerator is negative, again showing that papers written by minority scholars receive fewer citations on average. The racial citation gap is driven both the referee bias (or preference for diversity) and the differential size of the minority and the majority groups. Comparing Equations 25 and 28 therefore yields the relative size of the racial citation bias that is due to extreme referee bias:

$$\frac{E(s_1^* - s_2^*)_{a=1,b=0}}{E(s_1^* - s_2^*)_{a=b}} = \frac{2\gamma}{\gamma - \kappa} \tag{29}$$

Given that this quantity is larger than 1, it shows that the combination of differential group size and referee bias leads to a larger racial citation gap than just the differential group size.

We have seen that in the absence of referee bias, both minority and majority authors are homophilic to the same extent. However, when referee bias is present, majority authors tend to be more homophilic than minority authors. This implies that part of the racial citation gap in papers authored by minority scholars can be attributable to **strategic citation** (Angrist et al. [2020]). How big is the size of the racial citation gap due to strategic citation among minority authors? One can

calculate it by taking the difference of the absolute value of the racial citation gap between majority and minority authors.

When minority authors cite works authored by minority scholars more frequently than they do for works authored by White scholars, the racial citation gap attributable to strategic citation is equal to:

$$(s_1^{1*} - s_2^{1*})_{strategic} = |s_1^{2*} - s_2^{2*}| - |s_1^{1*} - s_2^{1*}| = \frac{\alpha[(n_2 - n_1)(\kappa + \gamma)]}{h(q,r)n\kappa\gamma} \tag{30}$$

Otherwise, the racial citation gap attributable to strategic citation is equal to:

$$(s_1^{1*} - s_2^{1*})_{strategic} = |s_1^{2*} - s_2^{2*}| - |s_1^{1*} - s_2^{1*}| = \frac{\alpha[n(\kappa + \gamma)]}{h(q,r)n\kappa\gamma} \tag{31}$$

Remark that this quantity also measures the extent to which majority authors are more coethnic scholars than minority authors. It is always positive, with the size of the citation gap due to strategic citation being clearly larger when minority authors cite majority authors more. Our empirical findings corroborate these predictions, as we have found that homophilic citation is higher for papers written by an all-White team than papers with at least one minority author.

B.4.2 The effects of citation costs, group size, and referee preferences

Does the differential cost of citing coethnics relative to non-coethnics alone drive the citation gap? The model suggests that the answer is no. In fact, from Equation 18, one can see that, as γ tends to κ (that is, the difference between the cost of citing coethnic authors versus non-coethnic authors is minimal), the racial citation gap $E(s_1^* - s_2^*)$ tends to $\frac{\alpha(n_1^2 - n_2^2)(a-b)}{h(q,r)n^2\kappa}$, which is negative. Similarly, referee preference for diversity alone does not drive the citation gap. In fact, when the referee is neutral ($a = b$), we have seen that the racial citation gap is negative (Equation 25). The differential group size ($n_1 - n_2$) seems to play a significant role in explaining the gap. However, its effect only appears when referee bias is present. It follows from the analysis that relative citation costs, differential group size, and referee preferences for diversity play complementary roles in determining the racial citation gap.

While the differential cost of citing coethnics relative to non-coethnics solely does not explain the racial citation gap, our empirical findings suggest that it partly explains the tendency to cite coethnic authors. In fact, if the differential citation cost is sufficiently small (γ tends to κ), then Equation 16 implies that minority authors cite works authored by majority authors more frequently than works from coethnics. This is not consistent with our empirical analysis. Similarly, our empirical findings suggest that both referee preference and the differential size of racial groups are important at explaining the tendency to cite works from coethnic scholars. In fact, our analysis implies that if referees are neutral ($a = b$) and/or the difference between in group size is very small, then minority and majority authors will display the same tendency, which is not what we find empirically.

B.4.3 Other predictions of the model: Heterogeneity analysis

In this section, we state other predictions of the model that are supported by our empirical analysis.

- From Equations 18, 25, and 28, the racial citation gap ($E(s_1^* - s_2^*)$) becomes more pronounced as paper quality (α) increases; this implies that the citation penalty experienced by minority scholars is larger for their high quality papers than for their low quality papers.[25]

[25] If one assumes that higher quality papers are published in higher ranked journals, this prediction implies that

- From Equations 18, 25, and 28, the racial citation gap ($E(s_1^* - s_2^*)$), becomes more pronounced for papers written by higher ability authors (r); this implies that the citation penalty is experienced more strongly by minority scholars who are more prominent.

- From Equations 18, 25, and 28, citing papers published in higher quality journals (q) discriminate less against the works of minority scholars; in other words, more prominent journals discriminate less against minority scholars.[26]

- From Equations 18, 25, and 28, the racial citation gap ($E(s_1^* - s_2^*)$) becomes more pronounced as the relative size of the minority group decreases. This prediction has an implication for how the racial citation gap varies across different fields (e.g., **theory vs. empirics**); it implies that in fields where minority scholars are significantly underrepresented, the racial citation gap is larger. The reduction of the gap over the past three decades that we find in the data is attributable to a greater representation of minority scholars in the empirical field, which corroborates the model's prediction.

B.5 A modification of the model

In this section, we show a modification of the model where author ability (r) instead enhances the quality of the paper, similar to the well-known total factor productivity in the Cobb-Douglas production function. Assume that a paper cites (s_1, s_2) papers of average quality α. Assume that the perceived value of this paper is $\alpha(aln(s_1) + bln(s_2)) + ln(r)$ to a scholar from the minority group and $\alpha(bln(s_1) + aln(s_2) + ln(r))$ to a scholar from the majority group, where a and b are positive, $a + b = 1$, and $a \geq b$. The average perceived value of the paper to a random referee is therefore:

$$\frac{n_1}{n_1 + n_2}\alpha[aln(s_1) + bln(s_2)] + \frac{n_2}{n_1 + n_2}\alpha[bln(s_1) + aln(s_2)] + ln(r)$$

Taking the exponential transform of this expression and letting $n = n_1 + n_2$, the overall average perceived quality of the paper is a function of the author's ability (r) and a combination of the paper's references, based on the notion that scholarly work is often influenced by existing literature:

$$v(s_1, s_2) = r \cdot s_1^{\alpha\frac{n_1 a + n_2 b}{n}} \cdot s_2^{\alpha\frac{n_1 b + n_2 a}{n}}$$

This is the traditional Cobb-Douglas production function. Assume that the total cost of producing the paper is defined as in the previous model, with the exception of the omission of the parameter capturing the author's quality:

$$c(s_1, s_2) = l(q)(c_1 s_1 + c_2 s_2)$$

where $l(q)$ is positive and increasing in q. The function $l(q)$ measures how hard it is to publish in a journal of quality q.

The author therefore solves the following maximization problem:

$$\max_{(s_1, s_2)} \quad u(s_1, s_2) = r s_1^{\frac{\alpha(n_1 a + n_2 b)}{n}} s_2^{\frac{\alpha(n_1 b + n_2 a)}{n}} - l(q)(c_1 s_1 + c_2 s_2)$$

the works of minority scholars published in more prominent journals suffer a higher citation penalty. Our empirical findings corroborate this prediction.

[26] This is different from saying that the racial gap in citations to papers published in higher quality journals is smaller.

As previously, let $k_1 = n_1 a + n_2 b$ and $k_2 = n_1 b + n_2 a$. It is easy to remark that this maximization problem is equivalent to:

$$\max_{(s_1, s_2)} \quad u(s_1, s_2) = r s_1^{\frac{\alpha k_1}{n}} s_2^{\frac{\alpha k_2}{n}} - l(q)(c_1 s_1 + c_2 s_2)$$

The F.O.Cs imply the following solution:

$$s_1^* = \left(\frac{\alpha k_1 r}{n l(q) c_1}\right)^{\frac{1}{1-\alpha}} \left(\frac{k_2 c_1}{k_1 c_2}\right)^{\frac{\alpha k_2}{(1-\alpha)k_1}} \tag{32}$$

$$s_2^* = \left(\frac{\alpha k_2 r}{n l(q) c_2}\right)^{\frac{1}{1-\alpha}} \left(\frac{k_1 c_2}{k_2 c_1}\right)^{\frac{\alpha k_1}{(1-\alpha)k_2}} \tag{33}$$

$$s_1^* - s_2^* = \left[\frac{(k_1 c_2 - k_2 c_1)}{k_1 c_2}\right] s_1^* \tag{34}$$

Assuming, as previously that a minority author and a majority author are symmetric in terms of their costs, we have $c_1^1 = c_2^2 = \kappa$ and $c_2^1 = c_1^2 = \gamma$. We note that κ is smaller than γ.

We find that author identity determines homophilic behavior, and that the findings are similar to those of the previous model. For a **minority author**, the racial citation gap is:

$$s_1^{1*} - s_2^{1*} = \frac{(k_1\gamma - k_2\kappa)}{k_1\gamma} s_1^* \tag{35}$$

This quantity could be positive (**homophily**) or negative (**heterophily**) depending on the relative size of the minority group.

For a **majority author**, the racial citation gap is:

$$s_1^* - s_2^* = \frac{(k_1\kappa - k_2\gamma)}{k_1\kappa} s_1^* \tag{36}$$

This is always negative (**homophily**).

The **average** racial citation gap is therefore:

$$E(s_1^* - s_2^*) = \frac{n_1}{n}[\frac{(k_1\gamma - k_2\kappa)}{k_1\gamma} s_1^*] + \frac{n_2}{n}[\frac{(k_1\kappa - k_2\gamma)}{k_1\kappa} s_1^*] \tag{37}$$

This quantity is negative, which means that works authored by minority scholars receive fewer citations that similar quality works authored by scholars from the majority group. This prediction is corroborated by our empirical findings.

