Race-related Research in Economics*

Arun Advani Elliott Ash Anton Boltachka David Cai Imran Rasul

September 2025

Abstract

Issues of racial justice and economic inequalities between racial and ethnic groups have risen to the top of public debate. Economists' ability to contribute to these debates is based on the body of race-related research. We study the volume and content of race-related research in economics. We base our analysis on a corpus of $225,000$ economics publications from 1960 to 2020 to which we apply an algorithmic approach to classify race-related work. We present three new facts. First, since 1960 less than 2% of economics publications have been race-related, representing a cumulative body of knowledge of just over 4000 race-related publications in economics from 1960. There is an uptick in such work in the mid 1990s. Among the top-5 journals this is driven by the *American Economic Review*, *Quarterly Journal of Economics* and the *Journal of Political Economy*. *Econometrica* and the *Review of Economic Studies* have each cumulatively published fewer than 20 race-related articles since 1960, corresponding to .3% of all publications in those two journals. Second, on content, while over 50% of race-related publications in the 1970s focused on Black individuals, by the 2010s this had fallen to 20%. There has been a steady decline in the share of race-related research on discrimination since the 1980s, with a rise in the share of studies on identity. Finally, we apply our algorithm to NBER and CEPR working papers posted over the last four decades, to study an earlier stage of the research process. We document a concentration of race-related research into a few fields, and its continued absence from many others – a result that holds even within the subset of research examining issues of inequality or diversity. We discuss implications of our findings for economists' ability to contribute to debates on race and ethnicity in the economy. *JEL: A11, B41*.

*We thank the ESRC CPP at IFS (ES/H021221/1, ES/Y002563/1) and Stone Centre at UCL for financial support, Lina Gakh, Kesten Laura, Ramachandran Manikandan and Peiyao Sun for excellent research assistance, and the authors of Angrist *et al.* [2020] for sharing their data on journal weights. We have benefited from comments from Oriana Bandiera, Timothy Besley, Francesco Caselli, Jeffrey Grogger, Jonathan Halket, Camille Landais, Attila Lindner, Andrew Oswald, David Schindler, Silvana Tenreryo, Eva Vivalt, Ashley Wong and numerous seminar participants. All errors remain our own. Advani: Warwick and IFS, a.advani.1@warwick.ac.uk; Ash: ETH Zurich, ashe@ethz.ch; Boltachka: ETH Zurich, aboltachka@gmail.com; Cai: LSE, david.jiyang.cai@gmail.com; Rasul: UCL and IFS, i.rasul@ucl.ac.uk.

1 Introduction

Economic ideas and concepts shape society through their impact on business, government and the media [Fourcade *et al.* 2015, Maesse *et al.* 2022]. Distinct features of economic methodology have enabled economists to tackle a widening array of subject matter, with a source of strength for economics being its diversity of subfields and the rise of empiricism, which has also led to economics increasingly influencing research in other disciplines [Lazear 2000, Angrist *et al.* 2020]. We study whether and how economists have leveraged this influence to contribute knowledge relevant to tackling a major social issue: large and persistent gaps in economic well-being between racial and ethnic groups. Using an array of newly matched bibliometric data from journal publications and working paper series, we provide novel evidence on the volume and content of race-related research that economists have produced.[1,2]

The spine of our analysis is built around the research content of academic journal publications in economics: these constitute the very subject matter of the discipline, laying the scientific foundation for economists to contribute to public debate. Publications are also the key metric along which career success is defined – they carry career rewards in terms of hiring, promotion, pay and tenure. We identify race-related research by taking an algorithmic approach to classify such work from a corpus of 225,000 publications in over 200 economics journals from 1960 to 2020.

The first step in our analysis is to identify race-related research. We fully recognize there is no definitive way to approach this given there can be reasonable differences in normative views on what such a body of work should constitute. Given the volume of publications considered, it is also infeasible to codify race-related research by hand. We thus take an algorithmic approach to classify publications as race-related, using keywords along two dimensions: (i) the racial or ethnic group being studied; and, (ii) the issue being studied, with a focus on five topic areas: discrimination, inequality, diversity, identity and historical studies. Examples of the 35 (case-insensitive) group keywords we use are *race*, *african-american, person of color*, and *ethnicity*. Examples of the 103 issue keywords we use are *discrimination*, *prejudice*, and *stereotype*. Our algorithm selects a publication as being race-related if: (i) at least one group keyword is in the title; or, (ii) at least one group keyword and at least one issue keyword are mentioned in the title or abstract (excluding

[1] Maesse *et al.* [2022] discuss four channels that economists have used to expand their influence: expert discourse, modalities of government linking policy and science, economists in academic, political and media networks, and economics as a social field. Lazear [2000] describes three features of economic methodology driving economic imperialism: the modelling of rational agents engaging in maximizing behavior subject to constraints, the importance of equilibrium, and the defined concept of efficiency. Angrist *et al.* [2020] document the rising influence of economics among other disciplines. They show economics is now the most widely cited social science in 7 of 16 disciplines, partly because different disciplines cite economics papers from different fields.

[2] The meaning of race and ethnicity have been extensively discussed in the social sciences. Ethnic differences apply across measurable group categories, and racial inequalities reflect racialized processes attributed to visible differences, with there being overlap in how these terms are used [Platt 2019]. While there is no biological basis for dividing people into 'races', race retains a social meaning. Throughout, for expositional ease, we refer to race-related research as work relevant for understanding racial and ethnic inequalities in well-being. We come back to this issue in Section 2.1 below.

the last line of the abstract).

Applying this algorithm to our corpus of publications, we reveal the following new facts on the volume and content of race-related research in economics. From nearly zero race-related publications in the early 1960s, the share of race-related publications rose to a peak of 2.9% in the mid 1970s, fell to 1.6% from the mid 1980s to the mid 1990s, and has risen steadily thereafter to around 2% today. This represents a cumulative body of knowledge of just over 4000 race-related publications in economics from 1960 to 2020. Accounting for changing journal influence over time, the *AER*-weighted share of race-related publications shows a more rapid rise, more than doubling since the mid-1990s. Hence although the share of publications studying issues of race has remained relatively flat since the mid 1990s, the prominence of such work has risen since the mid 1990s.

As a benchmark, we apply our algorithm to publications in sociology, noting that we likely under count race-related research in sociology given our use of economics-focused topic keywords. We find that: (i) in each year between 1960 and 2020, sociology journals have published a greater share of race-related research than economics journals – throughout the 2010s at least 12% of sociology publications have been race-related; (ii) more than 500 race-related articles have been published annually in sociology journals in the most recent years, and the cumulative number of race-related articles in sociology from 1960 to 2020 is $14,718$, more than three times the cumulative number in economics.

We zoom in on patterns of race-related research in the top-5 general interest journals in economics given these represent what is considered of broad interest to the discipline, reflect views of leading scholars, editors and referees, and are among the most highly cited publications [Heckman and Moktan 2020]. We document a major uptick in the share of race-related publications in the top-5 in the late 1990s, that has continued since – driven by increased numbers of such publications in the *AER*, *QJE* and *JPE*. In contrast, *Econometrica* and the *Review of Economic Studies* have each cumulatively published 17 race-related articles between 1960 and 2020 – relative to 5142 articles being published across those two journals since 1960.

Examining the evolution of content in race-related research, we find that in the 1970s race-related publications divided almost equally between those studying non-specific minority groups (so using keywords such as *ethnic minority* or *non-white*) and those focused on Black groups. By the 2010s the share of publications studying non-specific groups had risen to 75%, while those studying Black groups had fallen to 20%. Research studying Latinx groups emerged in the 1980s, but still only 3% of all race-related publications in the most recent years study this group. The study of other groups – including Asians and Native Americans – still comprises less than 2% of all race-related research. On topics, there has been a steady decline in the share of race-related research on discrimination since the 1980s, with a rise in the share of studies of identity.

To examine an earlier stage of the research production process, we apply our algorithm to NBER and CEPR working papers (WPs) posted over the last few decades. Around 3% of NBER WPs have been race-related, but the shares vary tremendously across fields. We document a concentration of race-related research into a few fields, and with such work being largely absent from many others. *Macroeconomics* (JEL Code E) – the most prominent field in the NBER series with 12% of all WPs – has the lowest share of race-related WPs (.35%). We do find some indication of positive time trends in the production of race-related research in fields in which it remains most scarce – such as *Macroeconomics* and *International Economics*. Fields with among the highest shares of race-related WPs are *Labor Economics* (J), *Urban Economics* (R) and *Economic History* (N). These each have at least 7% of WPs being race-related – three times the disciplinary average. For CEPR WPs, around 2% have been race-related over the 2000s, with similar patterns across fields being observed as for the NBER series.

Our work sheds new light on the ability of academic economists to scientifically contribute to public debates on racial justice and the causes, consequences and solutions to persistent economic inequalities between racial and ethnic groups. A preliminary version of the algorithm was described in our earlier work [Advani *et al.* 2024], that focused on constructing the aggregate time series of race-related work in economics, comparing this aggregate series to other disciplines, and using the *Social Science Prediction Platform* to examine whether economists were aware of trends in the publication of race-related research. In this paper, we refine the algorithm to describe the content of race-related research in economics in far more detail, for both published work and in prominent working paper series.

By placing the subject matter of economic research at the heart of our analysis, we add to a nascent literature classifying race-related research in economics, that does so either in terms of a specific area, such as discrimination [Bohren *et al.* 2025] or with regards to a specific journal, such as the *AER* [Horpendahl and Kling 2020]. In contrast, we take a disciplinary-wide perspective spanning a 60 year horizon of publications, to understand the content and publication outcomes of race-related research in economics. Our approach is thus more aligned to work describing corpi of work in economics [Angrist *et al.* 2017, 2020, Kleven 2018, Currie *et al.* 2020], or the representation of minorities in a large corpus of books [Adukia *et al.* 2023].[3]

Rather than take a normative stance, we aim to provide positive evidence that usefully informs normative questions on how economists can contribute to important societal debates. Throughout our analysis we discuss avenues for future research. We build on this in our conclusions by identifying areas of race-related research that might be relatively understudied in economics, and discussing interrelated issues such as the publications process, the allocation of research funds enabling the pursuit of race-related research, and the relationship between the production of race-related research and the entry of minorities into the economics academy.

The paper is organized as follows. Section 2 follows Advani *et al.* [2024] in describing how

[3] Bohren *et al.* [2025] study the miscategorization of types of discrimination in economics research. They find that between 1990-2018, 10 economics journals (including the top-5), published 105 empirical papers focused on topics such as discrimination, bias and disparities. Our algorithm identifies a broader set of topics for race-related research. Horpendahl and Kling [2020] document the rise in articles addressing issues of race in the *AER* (and in the *AEA Papers and Proceedings*) from 1991 to 2019.

we identify race-related publications. Section 3 documents the volume and content of this body of work in the last six decades of economics publishing, and Section 4 does the same for working papers. Section 5 draws together our findings to discuss implications for the discipline. The Appendix details data sources and robustness checks.

2 Identifying Race-Related Research

2.1 Defining Race-Related Research

Race and ethnicity are socially constructed categories used to classify humans based on perceived physical differences (race), particularly skin color, and based on shared cultural practices, language and ancestry (ethnicity) [Erikson 2010, Platt 2019]. Although there is little biological support for discrete racial or ethnic groupings [Lewontin 1972, Winther 2021], there is evidence across the social sciences for differences in average outcomes across racial and ethnic groups, as overviewed in Pager and Shepherd [2008], Chetty *et al.* [2020] and Mirza and Warwick [2024].

Conceptually, our target in this study is to identify work relevant for the study of the economic well-being of racial and ethnic groups, across countries and over time. Throughout, we refer to this body of work as 'race-related' research. To do this we do not seek to provide a single unifying definition for race or ethnicity. Instead we seek to capture the concept as it is used by economists and other social scientists. In Section 2.3 below we set out three measures, increasing in breadth, that seek to measure this use.

2.2 Corpus

Our corpus of academic publications is based primarily on the *JSTOR* database, using journals classified under the discipline of economics from 1960 through 2020. We fill gaps in this source using the *Web of Science* and *Scopus* databases. For each publication we extract information on the journal it is published in, publication date, its title and full text of the abstract. Our working sample covers $224,524$ publications in 230 economics journals.[4]

2.3 Algorithm

Given the volume of publications considered, it is infeasible to codify race-related research by hand. We thus take an automated approach using an algorithm to classify each publication. We do so using keywords along two dimensions: (i) the racial or ethnic group being studied; and, (ii)

[4] *JSTOR* has gaps in its publication series (especially in more recent years) and is missing some prominent journals. We fill these gaps using data from *Web of Science* (`webofknowledge.org`) and *Scopus* (`scopus.com`). The Data Appendix describes the procedure through which we access these databases and gives additional details on the construction of the corpus, including the fact that we do not use publications from the *AER Papers and Proceedings* as these lack abstracts.

the issue being studied. All keywords for classification purposes are considered in a case-insensitive manner and wildcards are used to capture different word spellings or forms. Examples of (case-insensitive) keywords for groups being studied are *race*, *african-american, person of color*, and *ethnicity*. Examples of issue keywords are *discrimination*, *prejudice*, and *stereotype*.[5]

Our algorithm selects a publication as race-related if: (i) at least one group keyword is in the title; or, (ii) at least one group keyword and at least one issue keyword are mentioned in the title or abstract – dropping the last sentence of the abstract to avoid false positives from publications that only mention race parenthetically; (iii) we declassify publications based on eliminating phrases such as *black market* and *horse race*.

We consider three clusters of keywords for group keywords, that gradually expand the racial/ethnic groups picked up by the algorithm. Beginning with the narrowest, Band 0 consists of 16 generic base keywords denoting racial and ethnic groups (e.g. *race*, *ethnic*, *under represented minority*). These non-specific keywords signify the study of minorities in general, rather than a particular group. Band 1 adds another 19 group base keywords relating to the main minority groups in the US (*African American*, *Latino* and *Native American*). Band 2 adds another 25 group keywords covering other 'hyphenated americans' (e.g. *Indian-American*, *Japanese-American*), less populous (in the US) ethnic groups (e.g. *arab*, *oriental*), and other minorities based on religious beliefs (e.g. *Muslim*, *Jewish*).

Both Band 1 and Band 2 are targeted at capturing the main racial and ethnic groups in the US, from where a large share of the corpus originates. To the extent that this misses academic research on racial/ethnic groups that are primarily not found in a US context (e.g. Algerians in France, Turks in Germany) this could potentially reduce the estimated levels of race related research, but is unlikely to affect the trends or the comparisons across disciplines that we make. Our core results combine the 35 group keywords in Band 0 and Band 1, and we show how results vary using narrower and broader bands. The full lexicon of group keywords is in Table A1.

Table A2 shows the lexicon of issue keywords: the 103 base keywords are designed to cover five topics: discrimination, inequality, diversity, identity, and historical issues. For example, discrimination includes *prejudice* and *stereotypes*, while inequality includes *disparity* and *disadvantage*.

Finally, we declassify publications containing any of the eliminated phrases in Table A3 in either the title or abstract. We derived this list of phrases iteratively by comparing the produced classification of race-related research against a hand-checked sample of publications.

Our algorithm is not designed to capture the universe of all race-related research and inevitably some gray areas remain (for example in topics related to immigration). However our algorithm is easily replicable, and can be extended to cover other groups and topics.[6]

---

[5] For example, the group wildcard rac* captures race, races, racial, racist, and racism. Wildcard issue keywords include discriminat*, prejudi*, and stereotyp*. The wildcards allow for both American and British English spellings.

[6] Our algorithm is not designed to capture two classes of work that could still be relevant for the study of racial/ethnic inequalities. First, papers that do not mention group keywords but refer to, say, 'blue' and 'red' groups instead. Second, research that is not specifically about race but could potentially be applied to understand racial

Our algorithmic approach still leads to misclassification errors in the form of false negatives and false positives. To reduce the rate of false negatives (race-related publications that are missed by our algorithm), we are relatively inclusive in the construction of the lexicon. To avoid false positives, using the combination of group and issue keywords removes many instances of not race-related research that might otherwise match our lexicon patterns. Dropping the eliminated phrases before applying the algorithm reduces false positives, where a term, e.g. race, is used with a different meaning. Dropping the last sentence of abstracts before applying the algorithm reduces false positives by excluding papers where race/ethnicity is not the primary focus, but mentioned parenthetically, often as a piece of heterogeneity analysis or robustness check.

2.4 False Positives and False Negatives

To quantify potential rates of false negatives and false positives, we hand code publications as being race-related in a validation sample from our original corpus. We construct this validation sample by first extracting a complete list of publications mentioning a group keyword in their title or abstract (excluding the final sentence, and not considering topic keywords and eliminated phrases) from the top-5 general interest journals from 1960 to 2020. This comprises 179 publications, which we then manually classify as being race-related or not. We find 81% of them to actually be race-related. Around one in five publications that contain a group keyword, but where no other restrictions are applied, is therefore *not* race-related.[7]

We compare the hand-coded classification in the validation sample to that generated by our algorithm to compute rates of false positives and false negatives. Following this approach, the rate of false positives is:

$$\frac{\#\text{Non race-related publications labeled as race-related (False Positive)}}{\#\text{False Positive} + \#\text{True Negative}} = 15.2\%, \tag{1}$$

and the rate of false negatives is:

$$\frac{\#\text{Race-related publications labeled as not race-related (False Negative)}}{\#\text{False Negative} + \#\text{True Positive}} = 5.5\%. \tag{2}$$

Combining both forms of misclassification error, the implied ratio of true race-related research to

inequalities – for instance, minorities might be more impacted by minimum wages [Derenoncourt and Montialoux 2021], polices with urban biases [Cook and Logan 2020], or through distributional effects of monetary policy [Bartscher *et al.* 2022].

