

# DISCUSSION PAPER SERIES

DP16115

## **Race-related Research in Economics and Other Social Sciences**

Arun Advani, Elliott Ash, David Cai and Imran Rasul

**LABOUR ECONOMICS**

**PUBLIC ECONOMICS**

**CEPR**

# Race-related Research in Economics and Other Social Sciences

*Arun Advani, Elliott Ash, David Cai and Imran Rasul*

Discussion Paper DP16115

Published 06 May 2021

Submitted 04 May 2021

Centre for Economic Policy Research  
33 Great Sutton Street, London EC1V 0DX, UK  
Tel: +44 (0)20 7183 8801  
[www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Labour Economics
- Public Economics

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Arun Advani, Elliott Ash, David Cai and Imran Rasul

# Race-related Research in Economics and Other Social Sciences

## Abstract

How does economics compare to other social sciences in its study of issues related to race and ethnicity? We assess this using a corpus of 500,000 academic publications in economics, political science, and sociology. Using an algorithmic approach to classify race-related publications, we document that economics lags far behind the other disciplines in the volume and share of race-related research, despite having higher absolute volumes of research output. Since 1960, there have been 13,000 race-related publications in sociology, 4,000 in political science, and 3,000 in economics. Since around 1970, the share of economics publications that are race-related has hovered just below 2% (although the share is higher in top-5 journals); in political science the share has been around 4% since the mid-1990s, while in sociology it has been above 6% since the 1960s and risen to over 12% in the last decade. Finally, using survey data collected from the Social Science Prediction Platform, we find economists tend to overestimate the amount of race-related research in all disciplines, but especially so in economics.

JEL Classification: A11, Z13

Keywords: N/A

Arun Advani - a.advani.1@warwick.ac.uk  
*University of Warwick, Institute for Fiscal Studies*

Elliott Ash - ashe@ethz.ch  
*ETH Zurich*

David Cai - jiyang.cai@gess.ethz.ch  
*ETH Zurich*

Imran Rasul - i.rasul@ucl.ac.uk  
*University College London and CEPR*

## Acknowledgements

We gratefully acknowledge financial support from the ESRC through CPP at IFS (ES/H021221/1) and CAGE at Warwick (ES/L011719/1). We thank the authors of Angrist et al. [2020] for sharing their data on journal weights. We thank Andrea Moro, Eva Vivalt, and seminar participants from the Marist School of Management and University of Warwick for valuable comments. We also thank the team at the Social Science Prediction Platform for their support in setting up the survey. This paper is based on evidence presented at the Econometric Society World Congress meeting in August 2020, in a policy session titled, 'What Can Economics do for Racial Justice?'

# RACE-RELATED RESEARCH IN ECONOMICS AND OTHER SOCIAL SCIENCES\*

ARUN ADVANI      ELLIOTT ASH      DAVID CAI      IMRAN RASUL<sup>†</sup>

APRIL 2021

## Abstract

How does economics compare to other social sciences in its study of issues related to race and ethnicity? We assess this using a corpus of 500,000 academic publications in economics, political science, and sociology. Using an algorithmic approach to classify race-related publications, we document that economics lags far behind the other disciplines in the volume and share of race-related research, despite having higher absolute volumes of research output. Since 1960, there have been 13,000 race-related publications in sociology, 4,000 in political science, and 3,000 in economics. Since around 1970, the share of economics publications that are race-related has hovered just below 2% (although the share is higher in top-5 journals); in political science the share has been around 4% since the mid-1990s, while in sociology it has been above 6% since the 1960s and risen to over 12% in the last decade. Finally, using survey data collected from the *Social Science Prediction Platform*, we find economists tend to overestimate the amount of race-related research in all disciplines, but especially so in economics. **JEL Classification: A11, Z13.**

---

\*We gratefully acknowledge financial support from the ESRC through CPP at IFS (ES/H021221/1) and CAGE at Warwick (ES/L011719/1). We thank the authors of Angrist *et al.* [2020] for sharing their data on journal weights. We thank Andrea Moro, Eva Vivalt, and seminar participants from the Marist School of Management and University of Warwick for valuable comments. We also thank the team at the Social Science Prediction Platform for their support in setting up the survey. This paper is based on evidence presented at the Econometric Society World Congress meeting in August 2020, in a policy session titled, ‘What Can Economics do for Racial Justice?’ All errors remain our own.

<sup>†</sup>Advani: University of Warwick and Institute of Fiscal Studies (IFS), a.advani.1@warwick.ac.uk; Ash: ETH Zurich, ashe@ethz.ch; Cai: ETH Zurich, jiyang.cai@gess.ethz.ch; Rasul: University College London and IFS, i.rasul@ucl.ac.uk.

# 1 Introduction

Twenty twenty is a year that will go down in history: not only was this the year of a global pandemic, but it was also the year in which issues of racial justice – driven in part by the Black Lives Matter movement – rose to the top of public debate on both sides of the Atlantic. This has prompted much discussion within the academy and economics. Exemplifying this trend, the August 2020 Econometric Society World Congress organized a special policy panel titled, “What can economics do for racial justice?” A simple answer to the question would be for economists to research and publish on the causes and consequences of racial inequality – an issue highlighted by Cook [2021]. From a practical perspective, then, it is important to know how much economists are already doing such research, and how that quantity compares to other disciplines.