Equations (21) and (22) also imply that the extent of homophily is larger for authors that belong to the majority group than for authors that belong to the minority group. This prediction is also consistent with our empirical findings. Overall, the analysis shows that the main insights of the theoretical analysis are not too sensitive to functional forms.

C Additional Tables

Table C.1: Race audit

	Predicted race		
	All White	At least one Black/Hispanic/Asian	Total
Hand-checked race			
All White	2,919	280	3,199
At least one Black/Hispanic/Asian	161	1,221	1,382
Total	3,080	1,501	4,581

This table presents the results of the race audit based on the manually verified sample of authors described in Section 2.3. The "predicted race" refers to the race inferred through the algorithmic procedure, whereas the "hand-checked race" corresponds to the classification obtained from manual review of curriculum vitae data, using visual cues and information on country of origin. To ensure comparability between the two classifications, we exclude observations for which the algorithmic procedure predicted the race as "Other."

Table C.2: Heterogeneous effect: Number of authors

	Outcome variable: Citation (asinh)		
	Solo Author	Two Authors	More than Two authors
	(1)	(2)	(3)
At least one Black/Hispanic/Asian	-0.050***	-0.072***	-0.032***
	(0.011)	(0.008)	(0.009)
Other races	-0.053***	-0.057***	-0.033***
	(0.010)	(0.009)	(0.010)
Observations	122779	122380	84072
R-sqr	0.458	0.535	0.613
Publications and team size controls	Y	Y	Y
Affiliation fixed effects	Y	Y	Y
Field fixed effects	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, focusing on the heterogeneous results depending on the number of authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Citations and race: baseline estimation on manually checked sample

	Outcome variable: Citation (asinh)						
	Checked race on checked sample						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
At least one Black/Hispanic/Asian	-0.335***	-0.303***	-0.235***	-0.084***	-0.100***	-0.094***	-0.075**
	(0.050)	(0.051)	(0.046)	(0.031)	(0.030)	(0.030)	(0.029)
Mixed races	-0.440	-0.392	-0.330	-0.093	-0.247	-0.214	-0.255
	(0.557)	(0.547)	(0.498)	(0.247)	(0.220)	(0.211)	(0.202)
Observations	5544	5544	5544	5544	5544	5544	5544
Adj. R-sqr	0.008	0.010	0.182	0.629	0.643	0.660	0.684
	Algorithmic-based race on checked sample						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
At least one Black/Hispanic/Asian	-0.357***	-0.318***	-0.252***	-0.132***	-0.144***	-0.129***	-0.099***
	(0.054)	(0.055)	(0.050)	(0.034)	(0.033)	(0.033)	(0.033)
Other races	-0.386***	-0.356***	-0.225***	-0.013	-0.023	-0.016	0.023
	(0.061)	(0.062)	(0.057)	(0.038)	(0.037)	(0.036)	(0.036)
Observations	5544	5544	5544	5544	5544	5544	5544
Adj. R-sqr	0.011	0.013	0.182	0.629	0.644	0.660	0.684
Team size controls		Y	Y	Y	Y	Y	Y
Publications controls			Y	Y	Y	Y	Y
Year fixed effects				Y	Y	Y	
Affiliation fixed effects					Y	Y	Y
Field fixed effects						Y	Y
Journal-Year fixed effects							Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The estimation compared the sample where the race of the authors was checked, with the manually annotated race, to the case of algorithm-based determination of the race. The reference category for the race variable is *All-White*, which represents papers written by only White authors. "Mixed races" refers specifically to individuals identified as belonging to more than one racial category. In contrast, "Other races" is used to capture uncertainty in classification as described in Section 2.2.1. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Robustness: Alternative race definitions

	Outcome variable: Citation (asinh)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
At least one Black/Hispanic/Asian	-0.038***								
	(0.005)								
At least one Black/Hispanic/Asian		-0.038***							
		(0.005)							
At least one Black/Hispanic/Asian			-0.043***						
			(0.005)						
Other races			-0.040***						
			(0.006)						
At least one Black/Hispanic/Asian				-0.051***					
				(0.005)					
Other races				-0.044***					
				(0.006)					
At least one Black/Hispanic/Asian					-0.044***				
					(0.005)				
Other races					-0.052***				
					(0.006)				
At least one Black/Hispanic/Asian						-0.048***			
						(0.005)			
Other races						-0.022***			
						(0.008)			
At least one Black/Hispanic/Asian							-0.044***		
							(0.006)		
Other races							-0.024***		
							(0.005)		
At least one Black/Hispanic/Asian								-0.042***	
								(0.006)	
Other races								-0.051***	
								(0.005)	
At least one Black/Hispanic/Asian									-0.050***
									(0.005)
Other races									-0.024***
									(0.008)
Observations	329231	329231	329231	329231	329231	329231	329231	329231	329231
R-sqr	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, using various race definitions. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Columns (1) and (2) define race using predictions from the US census and Florida voter registration, respectively. There is no threshold restriction, so the category "Other races" is not included. Column (3) employs Florida voter registration if the author's first name is available and the US census database for last names only. Column (4) is the baseline estimation in Table II. Black economists are additionally identified through manual checks and a list from the National Economic Association. Column (5) is similar to Column (4) but without the probability threshold. Column (6) also uses a combination of US census and Florida voter registration data, with only a 0.7 probability threshold, including manual checks for Black economists. Column (7) is like Column (6) with a 0.9 probability threshold. Column (8) is akin to Column (4) but excludes manual checks for Black economists. Column (9) resembles Column (6) but also omits manual checks for Black economists. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.5: Robustness: Alternative race definitions (2)

	Outcome variable: Citation (asinh)		
	(1)	(2)	(3)
Share of Black/Hispanic/Asian	-0.049***		
	(0.007)		
Share of other races	-0.068***		
	(0.007)		
All Black/Hispanic/Asian		-0.042***	
		(0.008)	
At least one Black/Hispanic/Asian (not all)		-0.057***	
		(0.007)	
Other races		-0.044***	
		(0.006)	
All Black/Hispanic/Asian			-0.043***
			(0.008)
One Black/Hispanic/Asian and one White			-0.044***
			(0.009)
50% White (not teams of 2)			-0.021***
			(0.008)
At least one Black/Hispanic/Asian (not all)			-0.063***
			(0.008)
Other races			-0.060***
			(0.007)
Observations	329231	329231	329231
R-sqr	0.526	0.526	0.526
Publications and team size controls	Y	Y	Y
Affiliation fixed effects	Y	Y	Y
Field fixed effects	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, using various race definitions. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Column (1) uses the share of authors identified as Black, Hispanic, and Asian and the share of authors that are neither Black, Hispanic, Asian, nor White. Column (2) is a categorical variable equal to 0 for all-White authors, 1 for all Black, Hispanic, or Asian authors, 2 for teams with at least one Black, Hispanic, or Asian author (excluding teams where all authors are Black, Hispanic, or Asian), and 3 for other races. Column (3) is a categorical variable equal to 0 for all-White authors, 1 for all Black, Hispanic, or Asian authors, 2 for teams with one Black, Hispanic, or Asian author and one White author, 3 for teams where 50% of the authors are White and 50% are non-White, 4 for teams with at least one Black, Hispanic, or Asian author (excluding teams in the previous categories), and 5 for other races. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Robustness: Readability index

	Outcome variable: Citation (asinh)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	-0.051***	-0.054***	-0.055***	-0.055***	-0.054***	-0.054***	-0.055***	-0.058***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Other races	-0.044***	-0.044***	-0.044***	-0.044***	-0.044***	-0.044***	-0.044***	-0.044***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Flesh reading index			-0.003***					-0.002***
			(0.000)					(0.000)
Flesh Kincaid index				0.011***				0.033***
				(0.001)				(0.003)
Smog index					0.016***			0.012***
					(0.001)			(0.001)
Gunning Fog index						0.005***		-0.035***
						(0.001)		(0.002)
Dale-Shall index							-0.018***	0.002
							(0.002)	(0.002)
Observations	329231	317649	317649	317649	317649	317649	317649	317649
R-sqr	0.526	0.530	0.531	0.531	0.531	0.531	0.531	0.532
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, adding controls for the readability of the abstract of the paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Columns (1) and (2) are related to the baseline estimation without restriction and with restriction on papers for which we could compute the readability scores. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.7: Robustness: Functional form of the citation variable

	Citation count	log(1+citation)	Poisson model	Top 2 percentile of citation distribution	Top 5 percentile of citation distribution	Top 10 percentile of citation distribution	Remove papers with 0 citation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
At least one Black/Hispanic/Asian	-4.484***	-0.048***	-0.126***	-0.004***	-0.006***	-0.010***	-0.058***
	(0.720)	(0.005)	(0.020)	(0.001)	(0.001)	(0.001)	(0.005)
Other races	-4.402***	-0.041***	-0.118***	-0.002***	-0.005***	-0.008***	-0.048***
	(0.759)	(0.005)	(0.021)	(0.001)	(0.001)	(0.001)	(0.005)
Observations	329231	329231	328852	329231	329231	329231	293435
R-sqr	0.086	0.520		0.101	0.160	0.201	0.440
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y

This table shows the relationship between citation and race, using alternative functional forms of the dependent variable. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Columns (4) to (6), in particular, estimate a linear probability model where the outcome variable is 1 if the paper's citations are in the top 2% (Column 4), top 5% (Column 5), or top 10% (Column 6) of the citation distribution and 0 otherwise. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: Robustness: Alternative measures of authors' prominence