[7] An alternative approach to constructing a validation sample would be to take a random sample of publications and hand code them as race-related or not. We do not follow this approach because race-related research comprises a small fraction of all publications. Hence an infeasibly large sample would either require to be hand-coded, or inferred rates of false positives and false negatives would be very imprecise.

identified race-related research is:

$$\frac{\#\text{Race-related publications}}{\#\text{Publications labeled as race-related}} = \frac{\#\text{False Negative} + \#\text{True Positive}}{\#\text{False Positive} + \#\text{True Positive}} = 102\%, \qquad (3)$$

To apply these rates of false positives and negatives to our full corpus of publications we need to assume: (i) no race-related research is conducted in these journals that excludes group keywords; and, (ii) misclassification rates found in the top-5 general interest journals apply equally to other journals. We underpin both assumptions in the next subsection, and later show how results vary by worst- and best-case scenarios for misclassification error.

2.5 Validation Using Chat-GPT

An alternative approach to classify whether publications are race-related or not is to use Chat GPT-3.5, a Generative Large Language Model created by OpenAI. We do so using the same validation sample described above, and then compare GPT's classification to our algorithm's. The Appendix describes the GPT prompt used. We summarize the contrast in approaches in Panels A and B of Figure A1. The confusion matrices demonstrate: (i) both approaches classify publications at around 90% accuracy, as shown in the diagonal matrix entries; (ii) as shown in the off-diagonal entries, GPT tends to falsely assign more publications into the not race-related category (false negatives), while our algorithm produces a more equal number of false positives and negatives. Of the 18 misclassified research publications using GPT, our algorithm makes similar errors in four of them.

Given the similarity in performance between our algorithm and GPT, we use GPT to: (i) show that the classification of publications is not sensitive to using additional information from the introduction of papers (as well as the title and abstract); (ii) underpin the earlier assumptions needed to calculate rates of false positive and negatives in our corpus.[8]

[8] On (i), we select 44 publications from the validation sample across all publication years. We have to reduce the sample because most introductions of publications are contained in a separate PDF file on *JSTOR*, which presents challenges for easy access. Moreover, given our validation sample covers publications over a long time period and across journals, there is considerable variation in their structure. We manually extract introductions, disregarding tables and figures. This resulting classification using GPT based on title, abstract and introductions is shown in the confusion matrix in Panel C of Figure A1. The classifications coincide for 89% of publications, but GPT's classification still exhibits a higher rate of false negatives, even when incorporating additional information from introductions. On (ii), we use two approaches. First, we note that to apply the rates of false positives and negatives to our full corpus of publications we assumed no race-related research in the top-5 journals excludes group keywords from its title and abstract. To test this we take a random sample of publications from the top-5 journals from 1960 to 2020 that were not in our validation sample but have a group keyword in their title or abstract, and use GPT to identify race-related papers within this group: none of these publications are classified as race-related. Second, we compare the classification of race-related research in non top-5 journals between our algorithm and using GPT based on a second validation sample, taking 86 random articles from the non top-5 that include group keywords in their abstracts or titles. This comparison is shown in the confusion matrices in Panels D and E of Figure A1. Both approaches yield a similar classification.

2.6 Journal Weights

We construct counts of race-related research based on the classification of individual publications. These counts make no adjustment for the quality of journals that work is published in. Given our 60-year study period has witnessed changing journal influence, to consider both the quantity and quality of race-related research we adjust for journal quality using the journal weighting scheme employed by Angrist *et al.* [2020] in their study of the intermural influence of economics. Journal weights are given by the relative frequency with which the journal is cited by the top 'trunk' journal in the economics discipline: the *American Economic Review*. Hence the weight of journal j in year t is given by:

$$w_j^t = \frac{\#\text{Citations to journal } j \text{ by trunk journal in year } t}{\#\text{Citations to all journals in the same discipline by trunk journal in year } t}. \tag{4}$$

These time-varying weights capture the rise and fall of the importance of journals in our corpus over time. Following Kleven [2018] and Angrist *et al.* [2020], when presenting time series evidence we plot five year moving averages to smooth variation but still pick up trends.[9]

3 Race-Related Journal Publications

3.1 Aggregate Trends

Panel A of Figure 1 shows the time series of race-related publications in economics from 1960 to 2020. While there are close to zero race-related publications in the early 1960s, there is a rapid growth in the share of race-related publications through the 1960s, so that by the end of the decade, close to 2% of all publications in economics were race-related – out of a total of $224,524$ publications in economics. The share rises to a peak of 2.9% in the mid 1970s, falls to 1.5% by the mid 1990s, and rises slightly thereafter to around 2% today. Panel B shows the corresponding number of publications: there is a steadily rising number of race-related publications each year, amounting to almost 120 publications annually from 2010 onwards. In 2020, the cumulative number of race-related publications in economics since 1960 stands at 4211.[10]

Panel C repeats the analysis using *AER*-weighted publications. Accounting for journal influence, the solid line shows the weighted-share of race-related publications replicates the pattern of rising race-related publications from the 1960s to mid 1970s and a decline until the mid 1990s.

[9] From the 1980s onwards, the set of journals in our sample is relatively stable. It is not the case that progressively higher or lower ranked journals over time are selected into the corpus. All economics journals that are not covered in Angrist *et al.* [2020] are given a zero weight.

[10] The total number of annual publications across all economics journals has risen from 1000 in the mid 1970s to over 7000 in the mid 2010s. The lower coverage of *JSTOR, WoS* and *Scopus* in the most recent years explains the slight downturn in the number of race-related publications in Panel B – so the actual cumulative number of race-related publications in economics until 2020 is likely closer to 4500.

However the weighted-share reveals a more rapid rise in race-related publications since the mid-1990s – a trend masked in Panel A when we do not account for journal quality. This is not because of changing weights for journals where race-related research is published, but rather because race-related research has been published in higher quality journals over time. To see this, the dashed line in Panel C fixes journal weights to their 2020 values, and shows similar trends since the 1980s as when we allow for time-varying journal weights.

Panel D shows the weighted number of race-related publications has risen steadily over time, reaching the equivalent of 1.5 *AER* publications annually since the mid-2010s. Hence although the share of all publications studying issues of race has remained relatively flat since the mid 1990s, the prominence of such work – as measured by the journals in it published in – has risen steadily since the mid 1990s.

Figure A2 confirms these time trends in the share and weighted-share of race-related research are similar when: (i) we drop the requirement of not using the final sentence of abstracts in our algorithm; (ii) we use alternative Bands for the group keywords. For example, utilizing the broadest set of all 60 group keywords (Bands 0, 1 and 2) we see that the share of race-related research lies around 2.5% since the 2000s.

3.2 Journals

Top-5 Journals It is useful to separately consider publications in the top-5 general interest economics journals: the *AER*, *Econometrica*, *QJE*, *JPE* and *Review of Economic Studies*. These represent what is considered of broad interest to the discipline, reflects views of leading editors and referees, and are among the most highly cited publications [Heckman and Moktan 2020]. Panel A of Figure 2 shows the share of race-related research in top-5 journals has lagged behind publication shares in other economics journals for most of our study period. However, we see a major uptick in the share of race-related publications in the top-5 from the mid-1990s, that has continued since. As a result, since the early 2000s there has been a convergence in the share of race-related research in top-5 and non top-5 journals.

Panel B shows the rise of race-related research in the top-5 journals has been driven by the *AER*, *QJE* and *JPE*. In nearly all years since 1960, the *QJE* has published a higher number of race-related articles than other top-5 journals, although there has been a rapid rise in the number of race-related publications in the *AER* since 2010, overtaking the *QJE* in the most recent years. Collectively, the *AER*, *QJE* and *JPE* have published 155 race-related articles since 1960, corresponding to 1.3% of all their publications. In contrast, *Econometrica* and the *Review of Economic Studies* have each cumulatively published fewer than 20 race-related articles since 1960, corresponding to .3% of all their publications over the same time period.[11]

[11] The rise in publications addressing issues of race in the *AER* (and *AEA Papers and Proceedings*) from 1991 to 2019 is documented in Horpendahl and Kling [2020]. Their classification of such articles is based initially on those

All Journals Figure A3 shows the extent to which a broader set of journals have published race-related research in the 2000s. General interest journals are at the top of the figure, with more specialized journals then ranked below by the share of race-related articles published from 1960 to 2020. Three points are of note. First, among general interest journals, the *Review of Economics and Statistics* has published the highest share of race-related articles in the 2000s, at just under 4%, with the *QJE* publishing the highest share in the 2010s, at just over 4%. Second, there is not much to suggest that European-based general interest journals – such as the *Review of Economic Studies*, *JEEA* or the *Economic Journal* – have published higher shares of race-related articles. Third, outside of general interest journals, the *JUE*, *JHR* and *JoLE* have all been traditional field-journal homes to race-related research. In the last decade, *EDCC*, the *Journal of Legal Studies* and the *AEJ: Applied* are some of the journals having 5% or more of their articles being race-related, double the disciplinary average.

Review of Black Political Economy The most prominent economics journal specialized in race-related research is the *Review of Black Political Economy* (*RBPE*) – that was indeed launched in response to concerns that mainstream economics journals were not open to publishing research on the political economy of race [Alexis *et al.* 2008]. The bars at the foot of Figure A3 show race-related publication rates for the *RBPE* – measured on a different x-axis scale to all other journals. Our algorithm assigns 77% of publications in the *RBPE* to be race-related in the 2000s, and 67% to be race-related in the 2010s – an order of magnitude higher than the other journals.[12]

To give a sense of the kinds of publication identified by our algorithm, Figure A4 shows some of the most highly cited race-related publications in economics, where we use citations/year to better compare publications over time. The upper part of the Figure picks out some of the most highly cited work from non top-5 journals, and the lower part does the same for top-5 publications. We noted before the uptick in race-related research in the top-5 around the mid 1990s – and this corresponds to when some of the most highly cited work was published. Two papers published in the *QJE* by George Borjas stand out – on intergenerational mobility and human capital externalities. In more recent times, the work by Bertrand and Mullanaithan using an audit study to understand labor market discrimination is among the most highly cited race-related publications.

with *JEL* codes *J15* and *J71*, and then hand-checking each identified article. They report 56 articles on race were published in the *AER* and *AER P&P* between 1991 to 2018 – closely matching our estimate only for the *AER* over this period of 48.

[12] This implies that up to one third of *RBPE* papers are not race-related. This reflects that our algorithm is relatively conservative, requiring both mentions of race groups together with race-related topics.

3.3 Groups and Topics Studied

Groups For each publication the algorithm classifies as race-related, we can use the group keywords to pinpoint which minority groups are studied.[13] Panel A of Figure 3 shows that in every year since 1975, the majority of race-related publications have covered non-specific groups (those in band 0 in our algorithm): today such work comprises around 75% of all race-related research in economics. While close to 50% of race-related publications in economics during the 1970s focused on Black groups, by the 2010s this had fallen to less than 20%. Research studying Latinx groups emerged in the 1980s, yet still only 3% of all race-related research in the most recent years has focused on this group. Research on other groups – including Asians or Native Americans – remains almost non-existent, that might be due to a lack of data, or inconsistent coding of disaggregated data for such groups. Hence, current trends still reflect a long-standing concern about the lack of research on smaller minorities (and on interactions between minority groups) [Altonji and Blank 1999].

Topics We can use the topic keywords used by our algorithm to pinpoint the issue studied, divided into the five areas covered: discrimination, inequality, diversity, identity, and historical issues.[14] Panel B of Figure 3 shows the majority of race-related research relates to inequality, comprising 59% of all race-related publications today. There has been a steady decline in the share of race-related research on discrimination since the 1980s with a rise in the share of studies on identity. Race-related historic research has increased slightly over time, while the share of race-related publications examining issues of diversity has remained relatively constant over our long study period.

3.4 Benchmarks

While we make no normative claim as to whether the share of race-related articles in economics is too high or too low, it remains useful to construct some benchmark comparisons. We approach this in three ways: making comparisons within discipline, across disciplines, and also comparing to the mention of race in fiction and the news.

Within Discipline: Using Machine Learning to Classify Topics We use machine learning to classify study areas in our corpus, and use this to measure the extent to which race and ethnicity has been studied relative to other identified topics in economics. To do so, we use Latent

[13] Publications can of course be classified as studying multiple groups: this occurs in 6.5% of cases (Black and Latinx groups are the groups most commonly studied together). When a publication mentions more than one group, we split the publication equally across groups.

[14] The algorithm identifies when publications study multiple topics: this occurs in 28% of cases. The most commonly combined topics are discrimination and diversity, while identity tends to be studied separately. When a publication mentions more than one topic, we split the weight of the publication equally across topics.

Dirichlet Allocation (LDA) modeling as an analytical tool to uncover hidden thematic structures of publications from their abstracts. LDA is a latent factor model that probabilistically assigns words in a document to one or more underlying topics, which are represented as distributions over words. LDA iteratively uncovers these hidden topics and their prevalence in each document. We would like the LDA model to learn the broadest possible set of social science topics, that may or may not be prevalent in economics. Hence, we build a broad corpus of 500,000 publications across social science disciplines: economics, sociology, political science, law, management, public policy, and history. Our benchmark model then identifies 30 distinct topics that are studied in this body of work. Figure A5 displays word clouds for the topics generated and we label each of the topics as shown in the lower part of Figure A5. One of the identified topics – Topic 8 – is labelled as 'race and ethnicity' where the most prominent keywords comprising this topic including *group*, *black*, *ethnic*, *white* and *racial*.[15]

We then use the LDA model to predict the topic of any given publication in our corpus of economics publications only. To be clear, while the topics are distinct, a given publication can be probabalistically assigned to multiple topics.[16]

Panel A of Figure 4 shows the distribution of LDA topics across publications in economics: 1% of them are classified under the race and ethnicity topic, which is less prevalent than nearly all other topics. Panel B shows the time series of the share of race and ethnicity topic papers, overlaid with the time series for the share of race-related research that our algorithm identified. Two points are of note. First, in most years since the early 1970s, our algorithm identifies a higher share of race-related research than is picked out by the LDA model – this might be due to the rich set of keywords used in our algorithm, and changing trends in groups and topics studied over time as highlighted in Figure 3. Second, trends in both time series both show an uptick in research on race/ethnicity in the mid 1990s that has continued until today. This is another reassuring validation of the real information picked up by our algorithmic classification of race-related research.[17]

[15] To implement LDA modeling, we use the `Gensim` library in Python, using its built-in tools to perform pre-processing tasks, such as removing punctuation and eliminating stopwords. During this process, we construct a dictionary, which is refined by excluding the 5% most and least frequent words. To determine the optimal number of topics, we analyze a combination of coherence score and perplexity measures across models with different numbers of topics. We also manually inspect the word distribution for each topic in each model. Two of the LDA topics are comprised of non-English words because some publications in English language journals still include non-English terms. The LDA model identifies these as separate groups, and we refer to them as Miscellaneous topics. These topics still also include English language words

[16] Figure A6 provides a granular view of topic co-occurrence. This analysis shows that the 'race and ethnicity' topic exhibits connections with the topics of 'labor market', 'health studies', 'social science theory', and 'income growth and inequality'.

[17] Cihak *et al.* [2020] also compare the extent to which topics have been studied in economics. They compile data on every race-related publication in the top-10 economics journals for the last decade, although the set of keywords they use to identify race-related research is far narrower than ours. They report .2% of those 7,920 articles cover issues of race, racial inequality, and racism. This is lower than what they find in terms of the share of articles devoted to monetary policy (7.4%), income distribution (2%), poverty (1.4%) and gender (.8%).

Across Disciplines: Comparison to Sociology An alternative benchmark is to compare across disciplines, as discussed in Advani *et al.* [2024]. We do so by applying our algorithm to publications in sociology, noting that we likely under count race-related research in sociology given our use of economics-focused topic keywords. We find that: (i) in each year between 1960 and 2020, sociology journals have published a greater share of race-related research than economics journals – throughout the 2010s at least 12% of sociology publications have been race-related; (ii) more than 500 race-related articles have been published annually in sociology journals in the most recent years, and the cumulative number of race-related articles in sociology from 1960 to 2020 is 14,718, more than three times the cumulative number in economics; (iii) accounting for journal influence, the weighted-share of race-related research has risen from the mid 1990s, reaching the equivalent of seven or more *ASR* publications annually since 2010.

News and Fiction One question is whether the trends observes in economics articles are reflecting changes in academic economics, or rather are mechanically reflecting changes in the broader public discourse. To address this issue and provide a third benchmark, we use the Corpus of Historical American English (COHA) [Davies 2021]. COHA contains 141,000 documents (475 million words of text) from the 1820s through the 2010s and is designed to be balanced by genre (fiction, magazine, news, academic, and other non-fiction) over time. COHA is over fifty times larger than other comparable historical corpora of American English. We take the news and fiction genres, available from the 1860s though the 2010s and apply our algorithm to this corpus.[18]

Figure 5 plots the shares of race/ethnicity-related 256-word segments for fiction (green) and news (orange) by decade. In fiction, race issues are consistently a low share, appearing in less than 2% of text snippets for the entire time period. For news, there is a striking pattern with two spikes in race-related mentions: one at the beginning of the sample in the 1860s (during the civil war), and another in the 1960s (during the civil rights movement). Neither time series looks entirely like the trend in academic economics articles (blue), suggesting that the trends we observe in economics are not fully reflecting broader changes in societal discourse.