To that end, this paper quantifies the volume and share of research in economics on race- and ethnicity-related issues, comparing this volume of research output to two other social sciences: sociology and political science. We make comparisons to these disciplines because they are social sciences where the study of race and ethnicity would *a priori* also be of central importance.<sup>1</sup>

Of course, the publications process is closely interlinked with: (i) incentives individual academics have to conduct to such research; (ii) the selection of individuals into economics (see also Charles [2021]). In a more extensive companion paper [Advani *et al.* 2021a], we examine these issues in greater detail. In this volume, Koffi and Wantchekon [2021] also highlight that even where scholars from under-represented backgrounds select in to economics – African scholars in their case – the “evidence points to some degree of discrimination” in terms of access to networks such as the NBER. In this paper our focus is on a simple cross-disciplinary comparison of the production of race-related research.

## 2 Method

**Corpus.** We build a corpus of publications for economics, political science, and sociology. The foundation for this corpus is the *JSTOR* database of academic journals ([jstor.org](https://www.jstor.org)). We consider all publications in journals that *JSTOR* classifies under the disciplines of economics, sociology, and political science. Although publication series are available back to the 1880s, our analysis focuses on publications from 1960 through 2020, where for each publication we use information on the journal it is published in, when it is published, the article title, and full text of the abstract. *JSTOR* has gaps in its publication series (especially in more recent years) and is missing some prominent journals. We fill these gaps using comparable publications data from *Web of Science* ([webofknowledge.org](https://www.webofknowledge.org)) and *Scopus* ([scopus.com](https://www.scopus.com)). The Data Appendix describes the procedure

---

<sup>1</sup>There are of course many normative views on what should constitute economics, including Lionel Robbins definition as “the science which studies human behavior as a relationship between ends and scarce means which have alternative uses.” Our approach takes the practical and positive view of economics as what economists do, whereby the subject matter is defined by what research is actually published.

through which we access these databases, other details on how we classify journals by discipline, and what is recorded for each publication.

Panel A of Figure A1 details the time series of journals covered in our corpus. For each discipline this rises steadily over time. Our working sample from 1960 to 2020 covers nearly half a million journal publications: 224,855 publications from 231 economics journals, 138,188 publications from 185 sociology journals, and 110,835 publications from 213 political science journals.

**Identifying Race-Related Research.** Given the volume of publications considered, it is infeasible to codify race-related research by hand. We thus take an automated approach and use an algorithm to classify race-related publications. We do so using keywords along two dimensions: (i) the racial or ethnic group being studied; and, (ii) the issue being studied. Examples of (case-insensitive) keywords along the group dimension are race, african-american, person of color, and ethnicity. Examples of (case-insensitive) issue keywords include discrimination, prejudice, and stereotype.<sup>2</sup>

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. For rule (ii) we drop the last sentence of the abstract to avoid false positives from research that only mentions race parenthetically, say because it is part of some robustness check rather than the primary focus of study.

The group and issue keywords are designed to reflect race-related research in the United States and other countries, and to allow for changing groups and issues over time and place. However, we fully recognize that there is no definitive way to classify race-related research, and indeed there might be reasonable normative differences on what such a body of work should constitute. Therefore we take an approach where we gradually extend our group keywords to cover broader notions of race and ethnicity.

Specifically, we define three bands of group keywords that gradually expand on the racial or ethnic groups being studied. Band 0 consists of only abstract or generic keywords denoting racial and ethnic groups (e.g. race, ethnic, under represented minority). Band 1 adds group keywords relating to the main minority groups in the US (African American, Latinos and Native Americans). Band 2 adds less salient group keywords (e.g. White, South Asian, Indian American, Japanese American) and other minorities based on religious beliefs (e.g. Muslim, Jewish). The full lexicon of group keywords used by Band are shown in Appendix Table A1.

The lexicon of issue keywords, shown in Appendix Table A2, are held constant and not split into bands. These words and phrases are broadly split across five broader topics: discrimination, inequality, diversity, identity, and historical issues. For example, discrimination includes prejudice

---

<sup>2</sup>Along each dimension we use wildcard keywords, so the keyword rac\* captures race, races, racial, racist, and racism. Similarly, wildcard issue keywords include discriminat\*, prejudi\*, and stereotyp\*. See Tables A1 and A2 for the full list of group and issue keywords.

and stereotypes, while inequality includes disparity and disadvantage.

Our primary results below are based on Band 1 group keywords, but we also show how our main results vary using other Bands.

Two further points on the selection of keywords are of note. First, the issue keywords are purposefully chosen to be economics-oriented. We are thus likely to under count race-related research in other disciplines. Second, we do not aim to capture research that is not specifically about race but could potentially be applied to understand racial inequalities.<sup>3</sup> There might also be related theoretical work, say on the formation of social preferences, or papers on discrimination that do not mention any group keywords but instead refer to ‘blue’ and ‘red’ groups. Such work is relevant to study racial gaps, but if those publications do not explicitly mention racial/ethnic groups or issues in their abstract or title, they will not be captured by our algorithm. If such work is more common in economics than other social sciences this can lead us to overstate gaps between disciplines in race-related research. We return to this point below, and make precise how voluminous such work would have to be in order to close race-related publication gaps between economics and other disciplines.