	Outcome variable: Citation (asinh)			
	>200	Top 35	Top 16	Top 5
	Measures based on the more prolific author			
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.053***	-0.078***	-0.067***	-0.091***
	(0.005)	(0.010)	(0.015)	(0.028)
Other races	-0.036***	-0.042***	-0.040**	-0.059*
	(0.006)	(0.012)	(0.018)	(0.031)
Observations	329231	89881	40322	16860
Adj. R-sqr	0.529	0.459	0.487	0.417
	Measures based on the average number of publications			
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.063***	-0.088***	-0.082***	-0.102***
	(0.006)	(0.011)	(0.016)	(0.028)
Other races	-0.038***	-0.055***	-0.060***	-0.085***
	(0.006)	(0.012)	(0.018)	(0.031)
Observations	329231	89881	40322	16860
Adj. R-sqr	0.524	0.451	0.480	0.405
	Measures based on the total number of publications			
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.054***	-0.079***	-0.077***	-0.103***
	(0.005)	(0.011)	(0.016)	(0.029)
Other races	-0.037***	-0.035***	-0.030	-0.062*
	(0.005)	(0.012)	(0.019)	(0.033)
Observations	329231	89881	40322	16860
Adj. R-sqr	0.533	0.463	0.492	0.419
Affiliation fixed effects	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, emphasizing the effect of seniority. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Seniority is measured by the authors' publication records and based on the maximum number of publications, the average number of publications, and the total number of publications by the team. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.9: Regression excluding self-citations

	Outcome variable: Citation (asinh)			
	> 200 journals	Top 35 journals	Top 16 journals	Top 5 journals
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.045***	-0.077***	-0.082***	-0.099***
	(0.006)	(0.012)	(0.018)	(0.030)
Other races	-0.042***	-0.044***	-0.052***	-0.071**
	(0.006)	(0.013)	(0.020)	(0.033)
Observations	329231	89881	40322	16860
Adj. R-sqr	0.518	0.480	0.513	0.444
Affiliation fixed effects	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, excluding self-citations from the total citation count. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.10: Selection on unobservables- Estimation with and without the full set of controls

	Outcome variable: Citation (asinh)							
	> 200 journals		Top 35 journals		Top 16 journals		Top 5 journals	
	Uncontrolled	Controlled	Uncontrolled	Controlled	Uncontrolled	Controlled	Uncontrolled	Controlled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	-0.139***	-0.051***	-0.211***	-0.086***	-0.231***	-0.085***	-0.068*	-0.096***
	(0.007)	(0.005)	(0.014)	(0.010)	(0.021)	(0.015)	(0.037)	(0.028)
				Disaggregation				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black	-0.147***	-0.067***	-0.107	-0.110**	-0.131	-0.186***	0.053	-0.200*
	(0.031)	(0.021)	(0.068)	(0.046)	(0.100)	(0.069)	(0.166)	(0.120)
At least one Hispanic	-0.218***	-0.062***	-0.351***	-0.073***	-0.323***	-0.070***	-0.055	-0.147***
	(0.012)	(0.008)	(0.026)	(0.018)	(0.041)	(0.027)	(0.066)	(0.044)
At least one Asian	-0.111***	-0.046***	-0.176***	-0.089***	-0.209***	-0.085***	-0.079*	-0.071**
	(0.008)	(0.006)	(0.015)	(0.012)	(0.023)	(0.017)	(0.043)	(0.032)
Other races	-0.166***	-0.044***	-0.259***	-0.050***	-0.304***	-0.055***	-0.287***	-0.073**
	(0.008)	(0.006)	(0.015)	(0.012)	(0.024)	(0.018)	(0.039)	(0.031)
Observations	329231	329231	89881	89881	40322	40322	16860	16860
R-sqr	0.002	0.526	0.005	0.455	0.006	0.485	0.003	0.413
Publications and team size controls		Y		Y		Y		Y
Affiliation fixed effects		Y		Y		Y		Y
Field fixed effects		Y		Y		Y		Y
Journal-Year fixed effects		Y		Y		Y		Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Uneven columns represent the regression model without control variables. Even columns represent the regression model with the full set of control variables. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.11: Citation and race: paper-time level regression

	Outcome variable: Citation (asinh)							
	> 200 journals		Top 35 journals		Top 16 journals		Top 5 journals	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	-0.012***	-0.012***	-0.018***	-0.019***	-0.023***	-0.028***	-0.032***	-0.037***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.006)	(0.010)	(0.010)
Other races	-0.005***	-0.004***	-0.004	-0.005	0.002	-0.001	0.007	0.003
	(0.001)	(0.001)	(0.003)	(0.003)	(0.006)	(0.006)	(0.009)	(0.009)
Observations	6606106	6606106	2251757	2251757	1070057	1070057	570892	570892
Adj. R-sqr	0.406	0.406	0.471	0.471	0.508	0.509	0.525	0.527
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Paper cumulative citation up to time t	Y	Y	Y	Y	Y	Y	Y	Y
Publications in top 5, top 16, top 35, overall at time t	Y	Y	Y	Y	Y	Y	Y	Y
Average cumulative citations of the authors up to time t		Y		Y		Y		Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Columns (1) and (2) use the full baseline sample. Columns (3) and (4) restrict the sample to 35 high-impact journals identified by Card et al. [2020]. Columns (5) and (6) limit the sample to 16 high-impact journals according to Koffi [2021b]. Columns (7) and (8) restrict the sample to the top five economics journals: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.12: US-based publications and citations

	Outcome variable: Citation (asinh)			
	> 200 journals	Top 35 journals	Top 16 journals	Top 5 journals
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.056***	-0.083***	-0.061***	-0.098***
	(0.009)	(0.013)	(0.019)	(0.033)
Other races	-0.043***	-0.046***	-0.021	-0.002
	(0.010)	(0.016)	(0.022)	(0.037)
Observations	118111	45690	21886	7258
Adj. R-sqr	0.539	0.530	0.572	0.559
Affiliation fixed effects	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Here, we analyze the institutions of the authors, focusing on a sample where at least one author in the paper is based in the United States (US), and those papers are also cited by other US-based authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.13: US-based publications and citations: Homophily

	Outcome variable: Citation (asinh)							
	> 200 journals		Top 35 journals		Top 16 journals		Top 5 journals	
	From non-White	From White	From non-White	From White	From non-White	From White	From non-White	From White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
At least one Black/Hispanic/Asian	0.206***	-0.306***	0.179***	-0.338***	0.199***	-0.296***	0.182***	-0.287***
	(0.008)	(0.008)	(0.013)	(0.013)	(0.019)	(0.018)	(0.034)	(0.033)
Other races	0.006	-0.184***	0.019	-0.178***	0.057**	-0.119***	0.095**	-0.072**
	(0.008)	(0.008)	(0.015)	(0.015)	(0.023)	(0.022)	(0.038)	(0.036)
Observations	118111	118111	45690	45690	21886	21886	7258	7258
R-sqr	0.450	0.483	0.452	0.489	0.447	0.475	0.465	0.463

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, emphasizing the race of the authors of the citing paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Columns (1), (3), (5), and (7) restrict the citation count to citations from only teams, including non-White scholars. Columns (2), (4), (6), and (8) restrict the citation count to citations from only teams including all-White scholars. Here, we analyze the institutions of the authors, focusing on a sample where at least one author in the paper is based in the United States (US), and those papers are also cited by other US-based authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.14: Citation dynamics: Gender vs Race

	Outcome variable: Citation (asinh)				
	Overall	From non-White	From White	From at least one woman	From all men
	(1)	(2)	(3)	(4)	(5)
At least one Black/Hispanic/Asian	-0.059***	0.232***	-0.303***		
	(0.006)	(0.004)	(0.004)		
At least one Woman	0.070***			0.203***	-0.090***
	(0.006)			(0.004)	(0.005)
Observations	329231	329231	329231	329231	329231
R-sqr	0.526	0.419	0.460	0.387	0.472

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial and gender composition of the paper's authors, emphasizing the race and gender of the authors of the citing paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The reference category for the race variable is *All-Men*, which represents papers written by only male authors. Column (1) considers all the citation counts. Column (2) restricts the citation count to citations from only teams including non-White scholars. Column (3) restricts the citation count to citations from only teams, including all-White scholars. Column (4) restricts the citation count to citations from only teams including at least one woman scholar. Column (5) restricts the citation count to citations from only teams, including all-men scholars. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.15: Effect of seminars and conference attendance: Author level seminars and conferences