[18] Unlike the article abstracts corpus, the documents in COHA are of highly varying lengths and do not have titles that can be consistently assigned. Therefore, we split the documents into 256-word segments and ran our abstract-matching algorithm on the segments (requiring both a group and topic keyword match), ignoring titles. The resulting corpus for matching is about 1.2 million short text snippets, adding up to about 300 million words. While our algorithm is designed for economics articles, we found through manual inspection of outputs that the algorithm was working well on COHA text snippets.

4 Race-Related Working Papers

4.1 Corpus

We now investigate an earlier stage of the research process: the production of working papers (WPs). We build a corpus of the two most prominent WP series in economics, from the NBER and CEPR. Our sample covers $22,056$ NBER WPs first posted from 1974 to 2015, and $10,306$ CEPR WPs first posted from 1984 to 2015. The Data Appendix further details each series. We apply our algorithm to this corpus to establish the extent to which WPs are race-related.[19]

4.2 Aggregate Trends

Panel A of Figure 6 shows the time series of race-related NBER and CEPR WPs. In each year, NBER WPs are more likely to be race-related than CEPR WPs. While both series show upward trends in the share of race-related WPs, the gap between them has remained relatively constant over time. Over the last decade, 3.5% of NBER WPs have been race-related, while the corresponding figure for CEPR WPs is closer to 2%. Comparing these to discipline wide time trends in journal publications, we see that: (i) NBER WPs have nearly always had a higher share of race-related research than journal publications in any given year since the 1980s (either across all journals or among the top-5); (ii) the uptick in the share of race-related research in the NBER and CEPR WP series – that occurs in the early 1990s – slightly predates the uptick previously documented in the weighted-share of such journal publications, that was noticeable from the mid-1990s.

4.3 Fields

Both series contain *JEL* classifications for each WP, unlike journal publications where such classifications are not consistently available. This allows us to examine the differential production of race-related WPs across subfields of economics. Aggregating over the available time period for each series, Panel B of Figure 6 shows for each high-level *JEL* code: (i) the share of all WPs which list this *JEL* code (gray bars); (ii) the share of WPs that are race-related for each *JEL* code (blue/red bars). In both panels, we order *JEL* codes in increasing shares of race-related research among NBER WPs.[20]

[19] The NBER and CEPR represent prominent networks for US- and Europe-based research economists respectively. The NBER was founded in 1920, currently has around 1600 members organized around 20 research programs and 13 working groups. Each year the NBER holds around 125 meetings and publishes over 1100 WPs. The CEPR was founded in 1983, has over 1700 members in 14 research programmes, organizes around 250 meetings and publishes over 1000 discussion papers annually.

[20] When a working paper has multiple *JEL* codes, we split the assignment equally across all listed codes. The ranking across fields helps to further validate our algorithm. For example, we see that our algorithm classifies fewer than 3% of NBER WPs in *Economic Development* (O) as being race-related.

For both WP series, we observe a concentration of race-related research into a few fields, with such work being largely absent from many other fields.[21]

Starting with NBER WPs, the share of race-related working papers vary from .35% to 13% across *JEL* codes. *Macroeconomics* (JEL Code E) has the lowest share of race-related WPs from 1973 to 2019 (.35%). *Financial Economics* (G) is the next field where race-related working papers are most scarce. These two fields are among the most prominent in the NBER series, comprising nearly a quarter of all WPs. Hence the low rates of race-related WPs in these fields has knock-on effects for the aggregate share of all NBER WPs that are race-related.

The field with the highest share of race-related research is *Other Special Topics* (Z), at 13%. This is not surprising given that stratification economics is listed under this category.[22] The pattern across other fields closely matches the field journals in economics that have published the highest shares of race-related research: the *JUE*, *JHR, JoLE* and *EEH* – the other fields with the highest shares of race-related WPs are *Labor Economics* (J), *Urban Economics* (R) and *Economic History* (N). These each have at least 7% of WPs being race-related, three times the disciplinary average. There is some gap to the next field, *Public Economics* (H) – that has 3% of WPs being race-related. This is noteworthy given wealth inequalities across groups can be more extreme than for labor market outcomes [Darity and Mullen 2020, Mirza and Warwick 2024].[23]

The Relevance of Race-related Research Across Fields To narrow the interpretation of these field differences, we first consider whether they reflect that issues of race and ethnicity are just far less relevant for core research questions in some fields, or whether such issues are harder to study given data constraints.

We start to examine the issue by first restricting attention to WPs that have at least one of the topic keywords (Table A2) in their title and/or abstract. For example, this includes all WPs studying inequality, just not necessarily through the lens of racial/ethnic differentials. Panel A of Figure 7 then repeats the analysis by fields for this subset of WPs. Although the share of race-related WPs increases in each field, from 2% in *Macroeconomics* to 24% in *Urban Economics* in the NBER series, overall the ranking across fields in the share of race-related WPs remains

[21] These patterns across fields are reminiscent of the concentration of women in economics into subfields at the time of PhD graduation[Fortin *et al.* 2021]. Using data on NBER SI submissions by program, Chari and Goldsmith-Pinkham [2018] find that over 2016-8, the share of women authors was 18% in programs related to finance and macro, and 31% in programs related to applied micro.

[22] Stratification economics views inter-group inequality as the long term result of historic factors. The field draws on economics, sociology, and social psychology and was crystallized in Darity [2005]. It was assigned *JEL* category Z13 (*Economic Sociology, Economic Anthropology, Language, Social and Economic Stratification*) and is cross-listed with D31 (*Personal Income, Wealth and Their Distributions*).

[23] For the US, Darity and Mullen [2020] document that the median net worth of Whites in the bottom 20% of the income distribution is higher than the median net worth of all Black households. For the UK, Mirza and Warwick [2024] document that all ethnic minority groups are under-represented in the top 20% of the wealth distribution. Types of wealth also differ dramatically: while the median White British household has £115,000 in property wealth, the median Black household has none, while Pakistani and Indian households have median property wealth greater than for White British households.

largely unchanged. For example, macroeconomics papers that have at least one topic keyword in their title and/or abstract constitute 7% of all NBER WPs, and 2% of this subset are identified to be race-related. Among CEPR WPs the same patterns emerge when we restrict WPs to those that have at least one of the topic keywords in their title and/or abstract.

A second explanation for field differences in the study of race-related issues is data constraints [Adjaye-Gbewonyo *et al.* 2014, Cronin *et al.* 2023, Heller *et al.* 2024]. For research questions focused on individuals/households, survey data on race and ethnicity can sometimes be lacking or aggregated, or racial-ethnic gaps in well-being are not studied due to small sample sizes. To the extent that such constraints are gradually being eased over time, we might pick this up in the share of race-related research WPs by field and decade. Panel B of Figure 7 shows how the production of race-related WPs has changed over the last three decades (still limited to those WPs that mention at least one topic keyword). Among the NBER series, we see steady increases in the share of such work over time in fields such as *International Economics*, *Industrial Organization*, and *Economic History*. This suggests data constraints might slowly being eased to allow for the study of group differences in some fields of economics, although patterns by decade are less clear within field for the CEPR WP series.

5 Discussion

Economists often view themselves as having an important role to play in informing societal debates [Fourcade *et al.* 2015, Spriggs 2020, Maesse *et al.* 2022]. This should include debates on racial and ethnic gaps in economic well-being. Our ability to do so depends on the scientific foundation of race-related research that economists have collectively produced. We quantify the volume and content of such work over the last six decades. We document that since 1960 less than 2% of economics publications have been race-related, with an uptick in work in the mid 1990s. There have also been changes over time in the groups and topics studied within race-related work. Across subfields of economics, we find that race-related research is prominent in a few fields, with such work being largely absent from many others. In comparison to sociology, our approach suggests the discipline has something like a 20-30 year lag in the production of race-related research. This difference might just reflect valid disciplinary differences in subject matter. If this is not the case however, then it is useful to see what evidence can be brought to bear to understand the economics-specific factors that can narrow this gap. We discuss: (i) understudied race-related topic areas in economics; (ii) the role of economics journals and funding; (iii) the selection and retention of minority faculty in the economics academy.

5.1 Understudied Race-related Topics in Economics

As emphasized throughout, our algorithm is designed to identify race-related research partly on the basis of topic keywords orientated towards economic issues. To shed light on the kinds of race-related work that might be missed, we examine topics that are studied in journals or disciplines focused on race and ethnicity, but that our algorithm does not pick up, as well as in publications in a discipline closely related to economics – sociology. To be clear, the explicit inclusion of these topics into our algorithm might well lead to more false positives, but they are still informative of race-related topics that are relatively understudied in economics.

Figure 8 shows the topics studied from three sources: (i) the *Review of Black Political Economy*; (ii) journals in the discipline of African American and American Indian Studies; (iii) sociology. Rather than using our two-part algorithm match (based on group and topic keywords) to assign 'is race-related' we assume that all the articles from *RBPE* or the ethnic studies journals (Panels A and B) are race-related, and we assume that sociology journal articles (Panel C) are race-related if they mention just a group keyword (regardless of whether there is a topic match). In each case we show the share of race-related research in the five topics picked up by our algorithm, and a residual category – labelled 'other topics' for the *RBPE* and labelled 'not economics orientated' for the other two sources. The right hand side of each Figure then picks out example keywords from this residual category.[24]

Examining the topic coverage in all *RBPE* articles (Panel A), we see that our economics topics tend to cover most of the associated articles, with the share of other topics comprising around 10% of publications in the last decade. Example keywords from this work include *inner-city*, *minority-owned*, *enterprises*, *finance* and *married*. This is of note because we saw earlier in the context of NBER working papers, *Financial Economics* (G) is a field where race-related working papers are scarce. For journals in African American and American Indian Studies we actually find the economics-focused topics comprise the majority of race-related research since the mid 1980s. Since the 1990s the share of not economics orientated publications has steadily risen to comprise around 30% of all publications in this discipline. Example keywords from this work include *curriculum*, *languages*, *teachers* and *art*. Finally, among publications in sociology with a least one group keyword in the title or abstract, around 75% relate to topics not captured by our algorithm. This share has remained relatively stable over our entire study period. Example keywords from this work include *couples*, *church*, *adolescents*, *husbands*, *happiness*, *personality*, *art*, and *religiosity*.

These findings complement existing work emphasizing that the lens through which economists study discrimination could be broadened [Small and Pager 2020, Spriggs 2020]. Others have argued a lack of recognition for minority economists has led to their perspective on mainstream

[24] The corpus of *RBPE* publications considered are from 1977, publications in any journal in the discipline of African American and American Indian Studies since 1986 (as classified by *JSTOR*), and race-related research in sociology from 1960 onwards.

topics being ignored – an example being within the economics of crime the lack of attention given to racial profiling, mass incarceration, and police use of force [Mason *et al.* 2022], or the design and impacts of public policy more broadly [Francis *et al.* 2022]. Stratification economics, that emphasizes competition and collaboration across groups to attain and maintain relative position in social hierarchies, has yet to enter mainstream areas of economic study [Darity 2022]. Finally, earlier work has suggested the discipline move away from the idea of race as exogenous, and build on the idea that racial self-classification may be endogenous to economic outcomes [Saperstein and Penner 2010, Charles and Guryan 2011, Spriggs 2020], or might reflect choices of identity [Darity *et al.* 2006, Akerlof and Kranton 2000].

5.2 Journals and Funding

The incentives academics have to produce race-related research partly depends on how such papers fare in the publications process. This is a topic we start to unpack in ongoing work, where we link NBER and CEPR working papers to their published versions (if published) and study how paths to publication vary for race-related research compared to other work [Advani *et al.* 2025]. In that work we examine paths to publication for a corpus of 22,056 NBER working papers (WPs) posted from 1974 to 2015. We compare paths for race-related WPs to various counterfactual sets of WPs. We find that conditional on date of posting, *JEL* codes, WP characteristics and author affiliation dummies, race-related NBER WPs are not differentially likely to be published in any journal, they are not differentially likely to be published in a economics journal, nor do they differ in their publication lag in economics journals. Conditional on the WP being published in an economics journal, race-related NBER WPs are not published in journals of differential quality. Finally, on citations for published WPs (where citations accumulate over both WP and published versions), we find no difference in total citations between race-related and non race-related WPs – a result robust to controlling for journal fixed effects. All this points to such work not being held to a lower standard of publication.

Another avenue that can cause a differential selection of race-related work is the even earlier process of research funding. Cruz-Castro *et al.* [2022] review the evidence on gender, race and ethnicity differentials in research funding in the US and Europe – so focusing on how the identity of individual researchers impacts funding outcomes (not the subject matter of funding proposals). While they find that gender gaps in funding have closed at the NSF, NIH and in Europe, for the US minorities remain far less likely to receive research funding than White individuals. There remain multiple possible explanations for this such as differences in applicant behavior, research productivity, peer review processes, and other inherent biases. Irrespective of the cause, the result might be the differential selection into the production of race-related working papers vis-à-vis non race-related work. This remains an important topic for future work.

5.3 Faculty

Differential outcomes in the process of research funding based on the race and ethnicity of researchers can lead to differential production of race-related research *if* there is a link between the racial/ethnic identity of researchers and the areas they research. While such links have been documented in the context of inventors [Einiö *et al.* 2023] and medical researchers [Dossi 2024], our findings lead naturally to the study of the relationship between the production of race-related research and the entry of minorities into the economics academy. The lack of entry of minorities – the pipeline problem – is well recognized, and this is a margin along which many of the initiatives of economic associations, such as the *AER*, *EEA* and *RES* are heavily directed [Bayer and Rouse 2016, Bayer *et al.* 2020].[25]

In ongoing work, we study the nexus between the racial and ethnic identity of individual faculty and the production of race-related research. We thus extend a line of work linking the subject matter of economic research and the subject matter studied by Black economists [Price and Sharpe 2020], most notably in relation to stratification economics and the economics of race, but also Black economists' distinct approaches and contributions to the study of areas of public policy – as discussed in a recent *JEL* symposium [Darity 2022, Francis *et al.* 2022, Mason *et al.* 2022]. Moreover, the documented concentration of race-related research in some fields might have knock-on effects for the formation of professional networks, that are so important for career progression in academia [Fourcade *et al.* 2015, Zinovyeva and Bagues 2015].[26]

Tackling this wider agenda, of understudied topics in economics, the publishing process, funding, and the entry and retention of minority faculty in the economics academy, can potentially all contribute in important ways to underpin economists' contribution to societal debates on the causes and consequences of large and persistent gaps in economic well-being across racial and ethnic groups.

[25] The under representation of minorities in the profession has long been recognized – the *AEA* established its *Committee on the Status of Minority Groups in the Economics Profession* over 50 years ago. More recently the *AEA*, *EEA* and *RES* have all been taken steps to promote inclusivity by establishing new initiatives, formalizing codes of conduct and surveying members. The 2019 *AEA* member survey found that 3% of economists identified as Black, 47% of Black respondents reported experiences of discrimination, and only 45% of all respondents (regardless of race) believed non-White economists are respected.

[26] Mason *et al.* [2005] shows that papers with at least one Black author are more likely to report a finding of racial discrimination than papers with no Black authors. Freeman and Huang [2015] show that papers by ethnically diverse coauthor teams receive more citations than papers written by same ethnic group teams. The link between selection and research has been documented more along lines of gender than race. For example, in a survey of 143 *AEA* members, men and women economists are found to differ on views on economic outcomes and policies, even after controlling for PhD vintage and employment type [May *et al.* 2014].

A Appendix

A.1 Data Sources

JSTOR We use *JSTOR* as our primary data source on academic publications. To classify journals to disciplines, we use *JSTOR*'s disciplinary definition for each journal except when Angrist *et al.* [2020] provide an alternative classification. For journals classified to be cross-disciplinary, we assign equal weights to journal publications across disciplines. For each *JSTOR* publication, we extract metadata such as the *JSTOR* ID, journal name, year, abstracts, and titles. *JEL* codes are not available for the metadata from *JSTOR* (while some publications do contain *JEL* codes, they are embedded within the PDF versions of publications and not published on *JSTOR*'s website).

Web of Science and Scopus The *JSTOR* publication series still has gaps, especially in recent years. These gaps relate to missing data for some journals in particular years, and also because the *JSTOR* publications series does not include certain journals, such as the *Journal of Public Economics* and the *Review of Black Political Economy*. To address these issues, we utilize data from the *Web of Science* (*WOS*) and *Scopus* publications series. The *WOS* dataset consists of articles published between 1970 and 2015. By employing ISSN numbers, we map articles to their respective disciplines. Within the *WOS* series there can still be missing abstracts (because the *WOS* API does not provide abstracts), that we then fill in using information from *Scopus*. We employ the Scopus API to retrieve publications from a specific journal and publication year. A matching process is then subsequently conducted to find similar titles between the *WOS* and *Scopus* series. For each ISSN and year combination, we compare all possible pairs of indices from the two datasets using fuzzy string matching. Pairs with a partial ratio above a threshold (.95) were considered matches and stored in a dictionary. We utilized the matching dictionary to map indices from the *WOS* dataset to their corresponding indices in the *Scopus* dataset. Abstracts from the *Scopus* dataset were added to the *WOS* dataset based on the matched indices.

Throughout, we exclude publications that have missing abstracts and that cannot be recovered using *Scopus* data. Additionally, our corpus does not include publications from *Paper and Proceedings* series, as these typically do not contain abstracts. We also note that certain journals, such as the *Economic Journal* prior to 1994, did not require abstracts and so those journal-years are not included in our final corpus.