**False Positives and False Negatives.** Using an automated approach to classify race-related research leads to misclassification errors in the form of false negatives and false positives. To reduce the rate of false negatives (publications that are truly race-related but missed by our algorithm), we are relatively inclusive in the construction of the lexicon. On false positives, qualitatively we have found that using the combination of group *and* issue keywords removes many instances of non-race-related research that might otherwise match our lexicon patterns. These include, for example, publications about econometric bias, black markets etc. Dropping the last sentence of abstracts before applying the algorithm also helps reduce false positives by excluding many papers where race/ethnicity is not the primary focus.

To quantify potential rates of false negatives and false positives, we start with the sample of all publications in top-five (T5) general-interest economics journals: the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies*. We take all publications in T5 journals that are classified as race-related using our automated classification procedure. We then compute rates of false positives and false negatives in this sample by comparing the machine-coded classifications to a hand-coded classification. Hand-coding was done by manually checking all T5 articles that contain a group *or* issue keyword, whether or not they contain both. As this is obviously a selected sample of all publications in economics, we have to make the following assumptions in order to apply these rates to our full corpus in economics: (i) no race-related research is conducted in T5 publications that

---

<sup>3</sup>Cook and Logan [2020] discuss many policies that have unintended impacts on racial inequality, including those with urban biases or those related to tax/benefit exemptions. Minorities might also be more impacted by changes in minimum wages [Derenoncourt and Montialoux 2021].

exclude both these group and issue keywords; and (ii) the resulting hand-coded misclassification rates found in the T5 apply equally to non-T5 journals.

Following this approach, the rate of false positives is given by:

$$\frac{\text{\#Non race-related publications labeled as race-related}}{\text{\#False Positive} + \text{\#True Negative}} = 3.65\%, \quad (1)$$

and the rate of false negatives is given by:

$$\frac{\text{\#Race-related publications labeled as not race-related}}{\text{\#False Negative} + \text{\#True Positive}} = 25.4\%. \quad (2)$$

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

$$\frac{\text{\#Race-related publications}}{\text{\#Publications labeled as race-related}} = \frac{\text{\#False Negative} + \text{\#True Positive}}{\text{\#False Positive} + \text{\#True Positive}} = 98.8\% \quad (3)$$

In the Appendix, we show what trends in the estimated volume of race-related research are in the worst-case and best-case scenarios based on these rates of misclassification error.

**Journal Weights.** Our counts of race-related research make no adjustment for the quality of journals that work is published in. It will however be useful to consider both the quantity and quality of race-related research. To adjust for journal quality, we adapt the journal weighting scheme employed by Angrist *et al.* [2020] in their study of the intermural influence of economics on other disciplines. Journal weights are given by the relative frequency with which the journal is cited by the top ‘trunk’ journal in the discipline. For the disciplines we focus on, these trunk journals are the *American Economic Review*, *American Sociological Review*, and *American Political Science Review*. The weight of journal  $j$  in year  $t$  is given by:

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

As these weights are time-varying they capture the rise and fall of the importance of journals within their discipline.

We use these weights to measure the quality of journals in our corpus over time. Panel B of Figure A1 shows that from the 1980s onwards, the quality 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, and this is so for all three disciplines considered.



## 3 Results

### 3.1 Comparing Economics To Other Disciplines

Figure 1 presents our main results. Panel A shows the total number of journal publications over time (covering any issue), by discipline. Following Kleven [2018] and Angrist *et al.* [2020], we plot five-year moving averages to smooth year-to-year noise but still pick up trends. There has been a steady rise in publications in all three social sciences, with the greatest increase being in economics. Economics has nearly always produced more publications than political science, and economics overtook sociology in the volume of publications in the mid-1980s. By the 2010s, there are around 6000 publications annually in economics, compared to something closer to 4000 publications annually in sociology and in political science.<sup>4</sup>

Despite this higher volume of total research output, Panel B shows that economics lags behind sociology and political science in the absolute number of race-related publications (using our Band 1 definition of such research). Indeed, for the most recent years, around five times as many race-related papers were published in sociology than in economics each year, with just under double the number being published annually in political science than economics. Since the start of our study period in 1960, a greater volume of race-related research has always been published in sociology than economics – there never was a golden age where economists published more of such work. Cumulatively, from 1960 to 2020, there have been 13,000 race-related publications in sociology, 4000 in political science, and 3000 in economics. The cumulative number of race-related publications in economics by 2020 is at the level that political science reached in 2014, and where sociology reached in 1996. These differences are all despite our classification lexicon being biased towards economics-related issue keywords (as shown in Appendix Table A2).

As mentioned earlier, we might under-count the volume of race-related research in economics by not including theoretical work that does not mention any group keywords but instead refers more abstractly to ‘blue’ and ‘red’ groups, say. The publication gaps shown in Panel B give a sense that such work would have to (implausibly) constitute the vast majority of race-related work in economics in order to close the gap with the other disciplines.