	Outcome variable: Citation (asinh)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
At least one Black/Hispanic/Asian	-0.087***	-0.086***	-0.086***	-0.085***	-0.091***	-0.075**	-0.072**
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Other races	-0.064	-0.050	-0.034	-0.038	-0.029	0.057	0.053
	(0.242)	(0.239)	(0.238)	(0.238)	(0.234)	(0.234)	(0.233)
Seminars/conferences count at time of publication		0.007***		0.003			0.002
		(0.002)		(0.003)			(0.003)
Prominent conferences at time of publication		0.045**		0.021			0.020
		(0.018)		(0.021)			(0.021)
Seminars in top-tier affiliations at time of publication		-0.012		-0.004			-0.008
		(0.010)		(0.011)			(0.011)
Seminars/conferences count 1 year before paper publication			0.010***	0.008*			0.008*
			(0.004)	(0.004)			(0.004)
Seminars in top-tier affiliations 1 year before paper publication			-0.002	-0.002			-0.003
			(0.012)	(0.012)			(0.012)
Prominent conferences 1 year before paper publication			0.010	0.005			0.006
			(0.021)	(0.022)			(0.022)
Seminars/conferences count 2 years before paper publication			0.001	0.001			0.002
			(0.003)	(0.003)			(0.003)
Seminars in top-tier affiliations 2 years before paper publication			-0.016	-0.016			-0.022*
			(0.012)	(0.012)			(0.012)
Prominent conferences 2 years before paper publication			0.017	0.011			0.003
			(0.022)	(0.023)			(0.023)
Seminars/conferences count 3 years before paper publication			-0.004	-0.004			-0.003
			(0.003)	(0.003)			(0.003)
Seminars in top-tier affiliations 3 years before paper publication			0.002	0.002			-0.004
			(0.011)	(0.011)			(0.011)
Prominent conferences 3 years before paper publication			0.060***	0.059**			0.051**
			(0.023)	(0.023)			(0.023)
Seminars/conferences count 4 years before paper publication			0.001	0.001			0.001
			(0.003)	(0.003)			(0.003)
Seminars in top-tier affiliations 4 years before paper publication			-0.025**	-0.024*			-0.022*
			(0.012)	(0.013)			(0.012)
Prominent conferences 4 years before paper publication			-0.008	-0.009			-0.005
			(0.023)	(0.023)			(0.023)
Number of editorial positions before paper publication					0.001**	0.001**	0.001**
					(0.001)	(0.001)	(0.001)
Any top 5 journals editorial positions before paper publication					0.170***	0.140***	0.137***
					(0.045)	(0.045)	(0.045)
Any top 35 journals editorial positions (not top 5) before paper publication					0.120***	0.092**	0.084*
					(0.043)	(0.043)	(0.043)
Total number of other visiting positions before paper publication					-0.010	-0.015	-0.023*
					(0.013)	(0.013)	(0.013)
Total number of visiting positions at top-tier instititions before paper pub.					0.065***	0.060***	0.058***
					(0.017)	(0.016)	(0.017)
Any authors from a top-tier PhD institition					0.055	0.037	0.041
					(0.038)	(0.038)	(0.038)
Any authors member of NBER, JPAL, CEPR, BREAD, IZA, CIFAR						0.138***	0.132***
						(0.020)	(0.020)
Observations	4427	4427	4427	4427	4427	4427	4427
Adj. R-sqr	0.705	0.706	0.707	0.707	0.709	0.712	0.713
Publications and team size controls	Y	Y	Y	Y	Y	Y	Y
Affiliation fixed effects	Y	Y	Y	Y	Y	Y	Y
Field fixed effects	Y	Y	Y	Y	Y	Y	Y
Journal-Year fixed effects	Y	Y	Y	Y	Y	Y	Y

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, emphasizing the effect of seminars and conference attendance. The reference category for the race variable is *All-White*, which represents papers written by only White authors. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.16: Effect of seminars and conference attendance: Univariate

	Outcome variable: Citation (asinh)			
	(1)	(2)	(3)	(4)
At least one Black/Hispanic/Asian	-0.305***	-0.253***	-0.225***	-0.222***
	(0.056)	(0.054)	(0.053)	(0.053)
Other races	0.196	0.119	0.069	0.069
	(0.621)	(0.561)	(0.555)	(0.555)
Seminars/conferences count before paper publication		-0.059***		-0.019***
		(0.005)		(0.006)
Prominent conferences before paper publication		0.161***		0.096**
		(0.037)		(0.041)
Seminars in top-tier affiliations before paper publication		0.057***		0.029
		(0.022)		(0.022)
Seminars/conferences count 1 before paper publication			-0.042***	-0.033***
			(0.009)	(0.009)
Seminars in top-tier affiliations 1 before paper publication			0.027	0.017
			(0.024)	(0.024)
Prominent conferences 1 before paper publication			0.000	0.000
			(.)	(.)
Seminars/conferences count 2 before paper publication			-0.027***	-0.023**
			(0.009)	(0.009)
Seminars in top-tier affiliations 2 before paper publication			0.008	0.001
			(0.024)	(0.024)
Prominent conferences 2 before paper publication			0.000	0.000
			(.)	(.)
Seminars/conferences count 3 before paper publication			-0.010	-0.008
			(0.007)	(0.007)
Seminars in top-tier affiliations 3 before paper publication			-0.006	-0.009
			(0.024)	(0.024)
Prominent conferences 3 before paper publication			0.000	0.000
			(.)	(.)
Seminars/conferences count 4 before paper publication			-0.033***	-0.030***
			(0.012)	(0.011)
Seminars in top-tier affiliations 4 before paper publication			-0.002	-0.006
			(0.026)	(0.026)
Prominent conferences 4 before paper publication			0.000	0.000
			(.)	(.)
Observations	4427	4427	4427	4427
Adj. R-sqr	0.006	0.063	0.091	0.092

This table shows the relationship between the citation (inverse hyperbolic sine—asinh) and the racial composition of the paper's authors, emphasizing the effect of seminars and conference attendance.

Table C.17: Effect of online availability

	Outcome variable: Citation (asinh)						
	All citations					By non-White authors	By White authors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
At least one Black/Hispanic/Asian × After online availability (0-7 years)	0.003						
	(0.006)						
At least one Black/Hispanic/Asian × After online availability (0-7 years)	0.010*						
	(0.006)						
At least one Black/Hispanic/Asian × After online availability (0-7 years)		-0.001	-0.001	-0.001	-0.002	0.007	-0.004
		(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)
At least one Black/Hispanic/Asian × After online availability (0-7 years)		-0.043***	-0.043***	-0.043***	-0.040***	-0.028***	-0.017***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.005)	(0.005)
At least one Black/Hispanic/Asian × After online availability (0-7 years)		0.029***	0.029***	0.029***	0.028***	0.020***	0.010**
		(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.005)
Observations	2911446	2911446	2911446	2911446	2911446	2911446	2911446
Adj. R-sqr	0.574	0.575	0.576	0.576	0.576	0.459	0.481
Paper fixed effects	Y	Y	Y	Y	Y	Y	Y
Year of citation fixed effects	Y	Y	Y	Y	Y	Y	Y
Journal fixed effects				Y			
Year of publication fixed effects				Y			
Journal*Year of publication fixed effects	Y	Y			Y	Y	Y
Authors' prior citations					Y		

The table reports average effects from estimating Equation 2, which assesses the impact of online availability on the racial citation gap. Each column represents a distinct regression specification. The coefficient at event time $\tau = -1$, the citation year immediately preceding online availability, is normalized to zero and serves as the reference period. All specifications control for the author's time-varying publication history and include flexible controls for paper age, modeled using age, age squared, age cubed, and age to the fourth power. Standard errors, clustered at the paper level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.18: Effect of non-coethnic collaborations

	Outcome variable: Share of bibliographical references to non-coethnic authors				
	All (White and non-White authors)	White authors	Non-White authors	All (White and non-White authors)	
	(1)	(2)	(3)	(4)	(5)
Before non-coethnic collaboration (Pre-treatment)	0.003	0.006***	0.008	0.002	0.002
	(0.002)	(0.002)	(0.015)	(0.002)	(0.003)
After non-coethnic collaboration (Post-treatment)	0.016***	0.015***	0.012	0.016***	0.013***
	(0.002)	(0.002)	(0.012)	(0.002)	(0.002)
Observations	232430	204012	28416	232430	153049
Adj. R-sqr	0.566	0.255	0.319	0.566	0.586
Same JEL with non-coethnic paper				Y	Y
Share of non-White authors in the field					Y

This table presents average effects from estimating Equation 3, which evaluates the impact of non-coethnic collaboration on the share of bibliographic references made to non-coethnic authors. Each column corresponds to a separate regression specification. The coefficient at event time $\tau = -1$—the year of publication immediately preceding the first non-coethnic collaboration—is normalized to zero and serves as the reference period. The coefficient labeled "Before non-coethnic collaboration (Pre-treatment)" represents the average of the estimated coefficients for all event times $\tau < -1$, while the coefficient labeled "After non-coethnic collaboration (Post-treatment)" corresponds to the average of the coefficients for all event times $\tau \geq 0$. All regressions include author fixed effects, publication year fixed effects, and controls for time-varying team size and the author's publication history. The sample size changes in Column 5 as JEL classification codes are not available for all publications. Standard errors, clustered at the author level, are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.19: Positionality in authorship listing and citations

Panel A: Paper level regression

	Outcome variable: Citation (asinh)						
	All citations	From non-White	From White	From non-White	From White	From non-White	From White
						Teams with more than 2 authors	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Paper's first author non-White	-0.012	0.044***	-0.093***	0.109***	-0.148***	0.106***	-0.142***
	(0.013)	(0.011)	(0.010)	(0.014)	(0.013)	(0.016)	(0.014)
Share of non-white except first author	-0.036	0.091***	-0.127***				
	(0.022)	(0.019)	(0.018)				
Paper's second author non-White				0.100***	-0.111***	0.117***	-0.097***
				(0.014)	(0.013)	(0.016)	(0.014)
Paper's third author non-White				0.120***	-0.080***	0.122***	-0.096***
				(0.014)	(0.013)	(0.016)	(0.014)
Observations	40751	40751	40751	40751	40751	23825	23825
Adj. R-sqr	0.594	0.479	0.513	0.480	0.514	0.507	0.545

Panel B: Author level regression

	Outcome variable: Citation (asinh)						
	All citations	From non-White	From White	From non-White	From White	From non-White	From White
						Teams with more than 2 authors	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First author	0.018	-0.020	0.134***	-0.045**	0.063***	-0.025	0.084***
	(0.025)	(0.023)	(0.022)	(0.018)	(0.017)	(0.022)	(0.021)
First author × Black/Hispanic/Asian	-0.022	0.060**	-0.278***	0.068**	-0.168***	0.032	-0.203***
	(0.033)	(0.030)	(0.028)	(0.030)	(0.028)	(0.036)	(0.033)
Share of non-white except first author	-0.042	0.076***	-0.288***				
	(0.028)	(0.026)	(0.024)				
Paper's second author non-White				0.036***	-0.069***	0.051***	-0.080***
				(0.013)	(0.012)	(0.015)	(0.014)
Paper's third author non-White				0.066***	-0.042***	0.070***	-0.044***
				(0.014)	(0.013)	(0.015)	(0.014)
Observations	73984	73984	73984	73984	73984	45920	45920
Adj. R-sqr	0.558	0.425	0.477	0.426	0.475	0.445	0.494
Author fixed effects	Y	Y	Y	Y	Y	Y	Y