Foreign Languages We drop all non-English language parts of abstracts, using an automated Python language detection method. We retain those journals that have paper titles and abstracts in both English and another language because our algorithm can be applied to such papers.

Deduplication Since we construct our corpus by combining different data sources, we face an issue that multiple versions of the same publication might exist. To address this issue we compare the titles of articles within each discipline, journal, and year to identify potential duplicates. The procedure utilizes string similarity measures to calculate the pairwise distance between article titles. If the similarity exceeds a predefined threshold (.90), the publications are considered duplicates. All duplicates are dropped from the analysis.[27]

Cleaning Abstracts Scraped abstracts from *JSTOR, WOS* and *Scopus* have varying formats. A challenge is that abstracts often contain copyright sentences or additional information. As our algorithm to identify race-related work relies on penalization based on the last sentence of abstracts, it is crucial to ensure abstracts are cleaned and standardized across platforms. Through manual inspection, we identified approximately twenty different patterns of copyright sentences used. Using string matching algorithms, we cleaned all abstracts for analysis to remove such extraneous information.

Working Papers For the NBER series, we construct a corpus starting from $28,206$ NBER WPs first posted from 1974 to 2019. Dropping articles published as WPs after 2015 for publication delay considerations, we are left with $22,056$ observations. For the CEPR series, we construct our corpus based on WPs first posted from 1984 to 2019. We start with $15,137$ WPs, and dropping articles published from WPs after 2015, we are left with $10,306$ WPs. WPs and their metadata are scraped using a publicly available API. In a few cases, multiple versions of WPs are posted over time. We use the first posted versions throughout, and also verify that almost no WPs change classification from race-related to non race-related (or *vice versa*) across posted versions.

When a WP lists multiple *JEL* codes, we split the assignment equally across codes. We omit WPs with no *JEL* classification and *JEL Code Y* (*Miscellaneous Categories*) because it is not represented in the NBER corpus and is associated with only eight papers in the CEPR series, among which none are race-related. 4722 (589) NBER (CEPR) papers do have not *JEL* codes.

A.2 Robustness Checks

Chat GPT The system prompt given to GPT-3.5 Turbo was based on our experience with similar tasks. The benchmark prompt was: *you are a helpful assistant. Determine in the most accurate way if the academic paper is related to race and/or ethnicity based on the given title and abstract. Respond with one word: Yes, No, or Unclear.* We set the temperature parameter to zero to ensure replicability. The output of GPT's classification was manually reviewed to check for hallucinations from the language model (i.e. where GPT provides answers that are not among our identified choices). We did not encounter any hallucinations. In one instance GPT provided

[27] While incorporating information from abstracts could in principle help to identify duplicates more reliably, in practice we see that abstracts themselves often undergo substantial revisions between versions, so relying on them could introduce noise rather than precision.

an answer that included an additional explanation of its choice: 'Unclear. The paper discusses various economic topics, but it is not clear if it specifically relates to race and/or ethnicity.'

References

[1] ADJAYE-GBEWONYO.D, R.A.BEDNARCZYK, R.L.DAVIS AND S.B.OMER (2014) "Using the Bayesian Improved Surname Geocoding Method (BISG) to Create a Working Classification of Race and Ethnicity in a Diverse Managed Care Population: A Validation Study," *Health Services Research* 49: 268-82.

[2] ADUKIA.A, A.EBLE, E.HARRISON, H.B.RUNESHA AND T.SZASZ (2023) "What We Teach About Race and Gender: Representation in Images and Text of Children's Books," *Quarterly Journal of Economics* 138: 2225-85.

[3] ADVANI.A, E.ASH, D.CAI AND I.RASUL (2024) "Race Related Research in Economics and Other Social Sciences," *Econometric Society Monograph*, forthcoming.

[4] ADVANI.A, E.ASH, D.CAI AND I.RASUL (2025) Paths to Publication for Race-related Research in Economics, mimeo, UCL.

[5] AKERLOF.G.A AND R.E.KRANTON (2000) "Economics and Identity," *Quarterly Journal of Economics* 115: 715-53.

[6] ALEXIS.M, T.SPRATLEN AND C.Z.WILSON (2008) "Robert Browne and the Caucus of Black Economists," *Review of Black Political Economy* 35: 61-6.

[7] ALTONJI.J.G AND R.M.BLANK (1999) "Race and Gender in the Labor Market," in O.C.Ashenfelter and D.Card (eds.) *Handbook of Labor Economics* Vol. 3C, North Holland.

[8] ANGRIST.J, P.AZOULAY, G.ELLISON, R.HILL AND S.F.LU (2017) "Economic Research Evolves: Fields and Styles," *AER Papers & Proceedings* 107: 293-7.

[9] ANGRIST.J, P.AZOULAY, G.ELLISON, R.HILL AND S.F.LU (2020) "Inside Job or Deep Impact? Using Extramural Citations to Assess Economic Scholarship," *Journal of Economic Literature* 58: 3-52.

[10] BARTSCHER.A, M.KUHN, M.SCHULARICK AND P.WACHTEL (2022) "Monetary Policy and Racial Inequality," *Brookings Papers on Economic Activity* Spring: 1-47.

[11] BAYER.A AND C.E.ROUSE (2016) "Diversity in the Economics Profession: A New Attack on an Old Problem," *Journal of Economic Perspectives* 30: 221-42.

[12] BAYER.A, G.A.HOOVER AND E.WASHINGTON (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: 193-219.

[13] BAYER.A, G.BRUICH, R.CHETTY AND A.HOUSIAUX (2020) Expanding and Diversifying the Pool of Undergraduates who Study Economics: Insights from a New Introductory Course at Harvard, NBER WP 26961.

[14] BOHREN.A, K.HAGGAG, A.IMAS AND D.POPE (2025) "Inaccurate Statistical Discrimination," *Review of Economics and Statistics* 107: 605-20.

[15] CHARI.A AND P.GOLDSMITH-PINKHAM (2018) Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute, mimeo, Yale.

[16] CHARLES.K AND J.GURYAN (2011) "Studying Discrimination: Fundamental Challenges and Recent Progress," *Annual Review of Economics* 3: 479-511.

[17] CHETTY.R, N.HENDREN, M.R.JONES AND S.R.PORTER (2020) "Race and Economic Opportunity in the United States: An Intergenerational Perspective," *Quarterly Journal of Economics* 135: 711-83.

[18] CIHAK.M, M.MLACHILA AND R.SAHAY (2020) "Race in Economics," *Finance and Economics* 57: 36-8.

[19] COLLINS.S.M (2000) "Minority Groups in the Economics Profession," *Journal of Economic Perspectives* 14: 133-48.

[20] COOK.L.D AND T.LOGAN (2020) Racial Inequality, Economists for Inclusive Prosperity, Policy Brief 27.

[21] CRONIN.J-A, P.DE FILIPPES AND R.FISHER (2023) Tax Expenditures by Race and Hispanic Ethnicity: An Application of the U.S. Treasury Department's Race and Hispanic Ethnicity Imputation, Office of Tax Analysis WP122.

[22] CRUZ-CASTRO.L, D.K.GINTHER AND L.SANZ-MENENDEZ (2022) Gender and Underrepresented Minority Differences in Research Funding, NBER WP 30107.

[23] CURRIE.J, H.KLEVEN AND E.ZWIERS (2020) "Technology and Big Data Are Changing Economics: Mining Text to Track Methods," *AEA Papers and Proceedings* 110: 42-8.

[24] DARITY.W.A JR (2005) "Stratification Economics: The Role of Intergroup Inequality," *Journal of Economics and Finance* 29: 144-53.

[25] DARITY.W.A JR, P.L.MASON AND J.B.STEWART (2006) "The Economics of Identity: The Origin and Persistence of Racial Identity Norms," *Journal of Economic Behavior & Organization* 60: 283-305.

[26] DARITY.W.A (2022) "Position and Possessions: Stratification Economics and Intergroup Inequality," *Journal of Economic Literature* 60: 400-26.

[27] DARITY.W.A AND K.MULLEN (2020) *From Here to Equality: Reparations for Black Americans in the Twenty-First Century*, Chapel Hill, NC: UNC Press.

[28] DAVIES.M (2021) *Corpus of Historical American English (COHA)*, https://doi.org/10.25346/S6/6I8JL1, UCLA Dataverse, V12.

[29] DERENONCOURT.E AND C.MONTIALOUX (2021) "Minimum Wages and Racial Inequality," *Quarterly Journal of Economics* 136: 169-228.

[30] DOSSI.G (2024) Race and Science, mimeo, LSE.

[31] EINIO.E, J.FENG AND X.JARAVEL (2023) Social Push and the Direction of Innovation, mimeo LSE.

[32] ELLISON.G (2013) "How Does the Market Use Citation Data? The Hirsch Index in Economics," *American Economic Journal: Applied Economics* 5: 63-90.

[33] ERIKSEN.T.H (2010) *Ethnicity and Nationalism: Anthropological Perspectives*, Pluto Books.

[34] FORTIN.N, T. LEMIEUX AND M.REHAVI (2021) "Gender Differences in Field of Specialization and Placement Outcomes among Ph.D. in Economics," *AEA Papers & Proceedings* 111: 74-9.

[35] FOURCADE.M, E.OLLION AND Y.ALGAN (2015) "The Superiority of Economists," *Journal of Economic Perspectives* 29: 89-114.

[36] FRANCIS.D.V, B.L.HARDY AND D.JONES (2022) "Black Economists on Race and Policy: Contributions to Education, Poverty and Mobility, and Public Finance," *Journal of Economic Literature* 60: 454-93.

[37] FREEMAN.R.B AND W.HUANG (2015) "Collaborating with People Like Me: Ethnic Coauthorship within the United States," *Journal of Labor Economics* 33: S289-318.

[38] HECKMAN.J.J AND S.MOKTAN (2020) "Publishing and Promotion in Economics: The Tyranny of the Top Five," *Journal of Economic Literature* 58: 419-70.

[39] HELLER.R, S.MOK, J.PEARCE, J.ROTHBAUM (2024) Using Multiple Data Sources to Learn About the Race and Ethnicity of Taxpayers, mimeo, CBO.

[40] HENGEL.E (2022) "Publishing While Female," *Economic Journal* 132: 2951-91.

[41] HORPENDAHL.J AND A.KLING (2020) "Gender, Race and Ethnicity, and Inequality Research in the *American Economic Review* and the *American Economic Association's Conference Papers*," *Econ Journal Watch* 17: 338-49.

[42] KLEEMANS.M AND R.L.THORNTON (2021) "Who Belongs? The Determinants of Selective Membership into the NBER," *AEA Papers & Proceedings* 111: 117-22.

[43] KLEVEN.H (2018) Language Trends in Public Economics, mimeo, Princeton University.

[44] LAZEAR.E (2000) "Economic Imperialism," *Quarterly Journal of Economics* 115: 99-146.

[45] LEWONTIN.R.C (1972) "The Apportionment of Human Diversity," in T.Dobzhansky, M.K.Hechtand W.CSteere (eds.) *Evolutionary Biology* Vol. 6, New York: Springer, pp 381-98.

[46] LOGAN.T.D AND S.L.MYERS JR (2020) "The Failure of Economics and the Marginalisation of Research on Race," in S.Galiani and U.Panizza (eds.) *Publishing and Measuring Sucess in Economics*, CEPR Press: London.

[47] LUNDBERG.S.J AND J.STEARNS (2019) "Women in Economics: Stalled Progress," *Journal of Economic Perspectives* 33: 3-22.

[48] MAESSE.J, S.PUHRINGER, T.ROSSIER AND P.BENZ (2022) *Power and Influence of Economists*, Routledge Frontiers of Political Economy: London and New York.

[49] MASON.P, S.MYERS JR. AND W.DARITY JR (2005) "Is There Racism in Economic Research?" *European Journal of Political Economy* 21: 755-61.

[50] MASON.P S.L.MYERS JR. AND M.SIMMS (2022) "Racial Isolation and Marginalization of Economic Research on Race and Crime," *Journal of Economic Literature* 60: 494-526.

[51] MAY.A.M, M.G.MCGARVEY AND R.WHAPLES (2014) "Are Disagreements Among Male and Female Economists Marginal at Best? A Survey of AEA Members and their Views on Economics and Economic Policy," *Contemporary Economic Policy* 32: 111-32.

[52] MIRZA.S AND R.WARWICK (2024) "Race and Ethnic Inequalities," *Oxford Open Economics* 3: i365-i452.

[53] OMEOKWE.A (2020) Economics Journals Faulted for Neglecting Studies on Race and Discrimination, Wall Street Journal, July 12th.

[54] PAGER.D AND H.SHEPHERD (2008) "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets," *Annual Review of Sociology* 34: 181-209.

[55] PLATT.L (2019) *Understanding Inequalities: Stratification and Difference* (Second edition), Polity.

[56] PRICE.G.N (2009) "The Problem of the 21st Century: Economics Faculty and the Color Line," *Journal of Socio-Economics* 38: 331-43.

[57] PRICE.G.N AND R.V.SHARPE (2020) "Is the Economics Knowledge Production Function Constrained by Race in the USA?" *Journal of the Knowledge Economy* 11: 614-29.

[58] SAPERSTEIN.A AND A.PENNER (2010) "The Race of a Criminal Record: How Incarceration Colors Racial Perception," *Social Problems* 57: 92-113.

[59] SLATER.R.B (2020) "The First Black Faculty Members at the Nation's Highest-Ranked Universities," *Journal of Blacks in Higher Education* 22: 97-106.

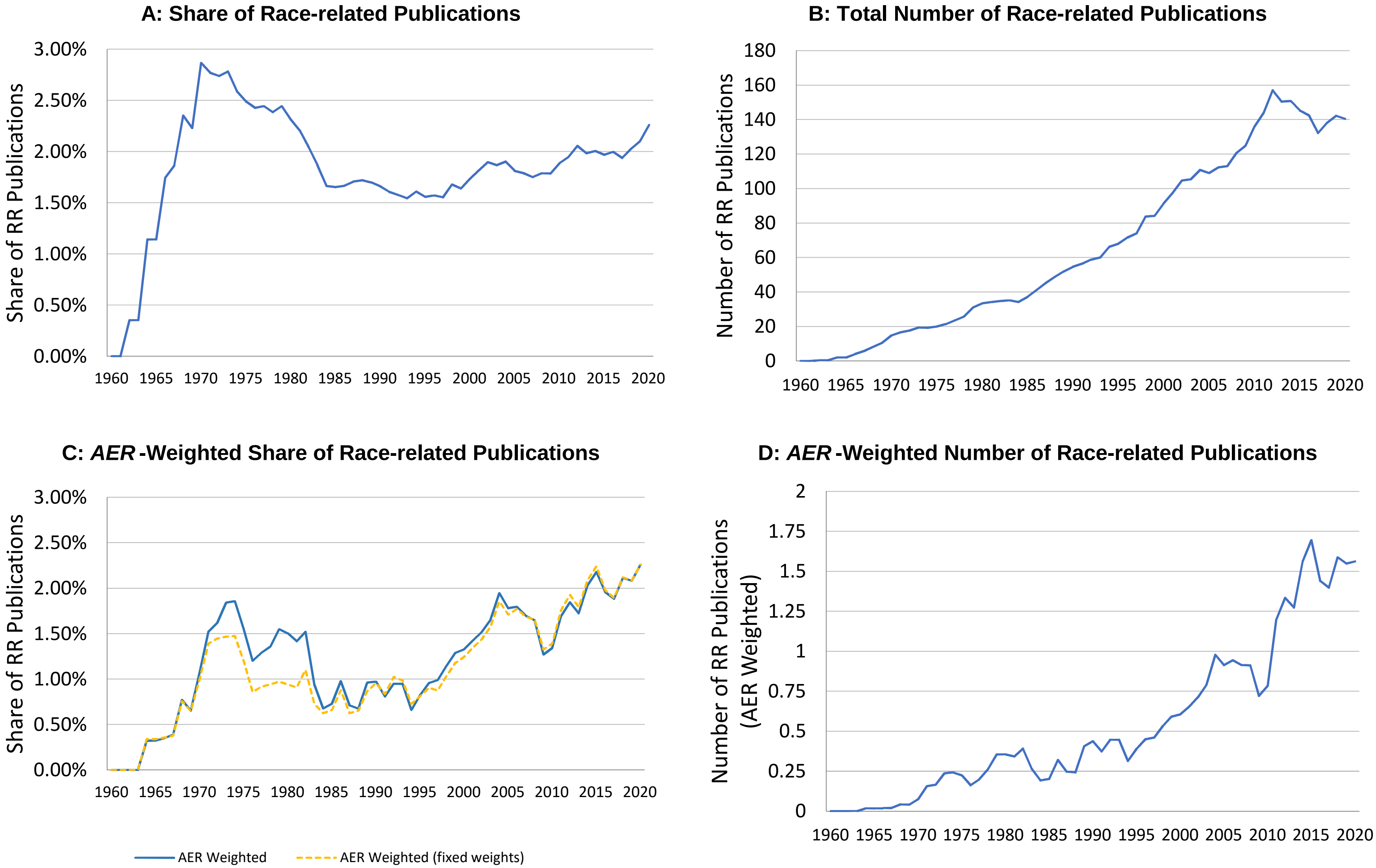
[60] SMALL.M.L AND D.PAGER (2020) "Sociological Perspectives on Racial Discrimiantion," *Journal of Economic Perspectives* 34: 49-67.

[61] SPRIGGS.W (2020) Is Now a Teachable Moment for Economists?, mimeo, Howard University.

[62] WINTHER.R.G (2021) "Lewontin (1972)," in L.Lorusso and R.G.Winther (eds.) *Remapping Race in a Global Context*, London: Routledge.