To move beyond the quantity of race-related research but also account for its quality, Panel C shows journal-weighted time series of race-related publications by discipline. As in the raw count data, each discipline displays a steady upward trend but the ranking across disciplines remains unchanged. By the end of our sample, each year economics produces around half the quality-adjusted volume of race-related research as political science, and a quarter of that in sociology. This suggests that the overall disparity in publication output between economics and the other disciplines is not compensated by publication in relatively higher-quality journals (as measured

---

<sup>4</sup>As the publication series start in the 1880s, the publication numbers do not start exactly at zero in 1960, the first year of our working sample.

by the Angrist trunk-journal citation weights).

Finally, Panel D shows the unweighted *share* of race-related publications by year and discipline. The lower number of race-related publications in economics (Panel B) is compounded by the higher number of total publications in the discipline (Panel A), so that the share of race-related publications in economics is far below that in sociology and political science. Moreover, there is no longer an upward trend in the share of race-related publications, in contrast to the (weighted) number of such publications in economics. Since around 1970, the share of publications in economics that are race-related has hovered just below 2%; in political science the share has been around 4% since the mid-1990s, while in sociology it has been above 6% since the 1960s, and risen to over 12% in the last decade. There is no period in our sample where economic produces a higher share of race-related research than the other disciplines.

**Robustness.** Panel A of Figure A2 shows that our measure of the broad level and trends in race-related research in economics is robust to: (i) narrowing the definition of race-related research to only use Band 0 group keywords; (ii) broadening the definition of race-related research to also use Band 2 group keywords; (iii) deriving the upper and lower bounds on our baseline definition using rates of false positives and negatives derived above; and (iv) dropping the rule requirement of not using the last sentence of abstracts to classify race-related research. Panel B shows similar robustness using the journal quality weighted time series for economics.

### 3.2 Top-5 Economics Journals

We next narrow in on the top-5 (T5) economics journals. It is useful to consider this top tier separately because it represents what is considered general interest and reflects the views of leading scholars, editors, and referees. In addition, these papers tend to be highly cited [Heckman and Moktan 2020].

Panel A of Figure 2 shows the total number of race-related publications in T5 journals. There appears to be a structural break in the number of race-related published articles around 2000, with a level-shift up since then. Pre-2000, there were fewer than three race-related papers published each year in this top-tier, while since 2000 this number has trebled. Still, the top-tier economics journals collectively struggle to publish 10 race-related papers per year.

The absolute number of publications reflects the restricted supply of slots in the top-tier journals. Panel B shows the share of all T5 publications that are race related, comparing it to the share in non-T5 journals. Here we see a more encouraging picture emerging: since the early 2000s, the share of race-related publications has been higher in T5 journals than other journals. If we view publications in the T5 as important for career incentives, this could be an important indication that the discipline is steadily providing greater incentives to work on such issues. We examine this in greater detail in Advani *et al.* [2021a].

## 4 Do Economists Recognize This?

To return to the central theme of the World Congress policy panel, if we view the lower publication rate of race-related research in economics relative to sociology and political science as concerning, then a first step in correcting this is for economists to recognize the problem. To assess the extent to which economists are aware of the facts described above, we used the *Social Science Prediction Platform* to survey economists on the share of race-related publications in the three disciplines. Academic participants were recruited via direct emails and Twitter advertising of the poll. Public sector economists were recruited via emails to the UK Government Economics Service and Bank of England internal mailing lists. We gathered a sample of 240 responses from academic and public sector economists in the 10 days prior to the ESWC in August.<sup>5</sup>

We asked participants to state their beliefs about the share of race-related research produced in economics and other disciplines, over time, and between T5 and other economics journals (see Advani *et al.* [2020a] for the complete survey). Questions were of the form, “What share of articles published in [discipline/journal type] between [dates] do you think were race-related?” These questions were asked after a brief explanation of our definition of a race-related publication (that is close to what is described as Band 1 above).

**Comparing Disciplines.** Figure 3A shows a box-and-whisker plot of the reported estimate of the share of race-related publications by discipline for the 2000-20 period. We show the 10th, 25th, median, 75th, and 90th percentiles of each prediction. The black line on each plot replicates the statistics from Figure 1D on the actual share of race-related research by discipline, as estimated using our algorithm and based on Band 1 group keywords. Economists correctly predict the disciplinary ranking but overestimate the share of race-related research in all three disciplines. Two-thirds of economists overestimate the share of race-related research in sociology, 80% do so for political science, while 90% overestimate it in economics. The ratio of the median estimate to the true rate is 4.04 for economics, 2.47 for political science, and 1.24 for sociology. Hence economists appear least informed and most optimistic about the publication of race-related research in their own discipline.

**Comparing Trends and T5 Journals.** Figure 3B shows predictions for different time periods across all economics journals, and also specifically for the T5 economics journals. As before, economists tend to overestimate the relative share of race-related publications: whether now or in the past, at least 80% of economists are overly optimistic. Economists also overestimate the increase in the share of race-related research since 2000. The median prediction for the period

---

<sup>5</sup>Of the 240 responses, 48% were from academics and 52% from economists working in the public sector. We had an additional 28 responses from non-economist academics, mainly sociologists and political scientists, and a further 28 from individuals in the private or charity sectors. We do not perform further analysis with these latter groups as they are too small to be considered separately.

since 2000 is 38% higher than the median prediction for the pre-2000 period. This is in the right direction, although almost four times larger than the true (10%) increase in this publication share.