This table examines the relationship between citation and author race, with a focus on the positionality of authors within mixed-race teams. The dependent variable is the citation count, measured using the inverse hyperbolic sine transformation. The analysis is restricted to papers authored by teams including at least one White and one non-White author. Panel A presents estimates from a paper-level regression that includes control variables for team size, authors' publication records, institutional affiliations, field of research, journal, and publication year fixed effects. Panel B reports results from an author-level regression, which incorporates the same controls as in Panel A, along with author fixed effects to exploit within-author variation in citation outcomes. Robust standard errors are reported in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.20: List of Journals

ACCOUNTING ORGANIZATIONS AND SOCIETY	1,194
ACTA STRUCTILIA	1
AGRICULTURAL ECONOMICS	1,581
AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS	421
AMERICAN ECONOMIC JOURNAL-ECONOMIC POLICY	407
AMERICAN ECONOMIC JOURNAL-MACROECONOMICS	336
AMERICAN ECONOMIC JOURNAL-MICROECONOMICS	393
AMERICAN ECONOMIC REVIEW	5,273
AMERICAN JOURNAL OF AGRICULTURAL ECONOMICS	4,920
AMERICAN LAW AND ECONOMICS REVIEW	194
ANNALS OF ECONOMICS AND FINANCE	303
APPLIED ECONOMIC PERSPECTIVES AND POLICY	299
APPLIED ECONOMICS	8,354
ECONOMETRIC REVIEWS	581
APPLIED ECONOMICS LETTERS	5,857
ASIAN JOURNAL OF LAW AND ECONOMICS	123
American Economic Journal-Economic Policy	2
B E JOURNAL OF ECONOMIC ANALYSIS & POLICY	915
B E JOURNAL OF MACROECONOMICS	469
BALTIC JOURNAL OF ECONOMIC STUDIES	850
BELL JOURNAL OF ECONOMICS AND MANAGEMENT SCIENCE	46
BROOKINGS PAPERS ON ECONOMIC ACTIVITY	438
CAMBRIDGE JOURNAL OF ECONOMICS	1,729
CANADIAN JOURNAL OF ECONOMICS-REVUE CANADIENNE D ECONOMIQUE	1,777
CESIFO ECONOMIC STUDIES	355
CHINA ECONOMIC REVIEW	1,280
CRITICAL FINANCE REVIEW	32
ECOLOGICAL ECONOMICS	5,046
ECONOMETRIC THEORY	1,355
ECONOMETRICA	3,605
ECONOMETRICS JOURNAL	339
ECONOMIC AND POLITICAL STUDIES-EPS	162
ECONOMIC DEVELOPMENT AND CULTURAL CHANGE	1,474
ECONOMIC INQUIRY	2,539
ECONOMIC JOURNAL	3,450
ECONOMIC POLICY	322
ECONOMIC RECORD	1,508
ECONOMIC SYSTEMS	422
ECONOMIC SYSTEMS RESEARCH	326
ECONOMIC THEORY	2,255
ECONOMICA	1,973
ECONOMICS & POLITICS	188
ECONOMICS LETTERS	11,175
ECONOMICS OF EDUCATION REVIEW	1,641
ECONOMICS OF GOVERNANCE	171
ECONOMICS OF INNOVATION AND NEW TECHNOLOGY	652
ECONOMICS OF TRANSITION	504
ECONOMIST	637
EMERGING MARKETS REVIEW	477
EMPIRICAL ECONOMICS	1,820
ENERGY ECONOMICS	4,329
ENERGY JOURNAL	1,089
ENERGY POLICY	10,968
ENVIRONMENT AND DEVELOPMENT ECONOMICS	659
ENVIRONMENTAL & RESOURCE ECONOMICS	1,933

EUROPEAN ECONOMIC REVIEW	3,411
EUROPEAN FINANCIAL MANAGEMENT	531
EUROPEAN JOURNAL OF POLITICAL ECONOMY	833
EXPERIMENTAL ECONOMICS	583
EXPLORATIONS IN ECONOMIC HISTORY	916
FEDERAL RESERVE BANK OF ST LOUIS REVIEW	381
FINANCE A UVER	166
FINANCE A UVER-CZECH JOURNAL OF ECONOMICS AND FINANCE	482
FINANCE AND STOCHASTICS	542
FINANCIAL MANAGEMENT	1,476
FISCAL STUDIES	416
FOUNDATIONS AND TRENDS IN ENTREPRENEURSHIP	78
GAMES AND ECONOMIC BEHAVIOR	2,566
HEALTH ECONOMICS	2,274
IMF ECONOMIC REVIEW	225
IMF STAFF PAPERS	298
INDUSTRIAL & LABOR RELATIONS REVIEW	1,445
INDUSTRIAL AND CORPORATE CHANGE	716
INTERNATIONAL BUSINESS REVIEW	1,105
INTERNATIONAL ECONOMIC REVIEW	2,304
INTERNATIONAL FINANCE	157
INTERNATIONAL JOURNAL OF CENTRAL BANKING	431
INTERNATIONAL JOURNAL OF FINANCE & ECONOMICS	821
INTERNATIONAL JOURNAL OF FORECASTING	1,807
INTERNATIONAL JOURNAL OF GAME THEORY	1,144
INTERNATIONAL JOURNAL OF INDUSTRIAL ORGANIZATION	1,717
INTERNATIONAL LABOUR REVIEW	1,779
INTERNATIONAL REVIEW OF ENVIRONMENTAL AND RESOURCE ECONOMICS	111
INTERNATIONAL TAX AND PUBLIC FINANCE	753
IZA JOURNAL OF LABOR POLICY	157
JOURNAL OF ACCOUNTING & ECONOMICS	973
JOURNAL OF ACCOUNTING AND PUBLIC POLICY	565
JOURNAL OF ACCOUNTING RESEARCH	1,593
JOURNAL OF AFRICAN ECONOMIES	623
JOURNAL OF AGRICULTURAL ECONOMICS	1,359
JOURNAL OF APPLIED ECONOMETRICS	1,330
JOURNAL OF APPLIED ECONOMICS	303
JOURNAL OF BANKING & FINANCE	5,246
JOURNAL OF BUSINESS	1,465
JOURNAL OF BUSINESS & ECONOMIC STATISTICS	1,597
JOURNAL OF BUSINESS VENTURING	1,111
JOURNAL OF COMPARATIVE ECONOMICS	1,205
JOURNAL OF CONSUMER RESEARCH	2,199
JOURNAL OF CORPORATE FINANCE	1,566
JOURNAL OF DEVELOPMENT ECONOMICS	2,555
JOURNAL OF DEVELOPMENT STUDIES	2,614
JOURNAL OF ECONOMETRICS	3,623
JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION	4,644
JOURNAL OF ECONOMIC DYNAMICS & CONTROL	3,474
JOURNAL OF ECONOMIC GEOGRAPHY	648
JOURNAL OF ECONOMIC GROWTH	260
JOURNAL OF ECONOMIC HISTORY	1,853
JOURNAL OF ECONOMIC INEQUALITY	312
JOURNAL OF ECONOMIC ISSUES	1,910
JOURNAL OF ECONOMIC LITERATURE	743
JOURNAL OF ECONOMIC PERSPECTIVES	1,262
JOURNAL OF ECONOMIC PSYCHOLOGY	1,783
JOURNAL OF ECONOMIC SURVEYS	563

JOURNAL OF ECONOMIC THEORY	3,882
JOURNAL OF ECONOMICS & MANAGEMENT STRATEGY	928
JOURNAL OF EMPIRICAL FINANCE	827
JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT	1,976
JOURNAL OF EVOLUTIONARY ECONOMICS	713
JOURNAL OF FINANCE	4,001
JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS	2,436
JOURNAL OF FINANCIAL ECONOMETRICS	265
JOURNAL OF FINANCIAL ECONOMICS	2,481
JOURNAL OF FINANCIAL INTERMEDIATION	506
JOURNAL OF FINANCIAL MARKETS	471
JOURNAL OF FINANCIAL SERVICES RESEARCH	474
JOURNAL OF FINANCIAL STABILITY	610
JOURNAL OF GLOBAL ECONOMIC ANALYSIS	45
JOURNAL OF HEALTH ECONOMICS	1,795
JOURNAL OF HOUSING ECONOMICS	560
JOURNAL OF HUMAN CAPITAL	173
JOURNAL OF HUMAN RESOURCES	1,307
JOURNAL OF INDUSTRIAL ECONOMICS	1,170
JOURNAL OF INTERNATIONAL BUSINESS STUDIES	1,699
JOURNAL OF INTERNATIONAL ECONOMICS	2,402
JOURNAL OF INTERNATIONAL FINANCIAL MARKETS INSTITUTIONS & MONEY	807
JOURNAL OF INTERNATIONAL MANAGEMENT	395
JOURNAL OF INTERNATIONAL MONEY AND FINANCE	2,332
JOURNAL OF LABOR ECONOMICS	1,028
JOURNAL OF LAW & ECONOMICS	1,138
JOURNAL OF LAW ECONOMICS & ORGANIZATION	751
JOURNAL OF LEGAL STUDIES	857
JOURNAL OF MACROECONOMICS	2,032
JOURNAL OF MANAGEMENT STUDIES	1,664
JOURNAL OF MATHEMATICAL ECONOMICS	1,828
JOURNAL OF MONETARY ECONOMICS	1,955
JOURNAL OF MONEY CREDIT AND BANKING	2,143
JOURNAL OF POLITICAL ECONOMY	3,044
JOURNAL OF POPULATION ECONOMICS	1,051
JOURNAL OF PRODUCTIVITY ANALYSIS	741
JOURNAL OF PUBLIC ECONOMICS	3,549
JOURNAL OF REAL ESTATE FINANCE AND ECONOMICS	1,017
JOURNAL OF REGIONAL SCIENCE	1,444
JOURNAL OF REGULATORY ECONOMICS	754
JOURNAL OF RETAILING	1,670
JOURNAL OF RISK AND INSURANCE	1,810
JOURNAL OF RISK AND UNCERTAINTY	740
JOURNAL OF TAXATION	2,367
JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION	7,941
JOURNAL OF THE ASSOCIATION OF ENVIRONMENTAL AND RESOURCE ECONOMISTS	232
JOURNAL OF THE EUROPEAN ECONOMIC ASSOCIATION	887
JOURNAL OF THE JAPANESE AND INTERNATIONAL ECONOMIES	619
JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES B-STATISTICAL METHODOLOGY	2,030
JOURNAL OF TRANSPORT ECONOMICS AND POLICY	1,056
JOURNAL OF URBAN ECONOMICS	1,557
JOURNAL OF WORLD BUSINESS	965
KYKLOS	1,336
LABOUR ECONOMICS	1,240
M&SOM-MANUFACTURING & SERVICE OPERATIONS MANAGEMENT	662