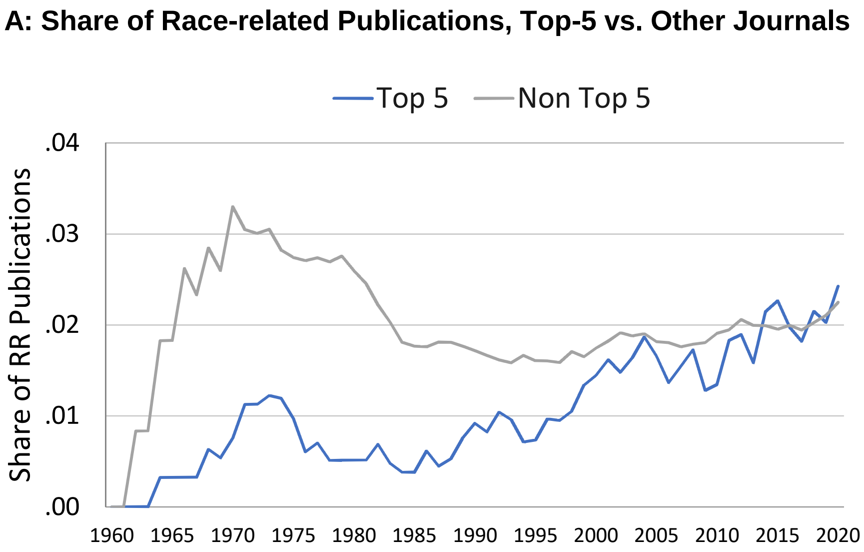
[63] ZINOVYEVA.N AND M.BAGUES (2015) "The Role of Connections in Academic Promotions," *American Economic Journal: Applied Economics* 7: 264-92.

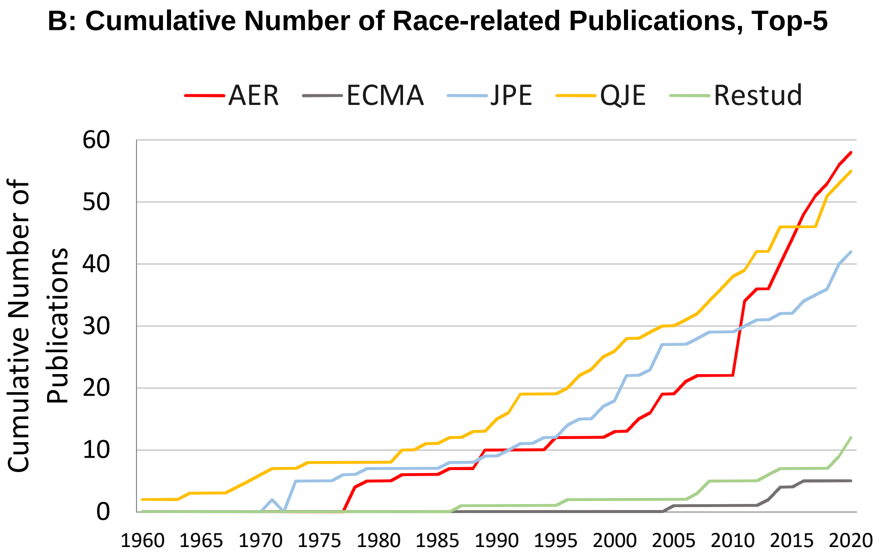
Figure 1: Race-related Publications in Economics, by Year



Notes: We use a corpus of publications in economics journals, based on data from *JSTOR*, *Web of Science* and *Scopus*. When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. We report five-year moving averages throughout. Panel A reports the share of total publications identified to be race-related by year of publication. Panel B reports the number of race-related publications by year of publication. Panels C and D report *AER*-weighted versions of Panels A and B, using the journal weights constructed in Angrist *et al.* [2020].

Figure 2: Race-related Publications, Top-5 Journals

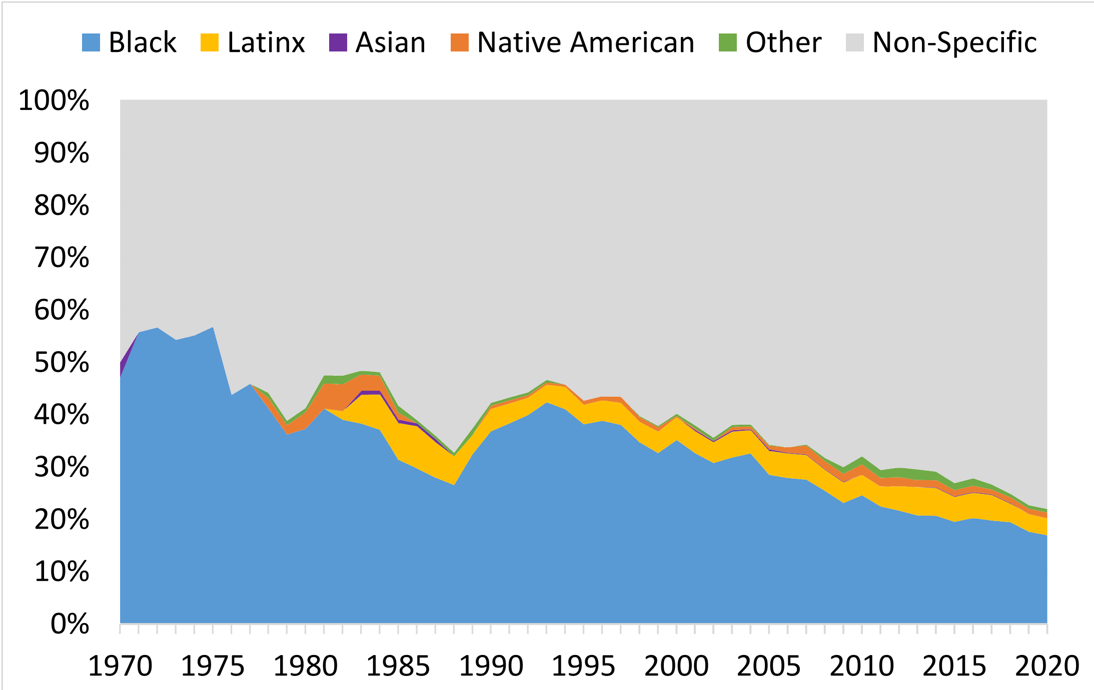




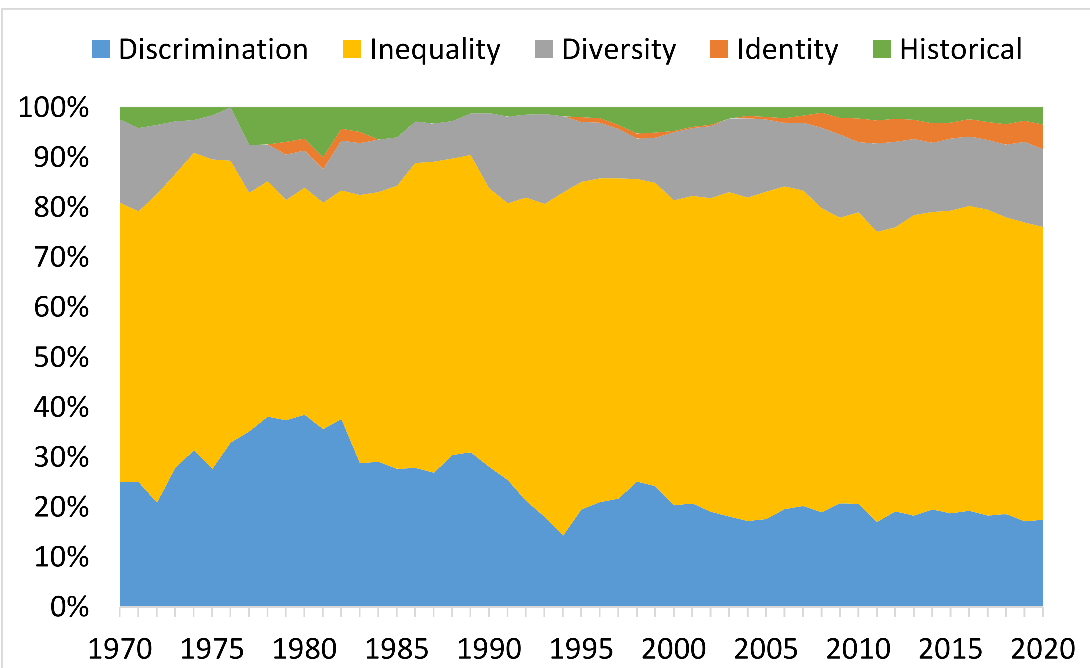
Notes: The top-5 general interest journals in economics are the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies*. For non-top 5 journals, we use a corpus of publications in economics journals, based on data from *JSTOR*, *Web of Science* and *Scopus*. When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. In Panel A we report five-year moving averages.

Figure 3: Race-related Publications in Economics

A. Groups Studied



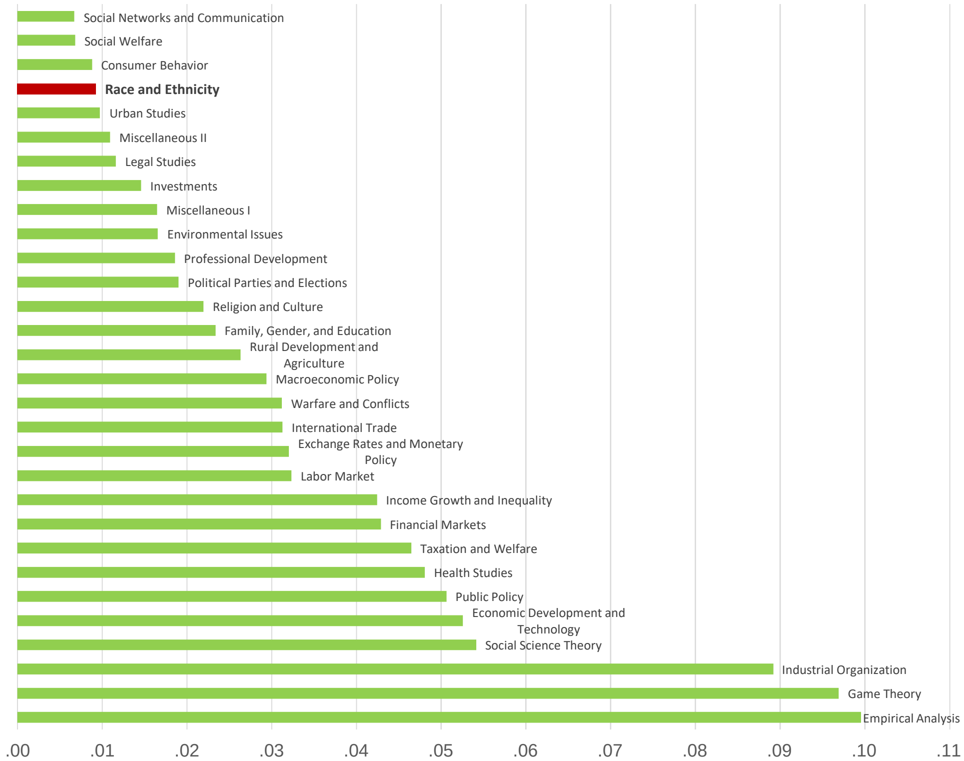
B. Topics Studied



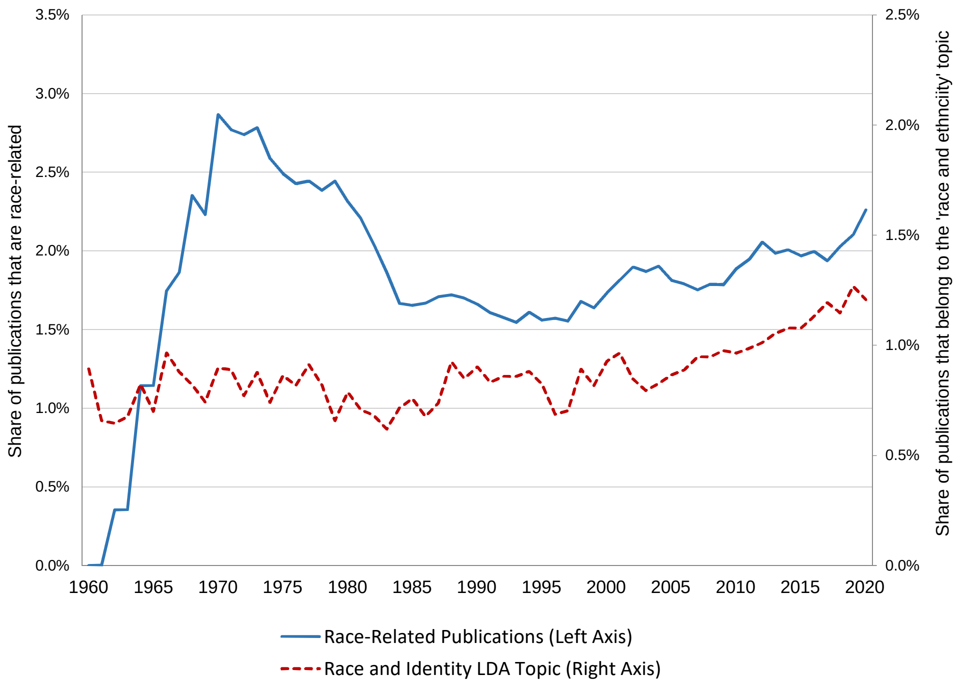
Notes: We use a corpus of publications in economics journals, based on data from *JSTOR*, *Web of Science* and *Scopus*. When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. We report five-year moving averages throughout. To construct the groups studies series in Panel A, for each year, we calculate the publications among all race-related ones that mention at least one group. When a publication mentions more than one group, we split the weight of the publication equally across those different groups. In Panel B, we make an analogous construction for publications that mention more than one race-related topic.

Figure 4: LDA Topic Model on Corpus of Economics Publications 1960-2020

A. Share of Publications by Topic

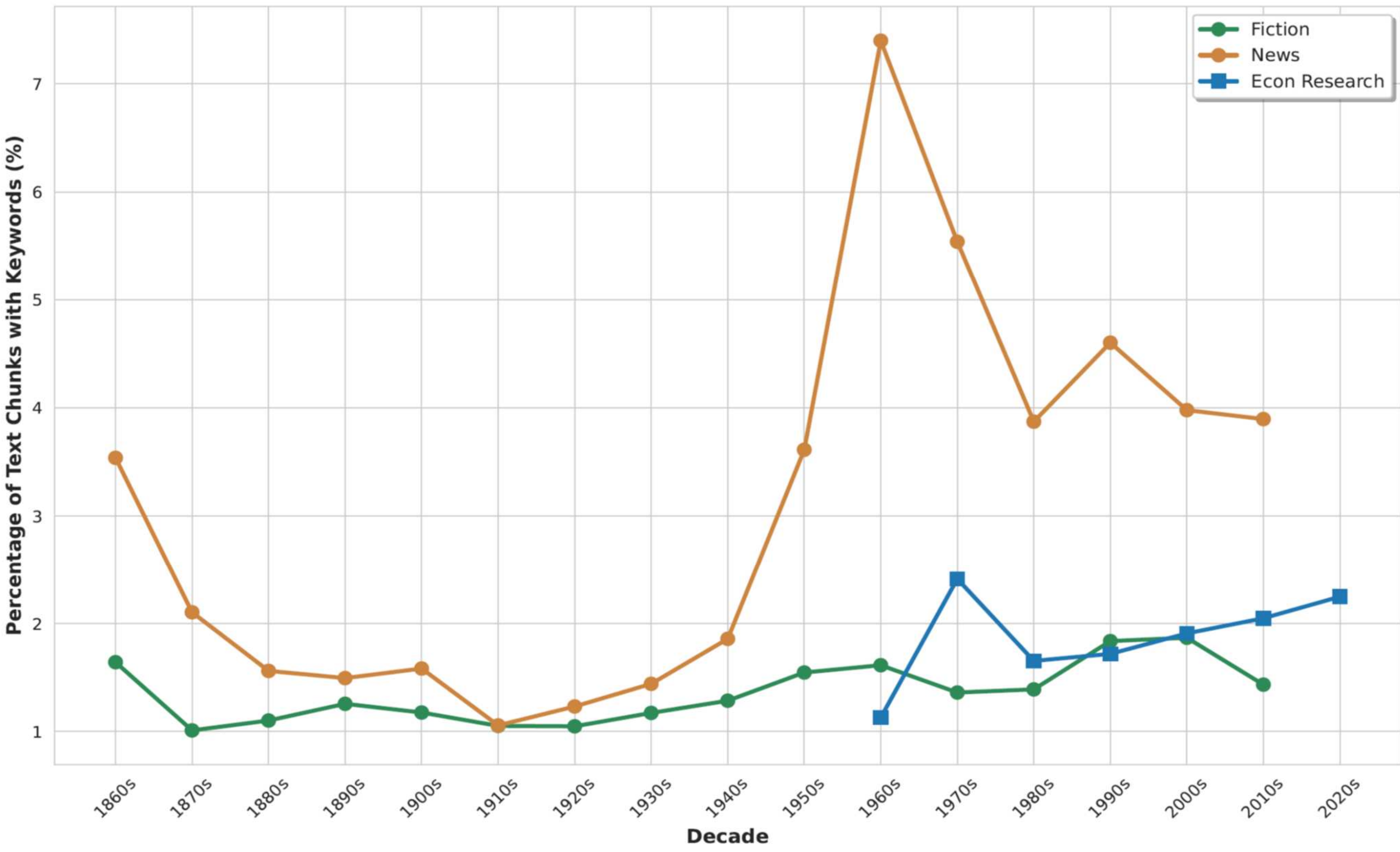


B. *AER*-Weighted Share of Race-related Publications and Share of Publications by Race and Identity Topic, by Year



Notes: Panel A shows the share of publications in economics by topic, based on the 30 topics derived from the LDA model. The model is run on a corpus of 493,972 publications in economics, sociology, political science, law, management, public policy and history, based on data from *JSTOR*, *Web of Science*, and *Scopus*. When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. In Panel A we then use predicted topic probabilities on our corpus of publications in economics from 1960. Each bar represents the average probability of a given topic across all economics publications in the corpus. As documents in an LDA model can belong to multiple topics simultaneously, these topics are not mutually exclusive at the individual article level. Panel B reports five-year moving averages for the share of all economics publications that are identified by our algorithm as race-related. This series is measured on the left-hand axis. On the right-hand axis we show the share of economics publications that are assigned the LDA topic of 'race and ethnicity'.