Moreover, economists incorrectly predict that race-related research is less prevalent in the T5 journals. At an individual level, 78% of economists make this prediction, and among this group the median prediction is a difference of 2pp (or around 40% less likely in relative terms). As seen in Figure 2B, while this was true pre-2000, T5 journals are now (slightly) more likely to publish race related research than non-T5 journals. Despite being pessimistic about the T5 relative to other journals, even here 70% of economists over predict the share of race-related research.

Finally, we note that there were no meaningful differences in the median reports between tenured faculty, untenured faculty, and students/postdocs. With the caveat that we have small sample sizes for these breakdowns (between 30 and 43), the pattern suggests that experience or cohort effects are unlikely to lead to improved recognition of the gaps in research priorities. Thus there is scope for information or policies to address incorrect beliefs about the production and publication of race-related research.

## 5 Discussion

So what can economics do for racial justice? The starting point for any answer is for economists to first study the causes and consequences of racial inequalities [Charles 2021]. In comparison to sociology and political science, economics potentially has a decades-long gap to make up – a gap that would be even wider if our lexicon of keywords extended beyond economics-focused issues, and one that is unlikely to be fully explained by under-counting some kinds of race-related research in economics. If efforts for racial justice can be measured by the volume of publications, economists might have some catching up to do.

Normatively speaking, these differences in research outputs may reflect acceptable differences in subject matter across economics, sociology, and political science. In the companion working paper [Advani *et al.* 2021a], we show that if the topics in economics articles are re-weighted to match those in sociology (using an unsupervised topic model applied to abstract texts), the gap closes somewhat but a large disparity remains. If subject matter can only explain some but not the bulk of the difference across disciplines, it is natural to ask: What are the economics-specific factors that can narrow this gap?

Our more extensive working paper examines the underlying factors in more detail, first by exploring the race-related topics that have been studied in economics and how those topics differ from other disciplines. This analysis helps shed light on under-researched areas that may provide important contributions on the causes and consequences of racial inequalities. Second, we examine in more detail the incentives researchers have to produce race-related research by looking at publication outcomes and citations. Third, we ask how this interacts with the selection of individuals

from minorities into economics. It is through understanding these three dimensions – production, incentives, and selection – that we might begin to identify the economics-specific interventions that may remedy the current situation.

## A Data Appendix

**Data Sources.** The main dataset of publications is from *JSTOR's DFR* service ([jstor.org/dfr](http://jstor.org/dfr)). We downloaded publication-level data using the *JSTOR* categorizations by subject: economics, sociology and political science. The list of journals by subject are available at [jstor.org/subjects](http://jstor.org/subjects). We use *JSTOR's* discipline definition for each journal, except when Angrist *et al.* [2020] specify an alternative classification. For example, *JSTOR* lists the *Journal of Legal Studies* in law as well as economics, while the Angrist *et al.* [2020] weights specify it only as a journal in economics. When a journal is deemed to belong to multiple disciplines, we split the number of publications in that journal equally across disciplines.

Our second data source is *Web of Science (WoS)*. We use the search API service offered by Clarivate. To download data, we started with the lists of journals from *JSTOR* and Angrist *et al.* [2020]. We searched for these journals in the *WoS* ‘source’ field. We then exported the resulting JSON data and extracted publication date, title, and abstract.

The third source for publications is *Scopus*, another API-based service hosted by Elsevier. Some recent articles, especially in the top-five economics journals, are available in *Scopus* but not the other databases.

For each source we retain the same fields for each publication: the subject(s), journal, publication date, title and full text of the abstract.

**Corpus Coverage.** We start with the set of journals covered by *JSTOR* and use *WoS* and *Scopus* to fill in missing publications. Most missingness occurs over the post-2009 period. Articles in *WoS* but not in *JSTOR* were detected based on normalized publication title and publication date. We insert publications from *Scopus* when we cannot find another publication in the same journal and same year with the same title. For example, some prominent economics journals, such as *Journal of Public Economics* and *Review of Black Political Economy*, are not included in *JSTOR* but are available on *WoS*.

A number of journals are excluded because they do not report abstracts. For example, the *American Economic Review Papers & Proceedings* never have abstracts. Other journals are excluded for those years in which abstracts are not included. For example, prior to 1994, the *Economic Journal* did not report abstracts.

The earliest article in *JSTOR* is from 1887, but coverage remains patchy until the mid-1900s. Hence our working sample starts in 1960. *WoS* includes abstracts starting in the 1990s, and these

are consistently provided until around 2018. *JSTOR* abstracts drop off around 2015. Many of these missing abstracts are found through *Scopus*.

As a result of these data issues, we have a slight rise in publication numbers in the 1990s, caused by the availability of *WoS* filling in missing data from *JSTOR* – this can be seen in Panel A of Figure A1. The drop-off after 2015 is from *JSTOR* dropping publications, and these being only imperfectly filled using *WoS* and *Scopus*.

**Identifying Race-Related Publications.** We performed string matching on the titles and abstracts for each publication. 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. For rule (ii) we drop the last sentence of the abstract to avoid false positives from research that only mentions race parenthetically, say because it is part of some robustness check rather than the primary focus of study.