MACROECONOMIC DYNAMICS	1,161
MACROECONOMICS ANNUAL 2002	5
MANAGEMENT SCIENCE	6,254
MATHEMATICAL FINANCE	681
NATIONAL TAX JOURNAL	2,035
NBER MACROECONOMICS ANNUAL	86
NBER MACROECONOMICS ANNUAL 1997	10
NBER MACROECONOMICS ANNUAL 1998	5
NBER MACROECONOMICS ANNUAL 1999	6
NBER MACROECONOMICS ANNUAL 2000	5
NBER MACROECONOMICS ANNUAL 2001	6
NBER MACROECONOMICS ANNUAL 2003	8
NEW ENGLAND REVIEW-MIDDLEBURY SERIES	548
ORGANIZATION SCIENCE	1,597
OXFORD BULLETIN OF ECONOMICS AND STATISTICS	1,543
OXFORD ECONOMIC PAPERS-NEW SERIES	2,167
OXFORD REVIEW OF ECONOMIC POLICY	981
POLITICAL ANALYSIS	469
PORTUGUESE ECONOMIC JOURNAL	158
PUBLIC CHOICE	3,528
QME-QUANTITATIVE MARKETING AND ECONOMICS	245
QUANTITATIVE ECONOMICS	284
QUARTERLY JOURNAL OF ECONOMICS	2,391
RAND JOURNAL OF ECONOMICS	1,339
REAL ESTATE ECONOMICS	690
REGIONAL SCIENCE AND URBAN ECONOMICS	1,859
REGIONAL STUDIES	3,166
RESEARCH POLICY	3,082
RESOURCE AND ENERGY ECONOMICS	793
REVIEW OF ECONOMIC DYNAMICS	883
REVIEW OF ECONOMIC STUDIES	2,547
REVIEW OF ECONOMICS AND STATISTICS	4,499
REVIEW OF ENVIRONMENTAL ECONOMICS AND POLICY	183
REVIEW OF FINANCE	538
REVIEW OF FINANCIAL ECONOMICS	369
REVIEW OF FINANCIAL STUDIES	1,825
REVIEW OF INCOME AND WEALTH	964
REVIEW OF INTERNATIONAL ECONOMICS	829
REVIEW OF NETWORK ECONOMICS	146
REVIEW OF WORLD ECONOMICS	563
SCANDINAVIAN JOURNAL OF ECONOMICS	1,150
SMALL BUSINESS ECONOMICS	1,809
SOCIAL CHOICE AND WELFARE	1,674
SOUTHERN ECONOMIC JOURNAL	2,888
STATA JOURNAL	809
STRATEGIC MANAGEMENT JOURNAL	2,549
THEORETICAL ECONOMICS	266
THEORY AND DECISION	1,639
TRANSPORTATION	1,522
TRANSPORTATION RESEARCH PART A-POLICY AND PRACTICE	1
WATER RESOURCES RESEARCH	1
WELTWIRTSCHAFTLICHES ARCHIV-REVIEW OF WORLD ECONOMICS	1,478
WORLD BANK ECONOMIC REVIEW	760
WORLD BANK RESEARCH OBSERVER	262
WORLD DEVELOPMENT	5,566

D Additional Figures

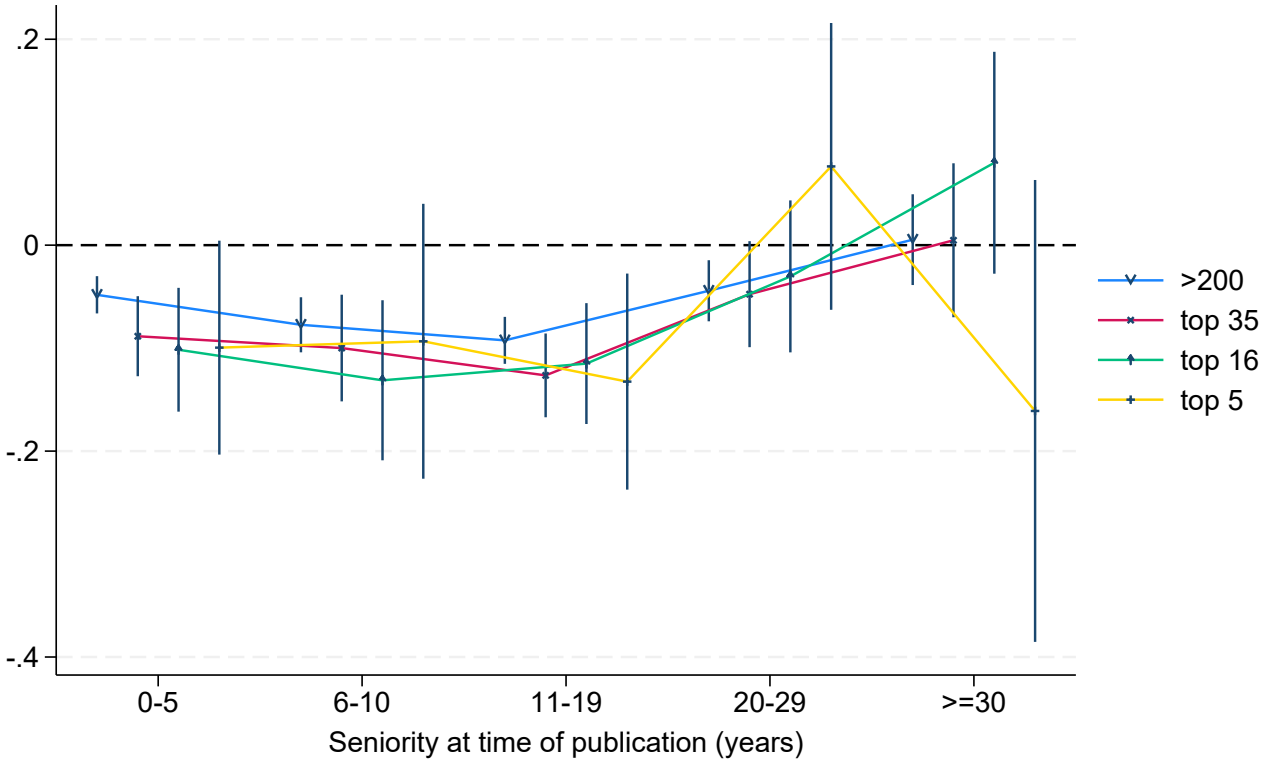


Figure D.1:
Citation, race and seniority

The figure presents the relationship between the citation (inverse hyperbolic sine—asinh), the outcome variable, and the racial composition of the paper's authors, the independent variable of interest, focusing on the effect of seniority. Seniority is measured by the number of years between the first publication of the more senior coauthor and the publication year. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The plots display the estimate for teams with at least one Black/Hispanic/Asian author by comparison to a team with all-White authors.

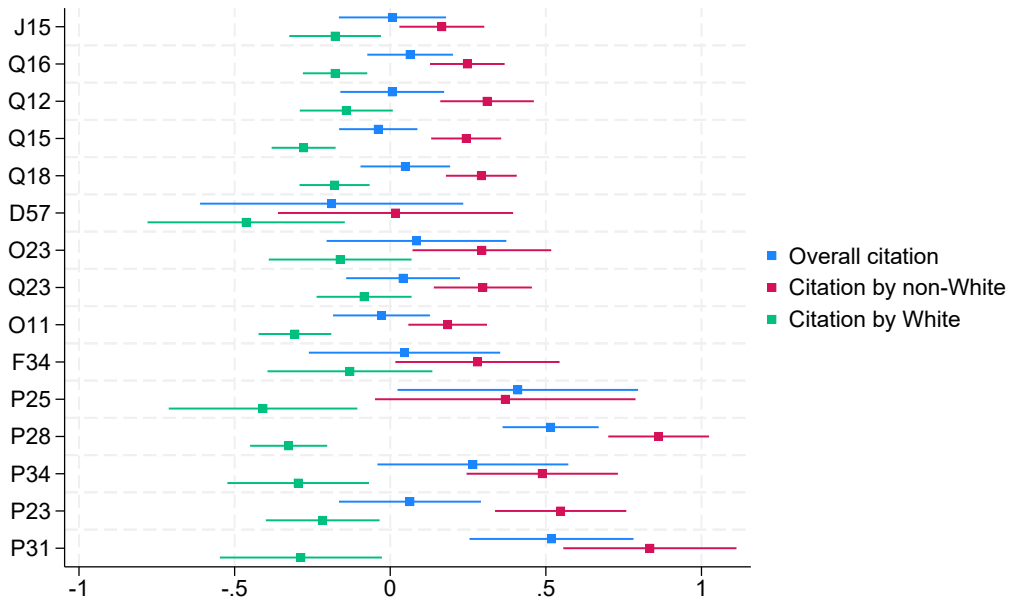


Figure D.2:
Citation and race of the authors in the citing paper

The figure presents the relationship between citation and race for the three-digit JEL code fields with higher fractions of non-White scholars, focusing on the racial composition of the authors of the citing paper. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The plots display the estimate for teams with at least one Black/Hispanic/Asian author by comparison to a team with all-White authors.