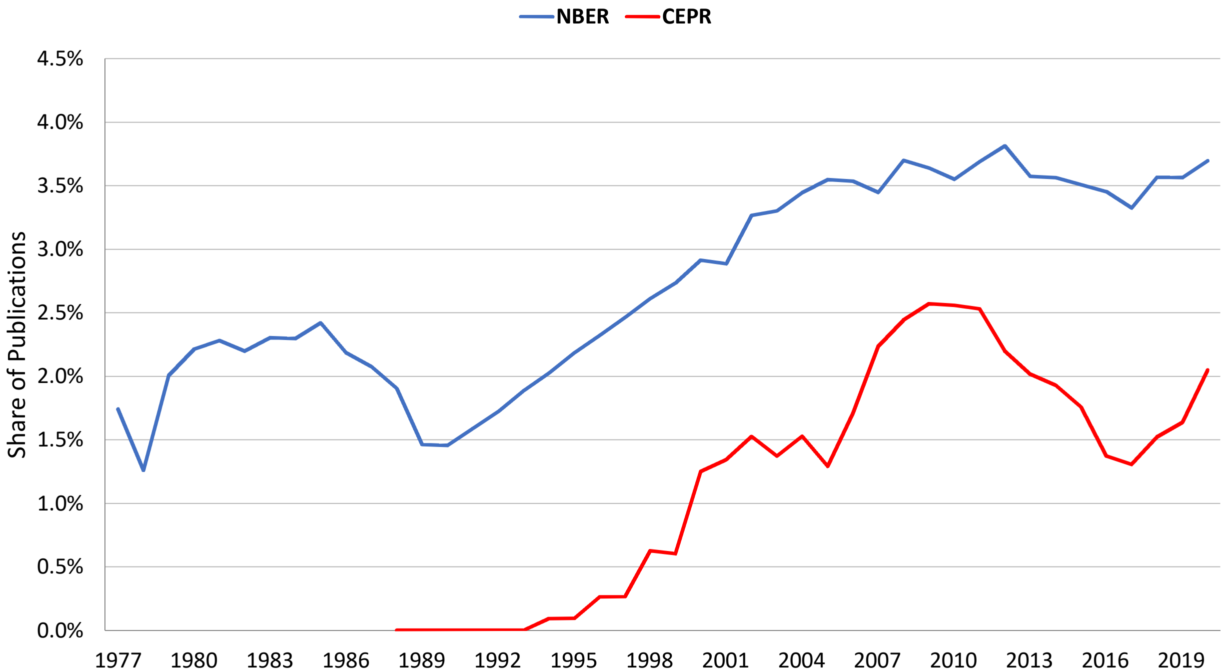
Figure 5: Text Snippets Mentioning Race/Ethnicity in Fiction and News, By Decade



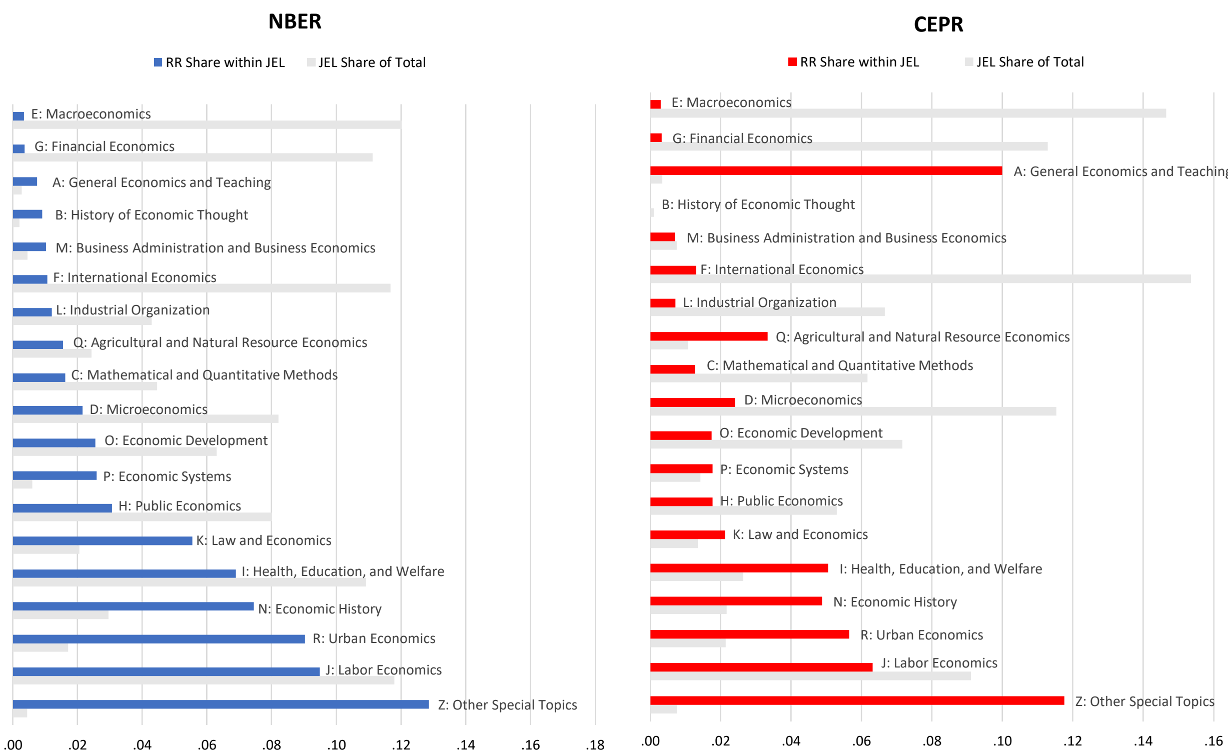
Notes: Each time series is based on the Corpus of Historical American English (COHA) that contains 141,000 documents (475 million words of text) from the 1820s through the 2010s and is designed to be balanced by genre (fiction, magazine, news, academic, and other non-fiction) over time. We split the documents into 256-word segments and run our abstract-matching algorithm on the segments (requiring both a group and topic keyword match), ignoring titles. We then plot the shares of race/ethnicity-related snippets for fiction (green) and news (orange) by decade.

Figure 6: NBER and CEPR Working Papers

A. Share of Race-related Working Papers



B. Share of Race-related Working Papers by JEL Category and JEL Share of Total

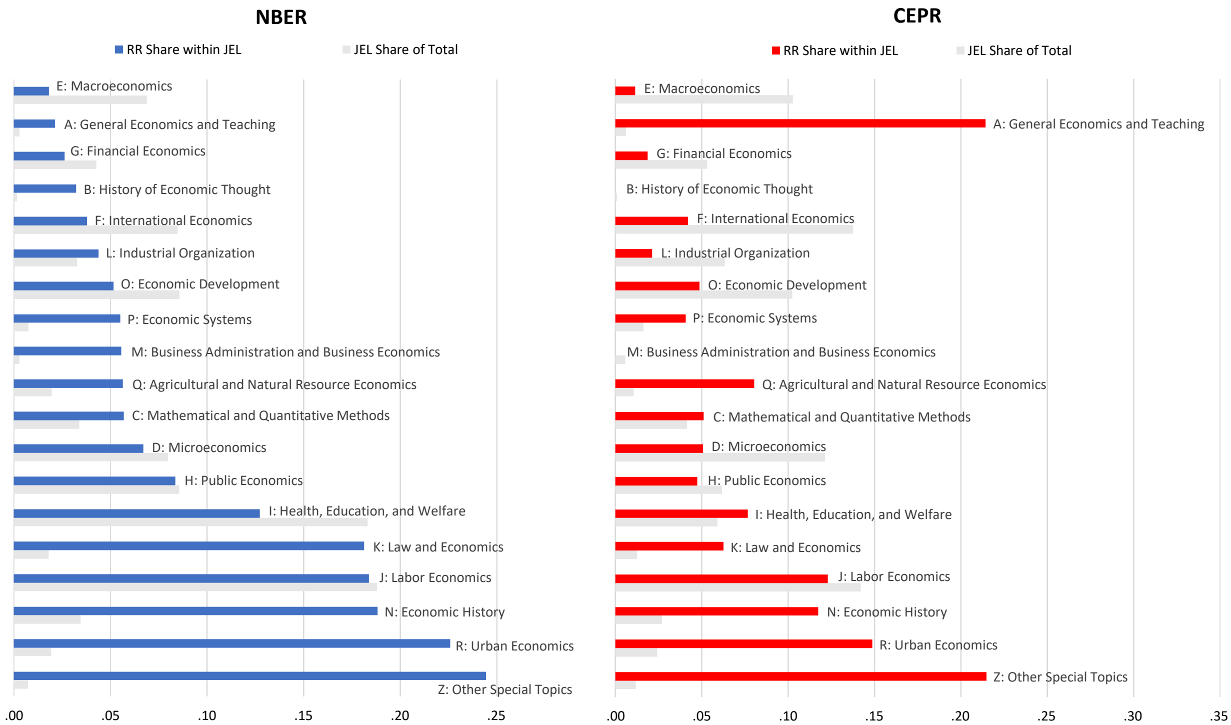


Notes: The sample is based on NBER working papers first released between 1974 and 2019, and CEPR working papers released between 1984 and 2019. Panel A shows the shares of working papers identified to be race-related in each series, in five-year moving averages - with the NBER (CEPR) series starting in 1977 (1987). Panel B shows the fraction of working papers that are identified to be race-related by JEL classification, as well as the share of all working papers in that series by JEL code. When a working paper has multiple JEL codes, we split the assignment article equally across all codes. We omit working papers with no JEL classification and JEL code Y, *Miscellaneous Categories* because this is not represented in the NBER corpus and is associated with eight papers in the CEPR series, among which none are race-related.

Figure 7: The Relevance of Race-related Research by Field

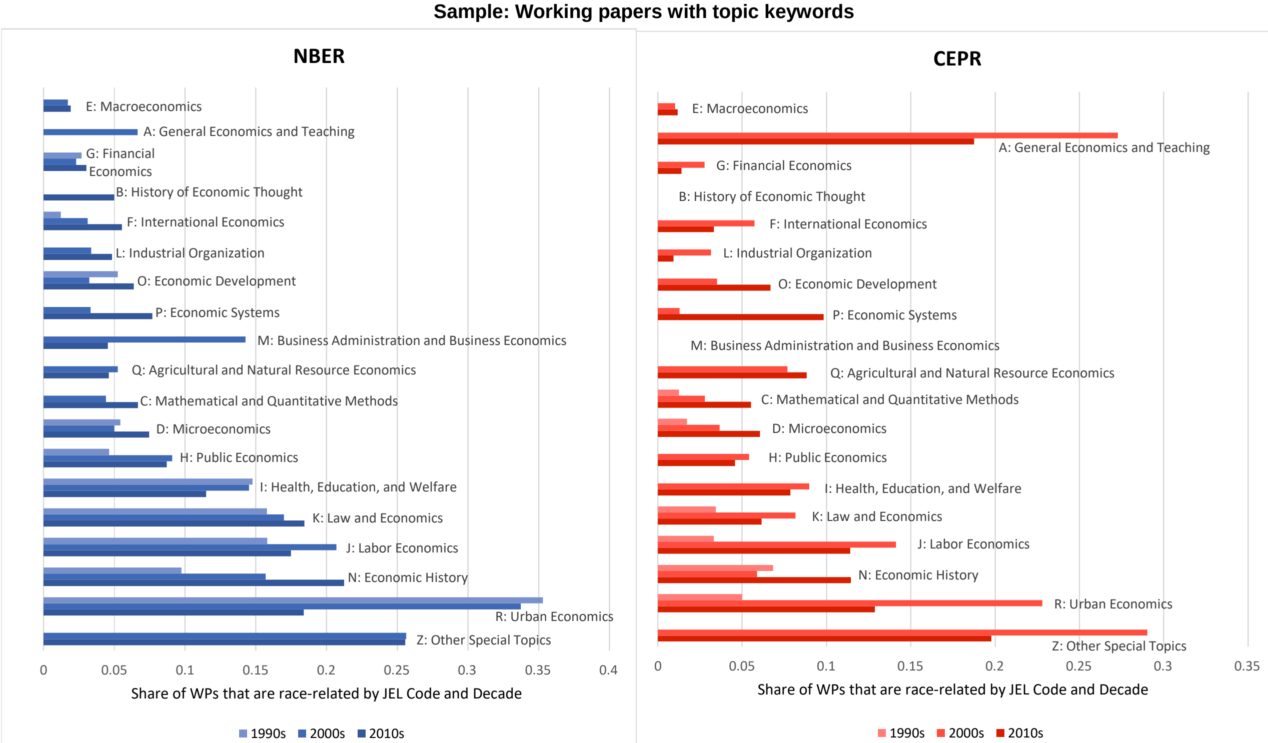
A. Share of Race-related Working Papers by JEL Category

Sample: Working papers with topic keywords



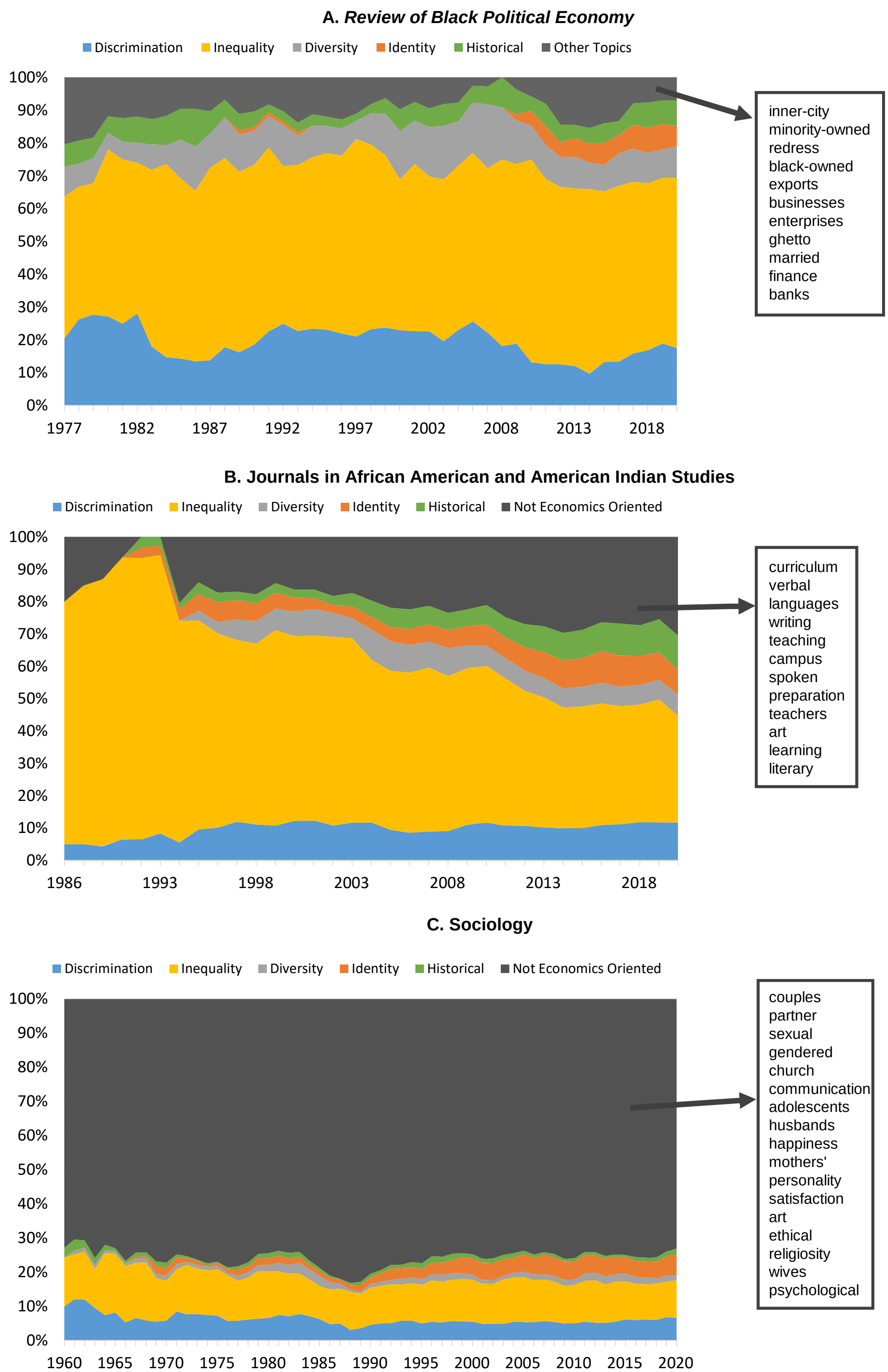
B. Share of Race-related Working Papers by JEL Category and Decade

Sample: Working papers with topic keywords



Notes: The sample is based on NBER working papers first released between 1974 and 2019, and CEPR working papers released between 1984 and 2019. In both cases we only consider working papers that mention at least one topic keyword in their title and abstract. Panel A shows the fraction of working papers that are identified to be race-related by JEL classification, as well as the share of all working papers in that series by JEL code. When a working paper has multiple JEL codes, we split the assignment article equally across all codes. We omit working papers with no JEL classification and JEL code Y, *Miscellaneous Categories* because this is not represented in the NBER corpus and is associated with eight papers in the CEPR series, among which none are race-related. Panel B shows the same information split by decade of posting.