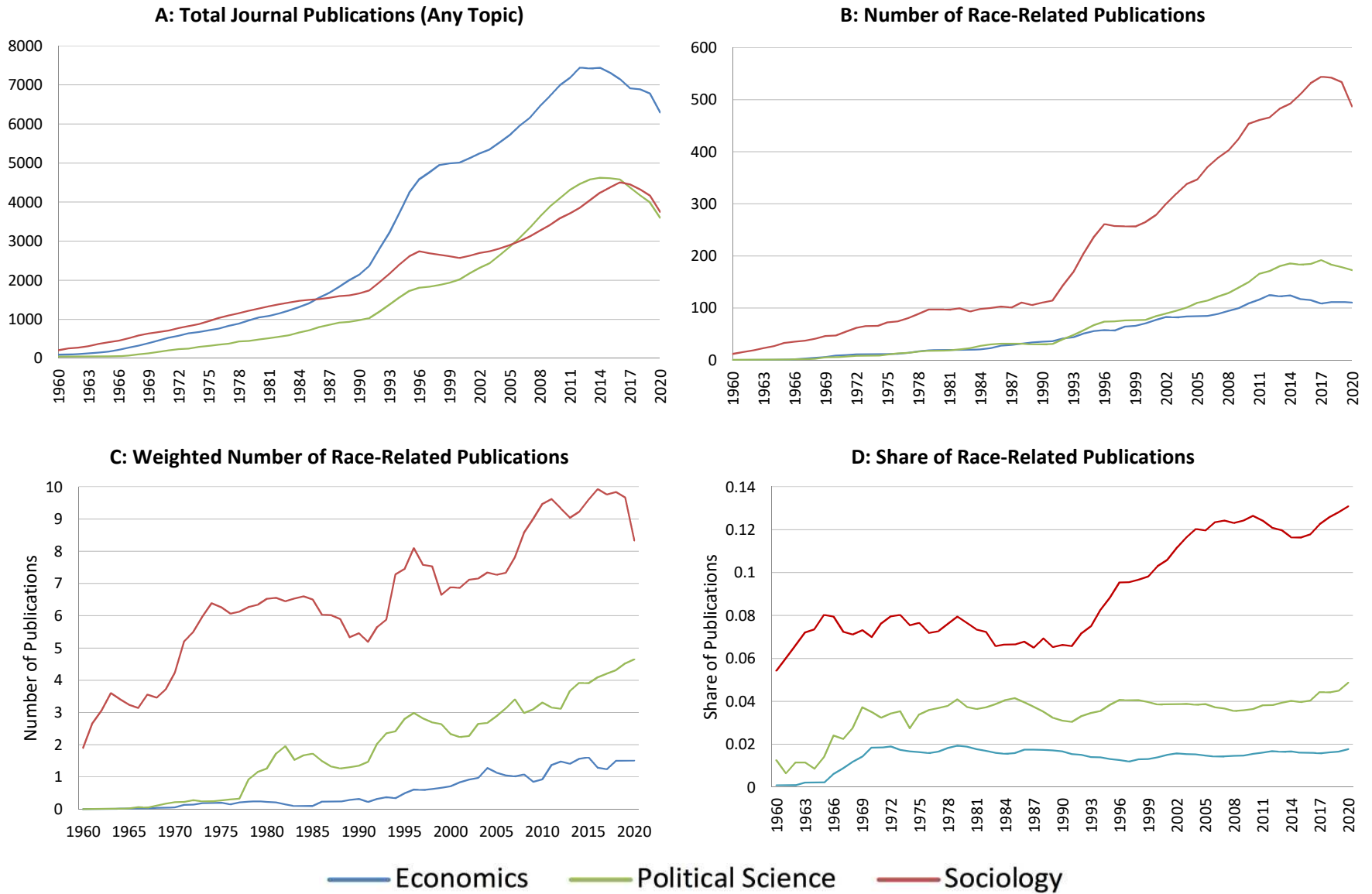
Each race/ethnicity keyword is divided into one of six groups: Non-Specific, African-American, Latinx, Native American, Asian, White, and Religious. Each race-related issue keyword is divided into one of five groups: discrimination, inequality, diversity, identity, and historical. The regular expressions for each pattern, as well as the division into categories, are listed in Tables A1 and A2 respectively.

## References

- [1] ADVANI.A, E.ASH, D.CAI AND I.RASUL (2020a) “The Study of Race in Economics, and Other Social Sciences,” Social Science Prediction Platform. <https://socialscienceprediction.org/s/dybafd>
- [2] ADVANI.A, E.ASH, D.CAI AND I.RASUL (2021a) On the Study of Race and Ethnicity in Economics: Production, Incentives and Selection, mimeo, UCL.
- [3] ANGRIST.J, P.AZOULAY, G.ELLISON, R.HILL AND S.FENG LU (2020) “Inside Job or Deep Impact? Using Extramural Citations to Assess Economic Scholarship,” *Journal of Economic Literature* 58: 3-52.
- [4] CHARLES.K (2021) Transcript of Remarks from, ‘What Can Economics do for Racial Justice?’ panel, *Econometric Society World Congress Monograph*.
- [5] COOK.L.D (2021) Transcript of Remarks from, ‘What Can Economics do for Racial Justice?’ panel, *Econometric Society World Congress Monograph*.
- [6] COOK.L.D AND T.LOGAN (2020) Racial Inequality, Economists for Inclusive Prosperity, Research Brief.