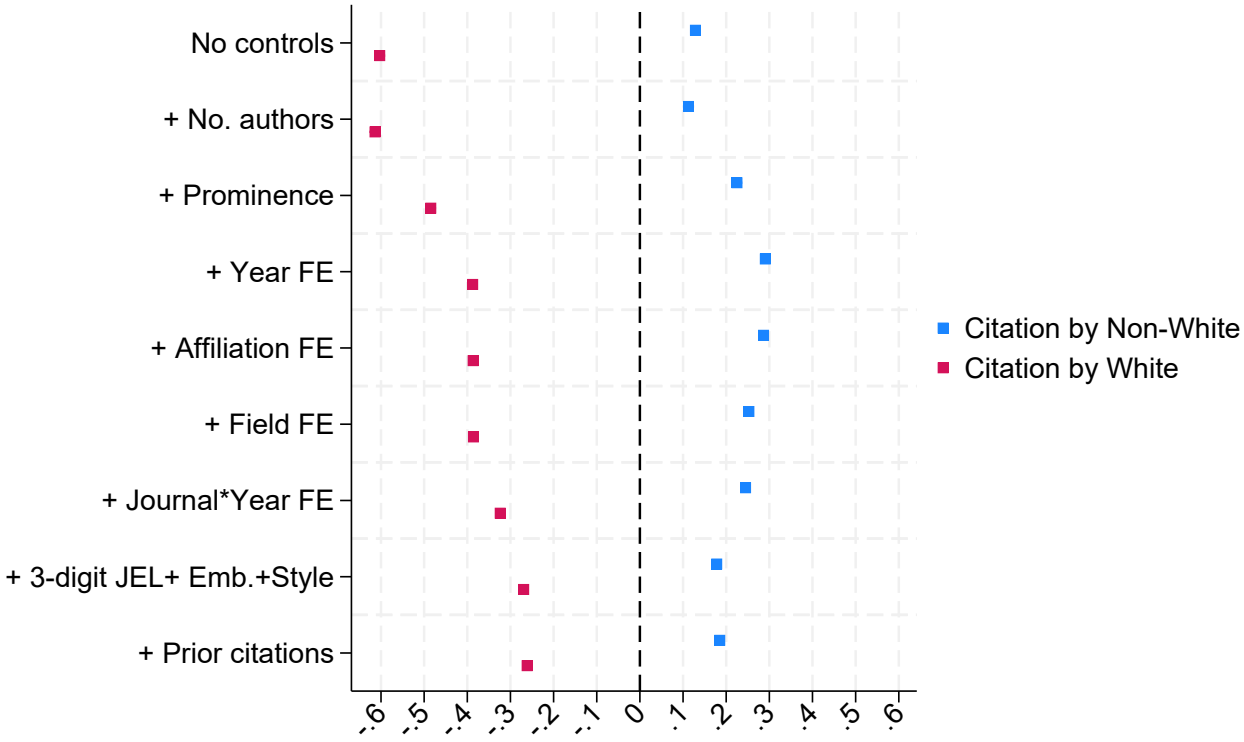


Figure D.3:
Citation and race of the authors: Incremental controls and saturated model

The figure presents the relationship between citation and race, focusing on the racial composition of the authors of the citing paper and adding the control variables incrementally. The reference category for the race variable is *All-White*, which represents papers written by only White authors. The plots display the estimate for teams with at least one Black/Hispanic/Asian author by comparison to a team with all-White authors. "No controls" is the unconditional value. "+ No. authors" adds team size. "+ Prominence" includes authors' publication history with team size. "+ Year FE" adds year fixed effects to team size and publication history. "+ Affiliation FE" incorporates affiliation fixed effects with the prior controls. "+ Field FE" adds broad field fixed effects. "+ Journal*Year FE" includes journal-by-year fixed effects with all prior controls. "+ 3-digit JEL+Emb.+Style" adds granular field controls (800 JEL dummies, 768 embedding vectors, 2 style dummies) to the previous controls. "+ Prior citations" adds authors' prior citations to all controls.

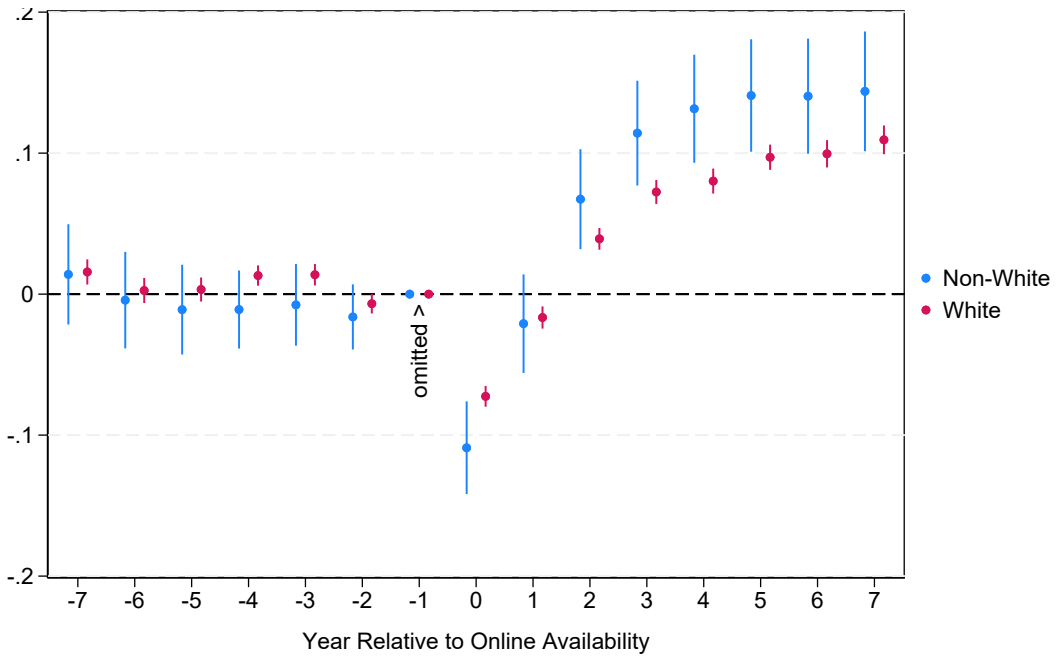
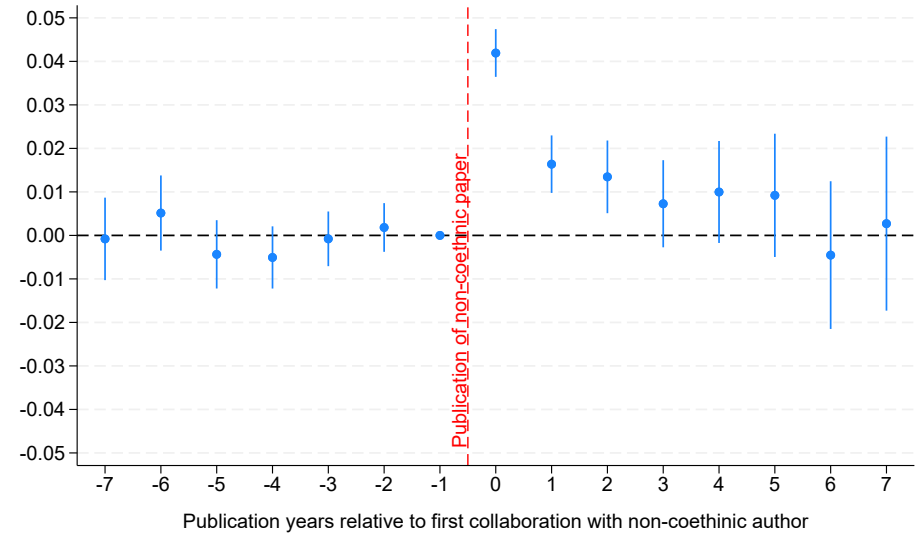
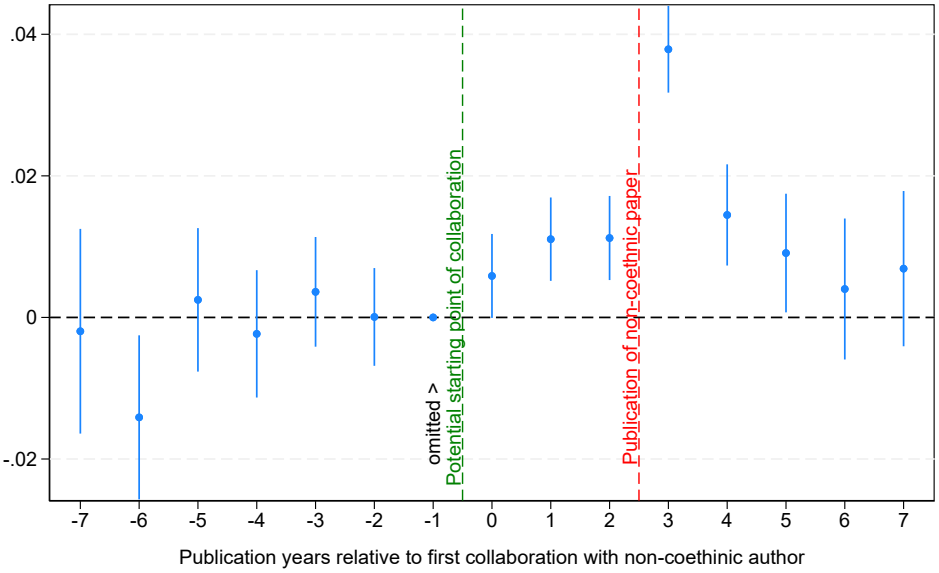


Figure D.4:
Effect of online availability on the racial citation gap (Sun and Abraham [2021] estimation)

The figure presents the event-study coefficients from estimating Equation 2, which assesses the impact of online availability on the racial citation gap, focusing on the robustness using the estimation method by Sun and Abraham [2021]. The coefficient at event time $\tau = -1$, the citation year immediately preceding online availability, is normalized to zero and serves as the reference period. The specification controls for the author's time-varying publication history, includes flexible controls for paper age, modeled using age, age squared, age cubed, and age to the fourth power, as well as paper fixed effects, citation year fixed effects, and journal times publication year fixed effects. Standard errors clustered at the paper level.