Figure 8: Race-related Topics Studied in Three Related Cases



Notes: We use a corpus of publications in based on data from *JSTOR* , *Web of Science* and *Scopus* . When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. We report five-year moving averages throughout. Panel A focuses on publications in the *Review of Black Political Economy* . Panel B focuses on journals from the discipline of African American and American Indian studies (as defined by *JSTOR*). Panel C focuses on journals in sociology. In Panels A and B, we assume all published articles are race-related, decomposing them into the broad topic areas and hence identifying those not covered by any topic area. We then select some prominent keywords in these other topic publications. For Panel C we start by restricting to publications with some group related keywords in the title and/or abstract.

Table A1: Group Keywords with Regular Expression Patterns

Non-Specific - Band 0	Decomposition Group
aboriginal	Non-Specific
advantaged[-]?group[a-zA-Z]{0,1}	Non-Specific
caste[a-zA-Z]{0,1}	Non-Specific
colou?red[a-zA-Z]{0,1}	Non-Specific
disadvantaged[-]?minor[a-zA-Z]{0,5}	Non-Specific
dominant[-]?group[a-zA-Z]{0,1}	Non-Specific
ethnic minorit[a-zA-Z]{0,3}	Non-Specific
ethnic[a-zA-Z]{0,4}	Non-Specific
indigenous	Non-Specific
natives	Non-Specific
non[-]?western[a-zA-Z]{0,1}	Non-Specific
non[-]?white[a-zA-Z]{0,1}	Non-Specific
people[-]?of[-]?colou?r	Non-Specific
person[a-zA-Z]{0,1}[-]?of[-]?colou?r	Non-Specific
rac[a-zA-Z]{0,3}	Non-Specific
underrepresented[-]?minorit[a-zA-Z]{0,3}	Non-Specific
Main Minority Groups - Band 1	**Decomposition Group**
african[-]?american[a-zA-Z]{0,1}	Black
afro[-]?american[a-zA-Z]{0,1}	Black
black[-]?american[a-zA-Z]{0,1}	Black
black[a-zA-Z]{0,1}	Black
negro[a-zA-Z]{0,2}	Black
hispanic[-]?american[a-zA-Z]{0,1}	Hispanic
hispanic[a-zA-Z]{0,1}	Hispanic
latino[-]?american[a-zA-Z]{0,1}	Hispanic
latino[a-zA-Z]{0,1}	Hispanic
mexican[-]?american[a-zA-Z]	Hispanic
spanish[-]?american[a-zA-Z]	Hispanic
american[-]?indian[a-zA-Z]{0,1}	Native American
cherokee[a-zA-Z]{0,6}	Native American
chippewa[a-zA-Z]{0,3}	Native American
choctaw[a-zA-Z]{0,3}	Native American
native[-]?american[a-zA-Z]{0,1}	Native American
navajo[a-zA-Z]{0,3}	Native American
siouan	Native American
sioux	Native American
Less Prominent Groups - Band 2	**Decomposition Group**
asian[-]?american[a-zA-Z]	Asian
chinese[-]?american[a-zA-Z]	Asian
indian[-]?american[a-zA-Z]	Asian
indo[-]?american[a-zA-Z]	Asian
japanese[-]?american[a-zA-Z]	Asian
korean[-]?american[a-zA-Z]	Asian
oriental[a-zA-Z]{0,1}	Asian
south[-]?asian[a-zA-Z]{0,1}	Asian
vietnamese[-]?american[a-zA-Z]	Asian
arab	Other
arab[-]?american[a-zA-Z]	Other
caucasian[a-zA-Z]{0,1}	Other
cuban[-]?american[a-zA-Z]	Other
ethiopian[-]?american[a-zA-Z]	Other
filipino[-]?american[a-zA-Z]	Other
hebrew[a-zA-Z]{0,1}	Other
islam[a-zA-Z]	Other
jew[a-zA-Z]{0,3}	Other
jewish[-]?american[a-zA-Z]	Other
muslim[-]?american[a-zA-Z]	Other
muslim[a-zA-Z]	Other
palestinian[-]?american[a-zA-Z]	Other
portuguese[-]?american[a-zA-Z]	Other
yiddish	Other

Notes: [a-zA-Z]{0,k} indicates that we allow any number of 0 to 'k' lowercase or uppercase characters to be matched. [-]? allows for an optional hyphen or space. We also account for American and British English spellings, for instance, in colou?red[a-zA-Z]{0,1}.

Table A2: Topic Keywords with Regular Expression Patterns

Discrimination (41)	Inequality (23)	Diversity (18)	Identity (4)	Historical (17)
-group bias	black youth[a-zA-Z]{0,1}	affirmative[-]?action[a-zA-Z]{0,1}	rac[a-zA-Z]{0,3} identit[a-zA-Z]{0,3}	black vot[a-zA-Z]{0,3}
animosit[a-zA-Z]{0,3}	black-white	desegregat[a-zA-Z]{0,3}	acting white	civil rights
animus	development	ethnic composition[a-zA-Z]{0,3}	identity	emancipat[a-zA-Z]{0,3}
anti[-]?black	disadvantage	ethnic[-]?diversity	identities	eugenics
anti[-]?discrimination	disadvantaged	ethnic[-]?fragmentation[a-zA-Z]{0,1}		jim crow
anti[-]?semitic	educat[a-zA-Z]{0,5}	ethnic heterogene[a-zA-Z]{0,5}		lynch[a-zA-Z]{0,5}
antisemitism	ethnic differen[a-zA-Z]{0,4}	ethnic integration[a-zA-Z]{0,1}		political disenfranchisement
apartheid	ethnic disparit[a-zA-Z]{0,3}	rac[a-zA-Z]{0,3} composition[a-zA-Z]{0,1}		postbellum
attitude[a-zA-Z]{0,1}	ethnic gap[a-zA-Z]{0,1}	rac[a-zA-Z]{0,3} integration[a-zA-Z]{0,1}		race relation[a-zA-Z]{0,1}
discriminat[a-zA-Z]{0,5}	ethnic inequalit[a-zA-Z]{0,3}	racial[-]?diversity		race riot[a-zA-Z]{0,3}
ethnic bias[a-zA-Z]{0,3}	gap[a-zA-Z]{0,1}	racial[-]?fragmentation[a-zA-Z]{0,1}		reconstruction[a-zA-Z]{0,1}
ethnic division[a-zA-Z]{0,1}	inequality	racial heterogene[a-zA-Z]{0,5}		slave[a-zA-Z]{0,2}
ethnic exclusion[a-zA-Z]{0,1}	living standard	representation		social[-]?activis[a-zA-Z]{0,1}
ethnic interact[a-zA-Z]{0,4}	standard of living	segregat[a-zA-Z]{0,3}		southern farm
ethnic stereotyp[a-zA-Z]{0,3}	negro-white	social[-]?diversity		the great migration
ethnic[-]?division[a-zA-Z]{0,1}	poverty	social[-]?fragmentation[a-zA-Z]{0,1}		tuskegee
ethnic[-]?exclusion[a-zA-Z]{0,1}	rac[a-zA-Z]{0,3} differen[a-zA-Z]{0,4}	tipping point		whitecapping
exploitation	rac[a-zA-Z]{0,3} disparit[a-zA-Z]{0,4}	underrepresent[a-zA-Z]{0,3}		
hatred	rac[a-zA-Z]{0,3} gap[a-zA-Z]{0,1}			
implicit bias[a-zA-Z]{0,4}	rac[a-zA-Z]{0,3} inequalit[a-zA-Z]{0,3}			
in-group	school[a-zA-Z]{0,3}			
ingroup	stratification			
institutional discrimination	welfare			
institutional racism				
inter-group				
intergroup				
oppress[a-zA-Z]{0,3}				
out-group				
outgroup				
prejudi[a-zA-Z]{0,4}				
rac[a-zA-Z]{0,3} bias[a-zA-Z]{0,4}				
rac[a-zA-Z]{0,3} interact[a-zA-Z]{0,4}				
rac[a-zA-Z]{0,3} profiling				
rac[a-zA-Z]{0,3} stereotyp[a-zA-Z]{0,3}				
racial[-]?division[a-zA-Z]{0,1}				
racial[-]?exclusion[a-zA-Z]{0,1}				
social[-]?division[a-zA-Z]{0,1}				
social[-]?exclusion[a-zA-Z]{0,1}				
statistical discrimination[a-zA-Z]{0,1}				
structural discrimination				
systemic racism				

Notes: [a-zA-Z]{0,k} indicates that we allow any number of 0 to 'k' lowercase or uppercase characters to be matched. [-]? allows for an optional hyphen or space.

Table A3: Eliminated Phrases with Regular Expression Patterns

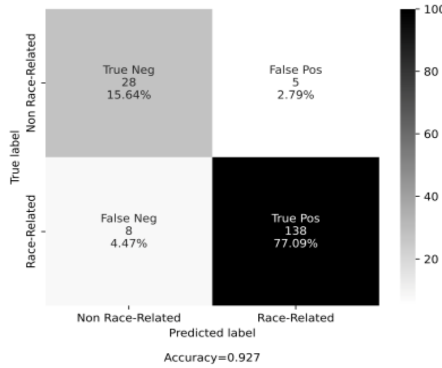
arms.{0,3}rac.{0,3}
black swan[a-zA-Z-]{0,1}
black.{0,3}box.{0,3}
black.{0,3}card[a-zA-Z]{0,1}
black.{0,3}economy
black.{0,3}market[a-zA-Z-]{0,3}
black.{0,3}scholes
electoral.{0,3}rac.{0,3}
horse.*rac.{0,3}
patent.{0,3}rac.{0,3}
priority.{0,3}rac.{0,3}
prize.*rac.{0,3}
r d.{0,3}rac.{0,3}
rac.*horse.{0,3}
rac.*prize.{0,3}
rac.*winner{0,3}
race[s]{0,1} between
rat.{0,3}.{0,3}rac.{0,3}
rd.{0,3}rac.{0,3}
rival
white.{0,3}collar
white.{0,3}noise
winner.*rac.{0,3}

Notes: [a-zA-Z]{0,k} indicates that we allow any number of 0 to 'k' lowercase or uppercase characters to be matched. [-]? allows for an optional hyphen or space.

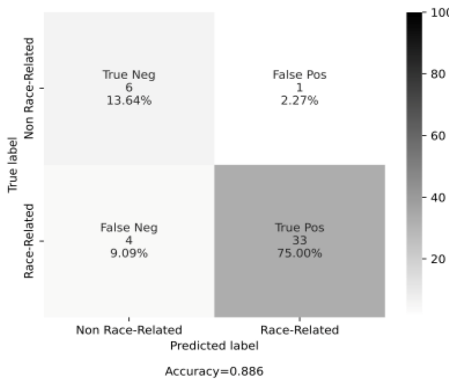
Figure A1: Validation Using GPT-3.5

A. Confusion Matrix for GPT

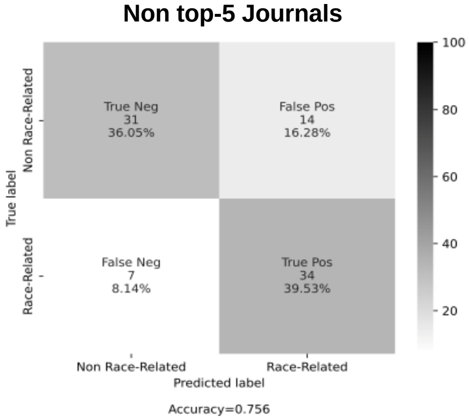
B. Confusion Matrix for Algorithm



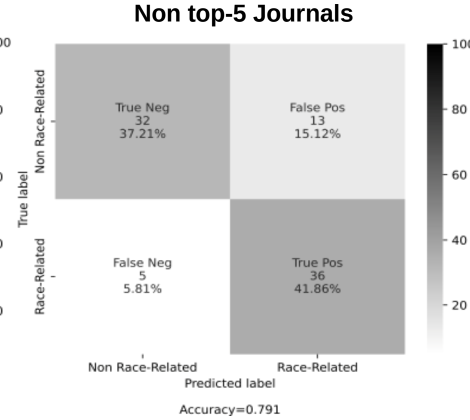
C. Confusion Matrix for GPT (Sample with Inroductions included)



D. Confusion Matrix for GPT, Non top-5 Journals



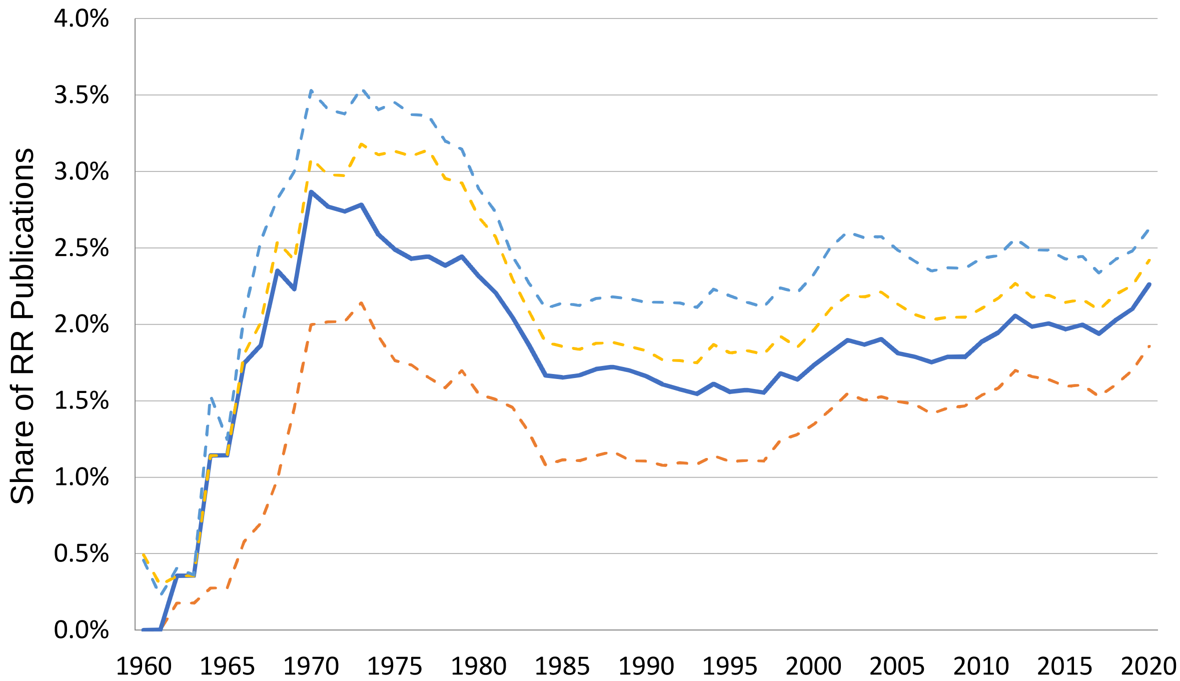
E. Confusion Matrix for Algorithm, Non top-5 Journals



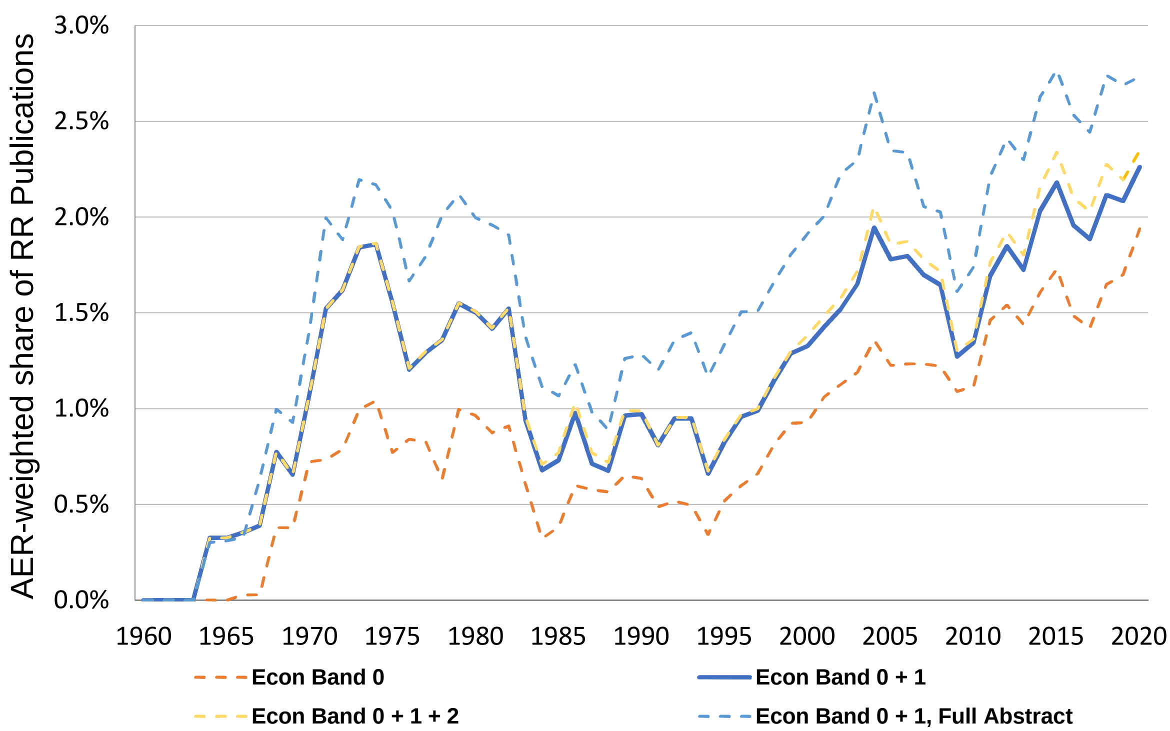
Notes: We use the OpenAI API to access GPT-3.5 for the validation exercise. Panels A and B show the output of classification of publications based on their titles and abstracts. The sample includes 179 publications mentioning a group keyword in their title or abstract (excluding the final sentence, and ignoring topic keywords and eliminated phrases) from the top-5 general interest journals from 1960 to 2020. We report confusion matrices for both the GPT classification output (Panel A) and the output obtained by implementing our algorithm on the same sample (Panel B), comparing them both to a hand-coded ground classification. These confusion matrices show the performance and efficacy of each classification model by summarizing the counts of true negatives (upper left quadrant), true positives (lower right quadrant), false positives (upper right quadrant), and false negatives (lower left quadrant). Panel C reports the confusion matrix for the GPT classification using additional information from the introduction of papers (as well as the title and abstract). This sample includes 44 publications selected from the validation sample described above. Panels D and E display confusion matrices for a random sample of 86 papers selected from non-top 5 economics journals. These papers specifically include group words in their abstracts or titles.