- [7] DERENONCOURT.E AND C.MONTIALOUX (2021) “Minimum Wages and Racial Inequality,” *Quarterly Journal of Economics* 136: 169-228.
- [8] 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.
- [9] KLEVEN.H (2018) Language Trends in Public Economics, mimeo, Princeton University.
- [10] KOFFI.M AND L.WANTCHEKON (2021) “Racial Justice from Within? Diversity and Inclusion in Economics,” *Econometric Society World Congress Monograph*.

**Figure 1: Race-Related Publications, by Year and Discipline**

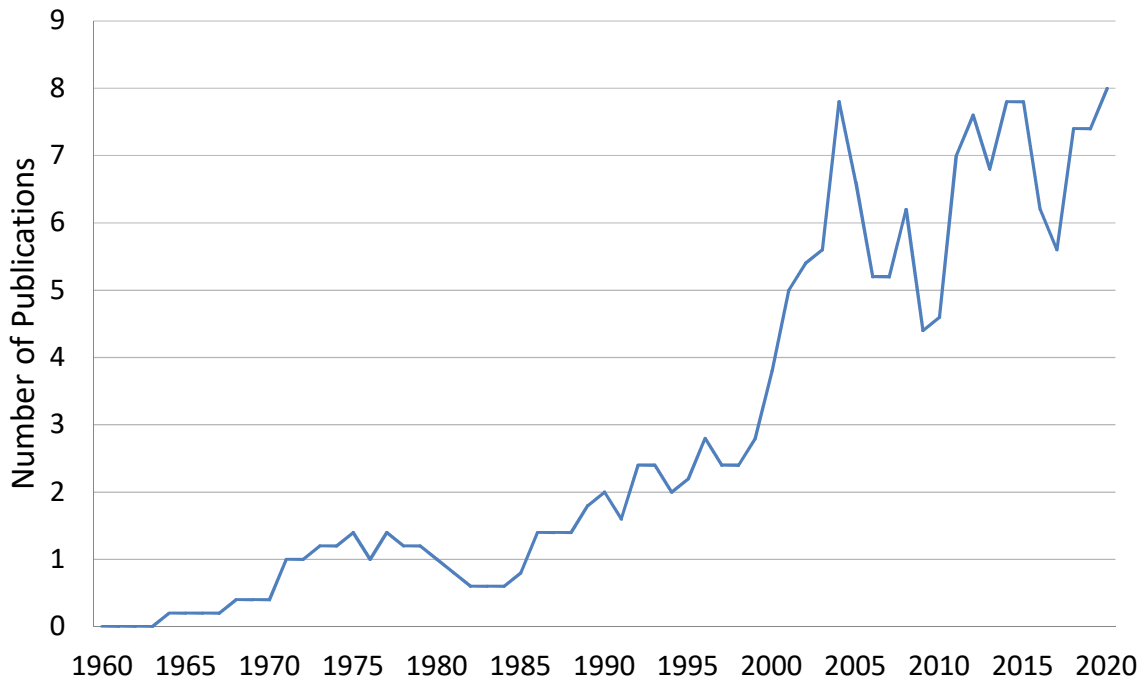


**Notes:** We use data from JSTOR, Scopus, and the Web of Science to construct the number and shares of race related publications in economics, political science, and sociology. Panel A reports the total number of publications in each discipline. As the publication series start in the 1880s, the publication numbers do not start exactly at zero in 1960, the first year of our working sample. Panel B reports the number of articles that are determined to be race-related by our algorithm. Panel C reports a journal-weighted version of Panel B using the journal quality weights from Angrist et al. [2020]. Panel D reports the share of articles determined to be race-related by our algorithm in each discipline. All series presented are 5-year moving averages.