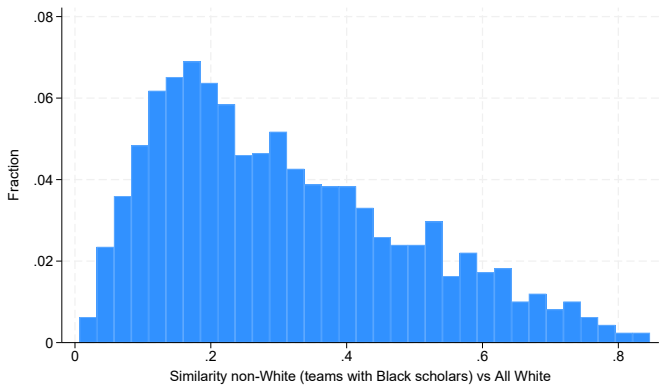
(a) Matching estimation



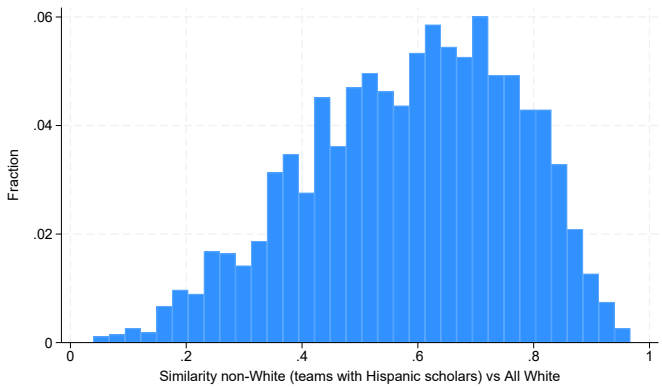
(b) Sun and Abraham [2021] estimation

Figure D.5:
Effect of collaboration with a non-coethnic author

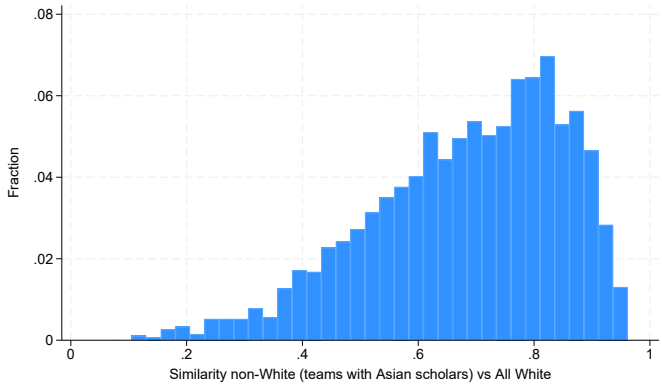
The figure presents the event-study coefficients from estimating the impact of collaboration with a non-coethnic coauthor on the share of bibliographic references to papers authored by members of the out-group based on equation 3. Panel A reports results using propensity score matching, while Panel B accounts for treatment effect heterogeneity using the estimator proposed by Sun and Abraham [2021]. Due to computational constraints, the Sun and Abraham [2021] estimation is restricted to a sample beginning in 1990. The analysis is conducted at the author level. Following an author's first collaboration with a non-coethnic coauthor, we track all subsequent publications involving only coethnic coauthors, truncating the sample at the point of the next non-coethnic collaboration.



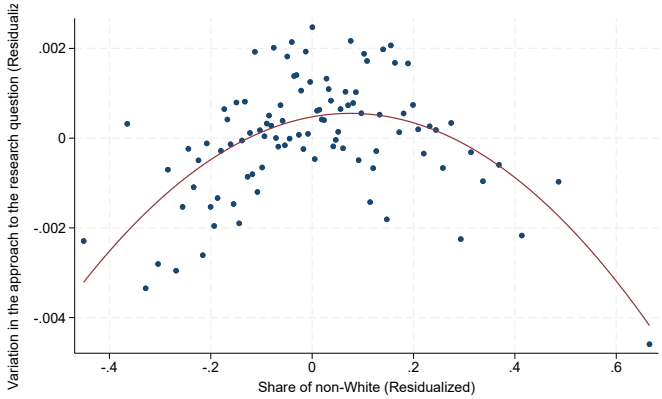
(a) Similarity distribution: Black versus White



(b) Similarity distribution: Hispanic versus White



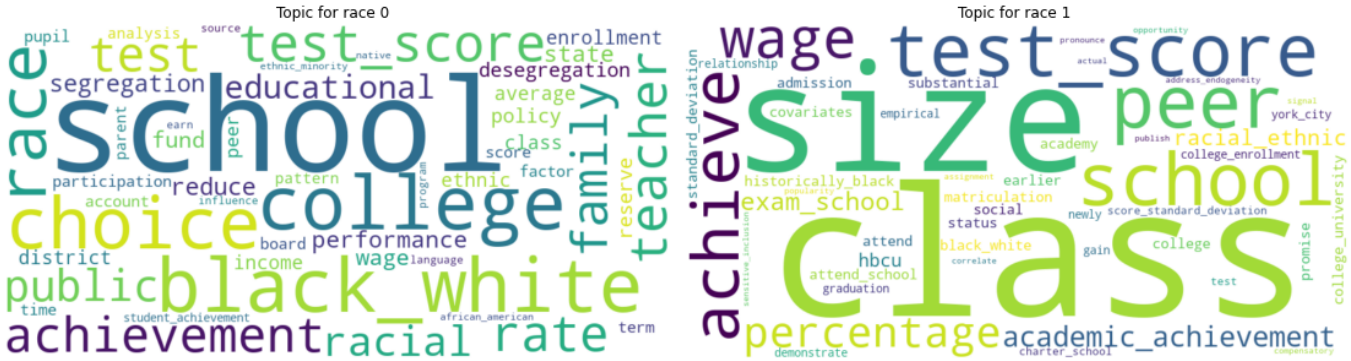
(c) Similarity distribution: Asian versus White



(d) Residualized plot

Figure D.6:
Racial categories and variation in the approach to the research question

The figure illustrates variations in research approaches across racial categories. Within each JEL code, papers are clustered into groups averaging 200 observations based on embedding representations. For each cluster, racial categories are treated as subgroups, and semantic differences are analyzed using the term frequency-inverse document frequency (TF-IDF) metric. The category "All-White teams" is used as the reference category. Figures (a), (b), and (c) show the distribution of this similarity metric between teams with Black, Hispanic, and Asian scholars, respectively, and all-White teams. Figure (d) examines the similarity distance between a paper and the cluster centroid within each cluster. It then computes the standard deviation of these similarity scores within a cluster and correlates this with the proportion of non-White scholars in the cluster. The similarity metric and the proportion of non-White scholars are calculated over time and then residualized using cluster and time-fixed effects to generate the plot.



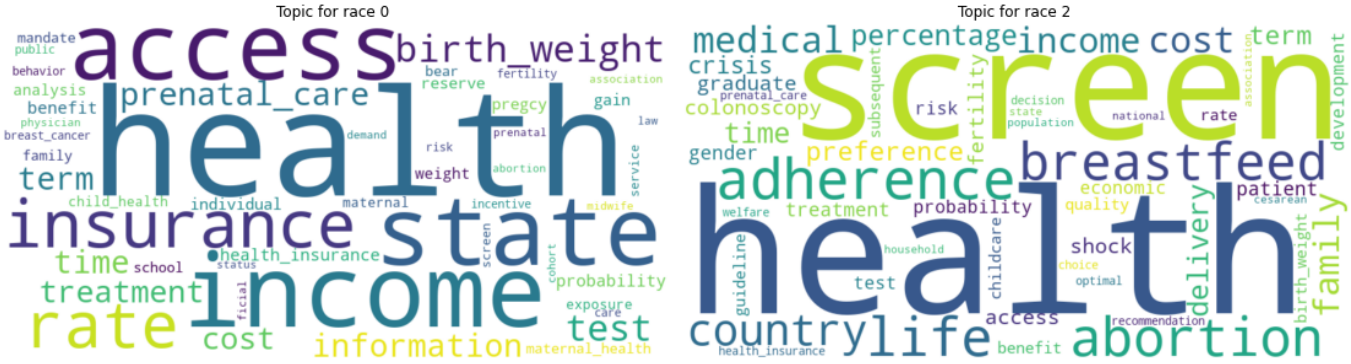
White Non-White

Panel A: Topic: Economics of Minorities; Cluster: Child, student, education



White Non-White

Panel B: Topic: Economics of Gender; Cluster: Women education



White Non-White

Panel C: Topic: Economics of Gender; Cluster: Women health

Figure D.7:
Examples: Racial categories and variation in the topic angle

The figure presents examples of the word clouds of each cluster by different racial categories.



White Non-White

Panel D: Topic: Wage level and structure; Cluster: Worker earning



White Non-White

Panel E: Topic: Agricultural Policy; Cluster: Country farmer market



White Non-White

Panel F: Topic: International Trade; Cluster: China's export

Figure D.8:
Examples: Racial categories and variation in the topic angle

The figure presents examples of the word clouds of each cluster by different racial categories.