Figure A2: Bounds on Race-related Publications in Economics

A. Unweighted Share Bounds

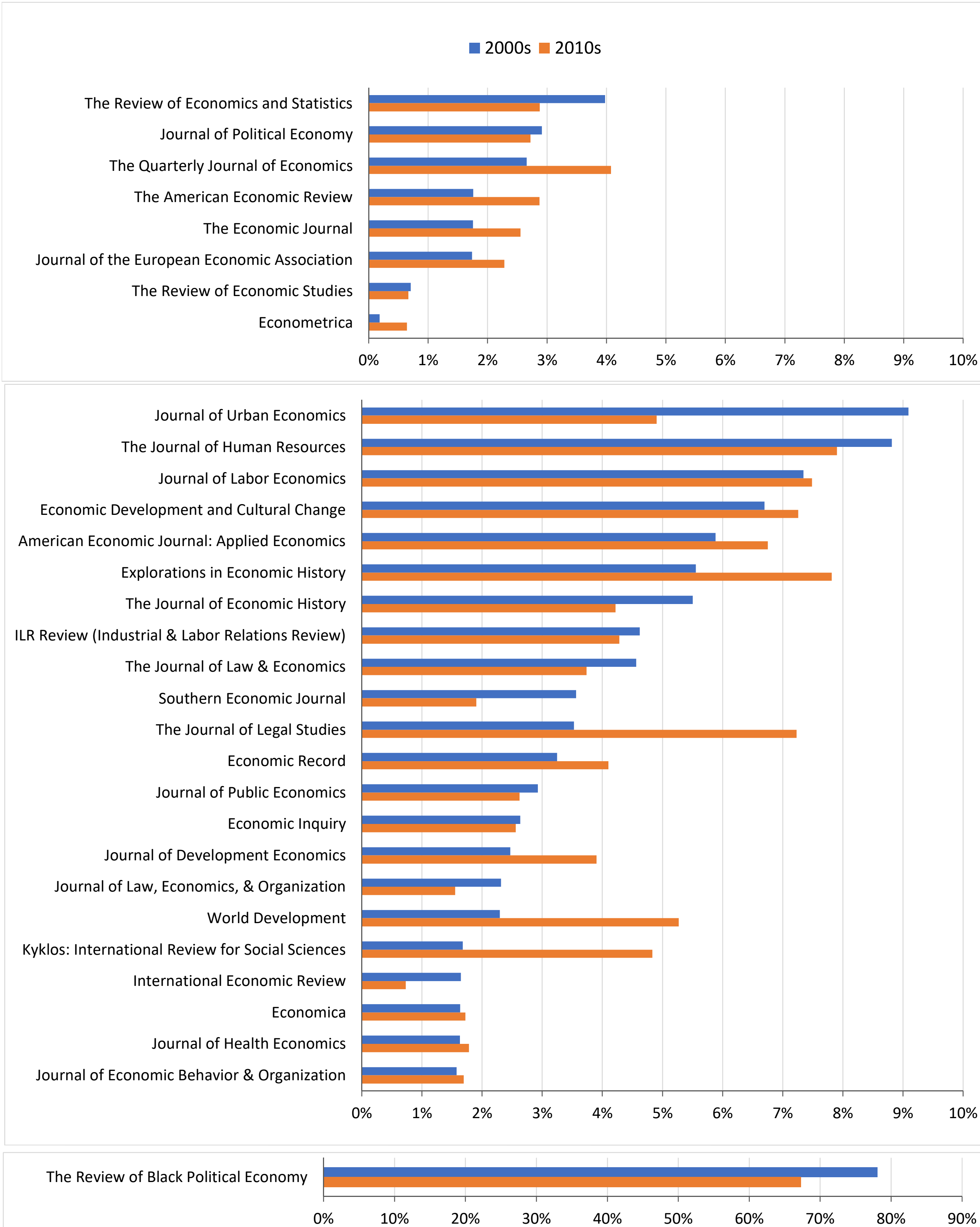


B. *AER* -Weighted Share Bounds



Notes: We use a corpus of publications in economics journals, based on data from *JSTOR*, *Web of Science* and *Scopus*. When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. We report five-year moving averages throughout. Panel A reports the share of total publications identified to be race-related by year of publication. Panel B reports the *AER*-weighted version of Panel A, using the journal weights constructed in Angrist *et al*. [2020]. Each Panel shows the resulting time series using alternative group keyword bands (see Table A1), or by using bands 0 and 1 and also including the last line of the abstract.

Figure A3: Race-related Publications, by Economics Journal



Notes: Eight general interest journals in economics are ranked separately and placed at the top. Within these eight and the other economics journals shown, the panels are ordered according to the share of race-related articles in that journal from 1960 to 2020. Each bar then shows the share of publications in the journal that are race-related (as identified by our algorithm), for publications in the 2000s and for the 2010s. The final series of bars are for the *Review of Black Political Economy* , for which the scaling of the x-axis differs.

Figure A4: Highly Cited Race-related Publications

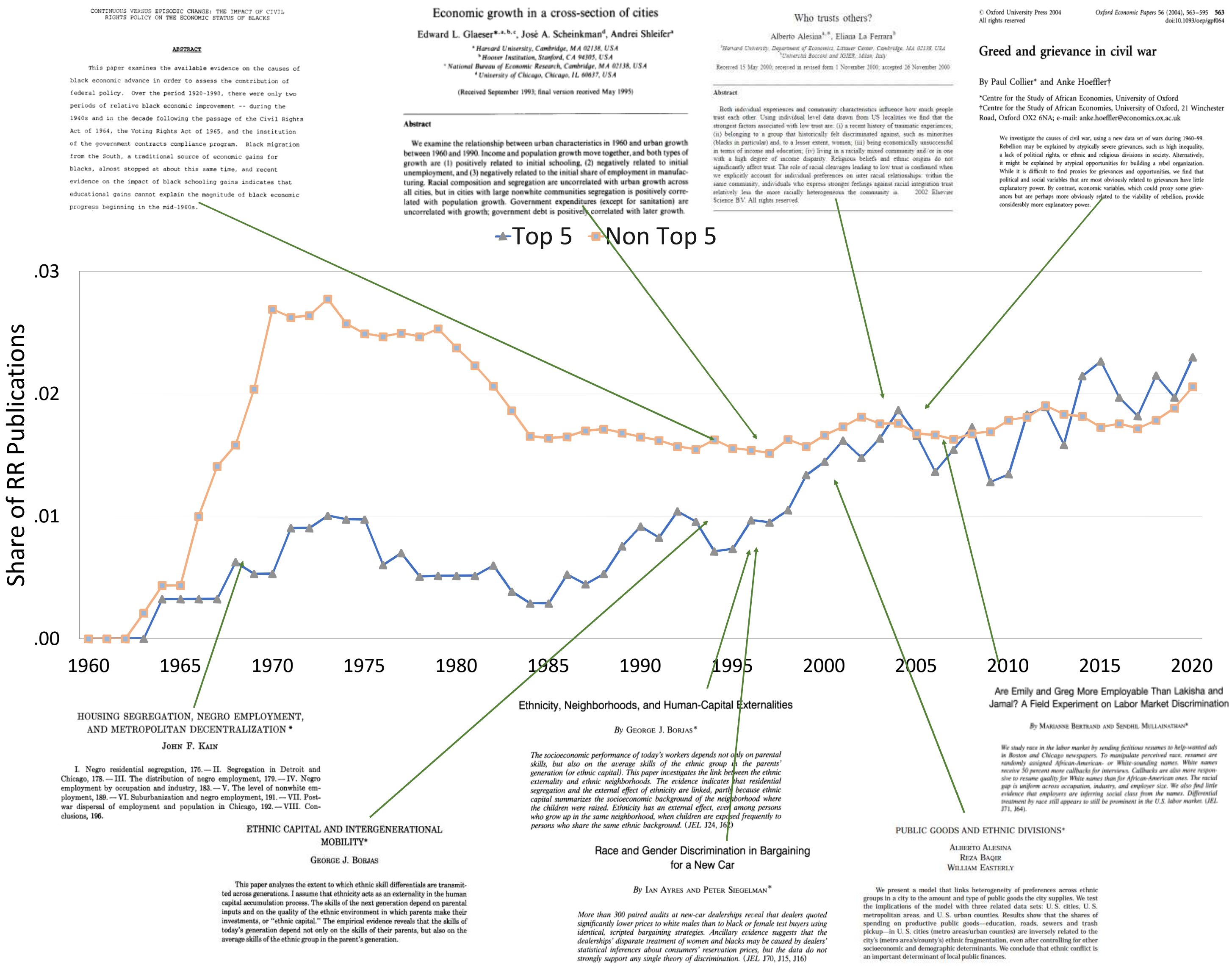
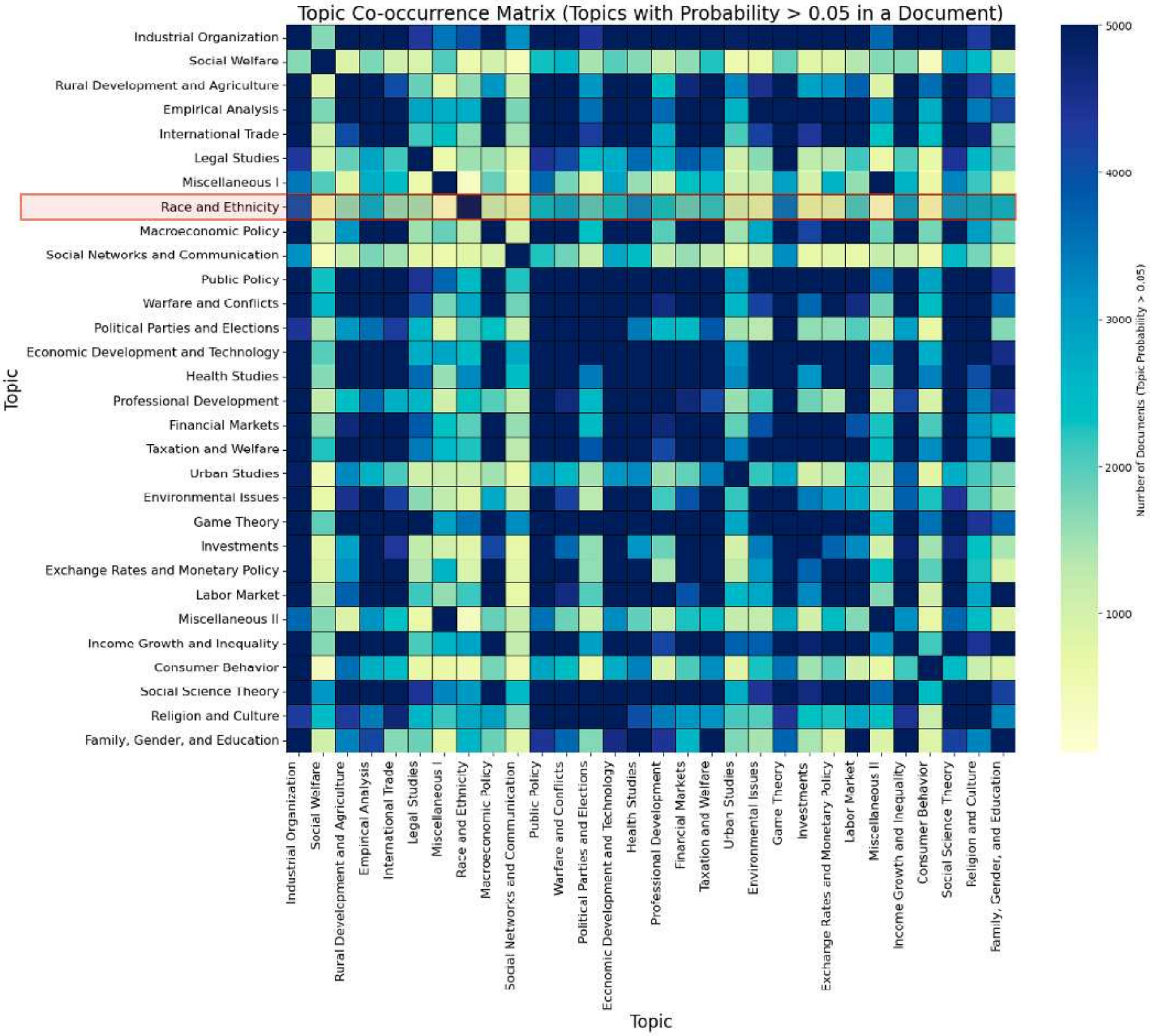


Figure A5: LDA Topics in the Full Corpus of Publications

Topic	Term1	Term2	Term3	Term4	Term5	Topic Label
1	model	firm	effect	datum	result	Industrial Organization
2	retirement	pension	french	plan	communist	Social Welfare
3	land	rural	resource	area	agricultural	Rural Development and Agriculture
4	model	test	method	paper	reserve	Empirical Analysis
5	country	trade	international	foreign	domestic	International Trade
6	law	right	court	legal	crime	Legal Studies
7	et	politique	les	ce	que	Miscellaneous I
8	group	black	white	ethnic	racial	Race and Ethnicity
9	long	run	shock	cycle	term	Macroeconomic Policy
10	network	social	trust	communication	medium	Social Networks and Communication
11	policy	public	government	reform	financial	Public Policy
12	war	conflict	year	state	more	Warfare and Conflicts
13	political	party	state	election	voter	Political Parties and Elections
14	economic	development	technology	new	research	Economic Development and Technology
15	datum	study	health	effect	measure	Health Studies
16	work	worker	job	more	organization	Professional Development
17	risk	market	asset	financial	insurance	Financial Markets
18	tax	income	welfare	government	household	Taxation and Welfare
19	city	urban	migration	migrant	housing	Urban Studies
20	environmental	cost	pollution	industry	plant	Environmental Issues
21	equilibrium	game	reserve	right	agent	Game Theory
22	capital	investment	energy	copyright	reserve	Investments
23	rate	exchange	price	monetary	inflation	Exchange Rates and Monetary Policy
24	labor	wage	employment	market	worker	Labor Market
25	que	et	este	por	article	Miscellaneous II
26	growth	income	inequality	population	increase	Income Growth and Inequality
27	food	climate	change	consumer	adaptation	Consumer Behavior
28	social	article	theory	research	approach	Social Science Theory
29	identity	class	cultural	society	culture	Religion and Culture
30	family	child	woman	gender	school	Family, Gender, and Education

Notes: We use Latent Dirichlet Allocation (LDA) modeling to identify topics in the corpus of 493,972 publications in economics, sociology, political science, law, management, public policy and history, based on data from *JSTOR*, *Web of Science*, and *Scopus*. When a journal is assigned to multiple disciplines, we split the number of publications in that journal equally across disciplines. We retain those journals that have paper titles and abstracts in English because our algorithm can be applied to such papers (even if the main text is then in a non-English language). Our benchmark model then identifies 30 topics. The Figure displays word clouds for the topics generated and we label each of the topics as shown in the lower part of the Figure.

Figure A6: Co-occurrence of Topics from the LDA Topic Model



Notes: This figure presents a Topic Co-occurrence Matrix based on the 30 topics derived from a LDA model, estimated on a corpus of 493,972 publications across economics, sociology, political science, law, management, public policy, and history. The data are sourced from JSTOR, Web of Science, and Scopus. When a journal is assigned to multiple disciplines, its publications are split equally across them. The co-occurrence is identified by checking whether two topics each have a posterior probability greater than 0.05 in the same document. Each cell has color based on the number of documents in which a given pair of topics co-occur above this threshold. Brighter colors indicate more frequent co-occurrence. As the LDA model allows documents to belong to multiple topics simultaneously, topics are not mutually exclusive.