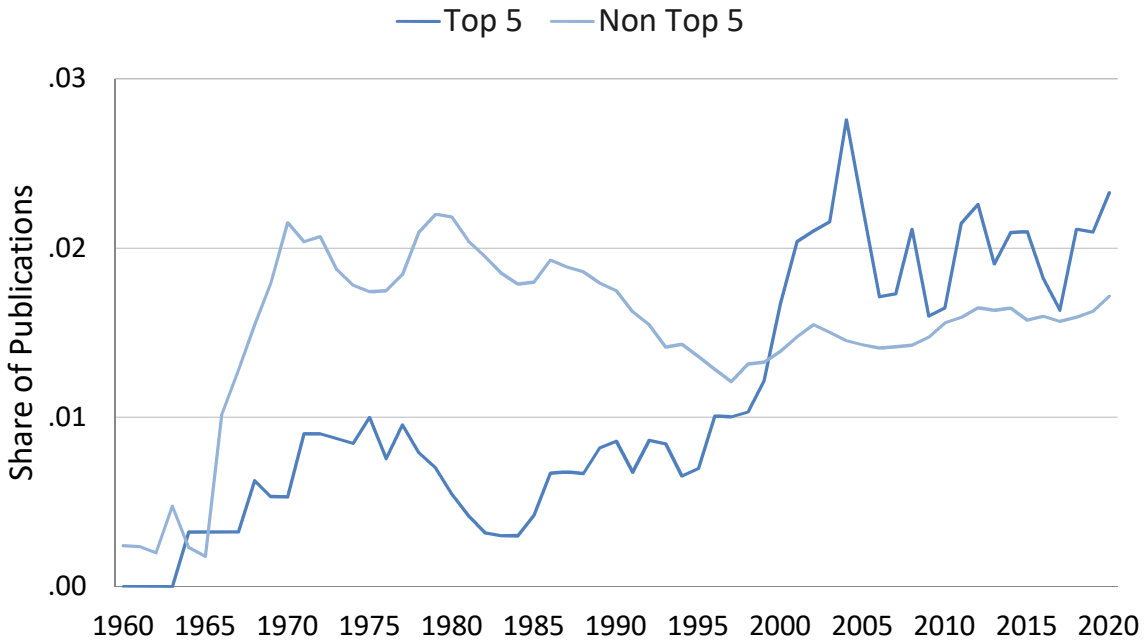


## Figure 2: Race Related Publications, Top-5 Journals

### A: Total Number of Race-Related Publications, Top-5 Journals



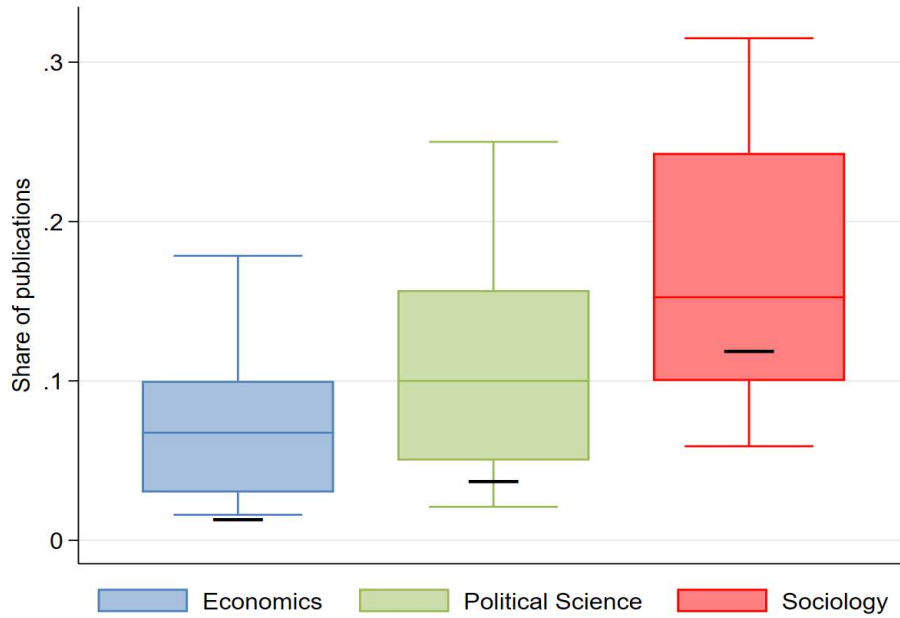
### B: Share of Race-Related Publications, Top-5 Journals vs. Other Journals



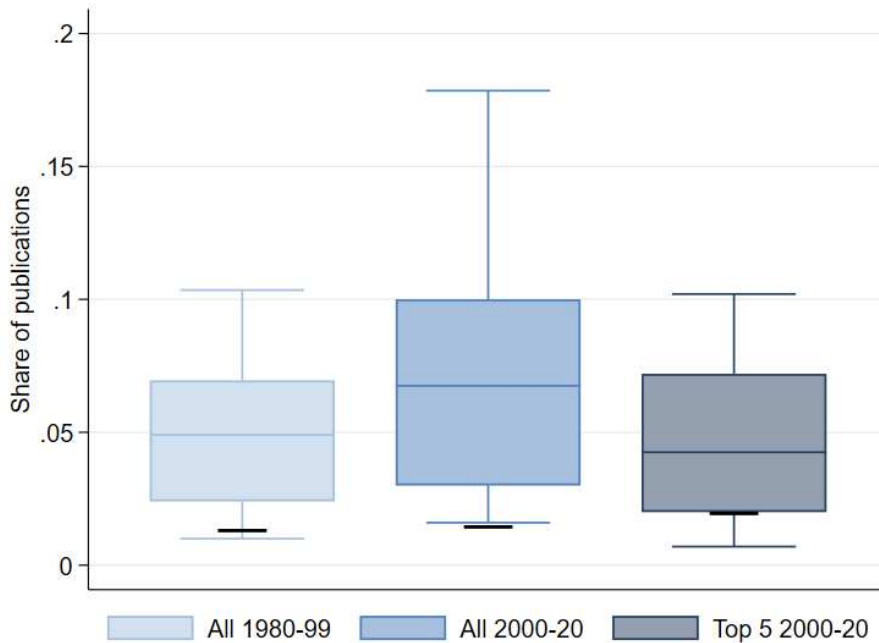
**Notes:** We show the presence of race-related research in Top-5 economics journals: the American Economic Review, the Journal of Political Economy, the Quarterly Journal of Economics, Econometrica, and the Review of Economic Studies. All series presented are 5-year moving averages.

**Figure 3: Social Science Prediction Platform Results (N=240)**

**A. Predicted Share of Race Related Research 2000-2020, by Discipline**



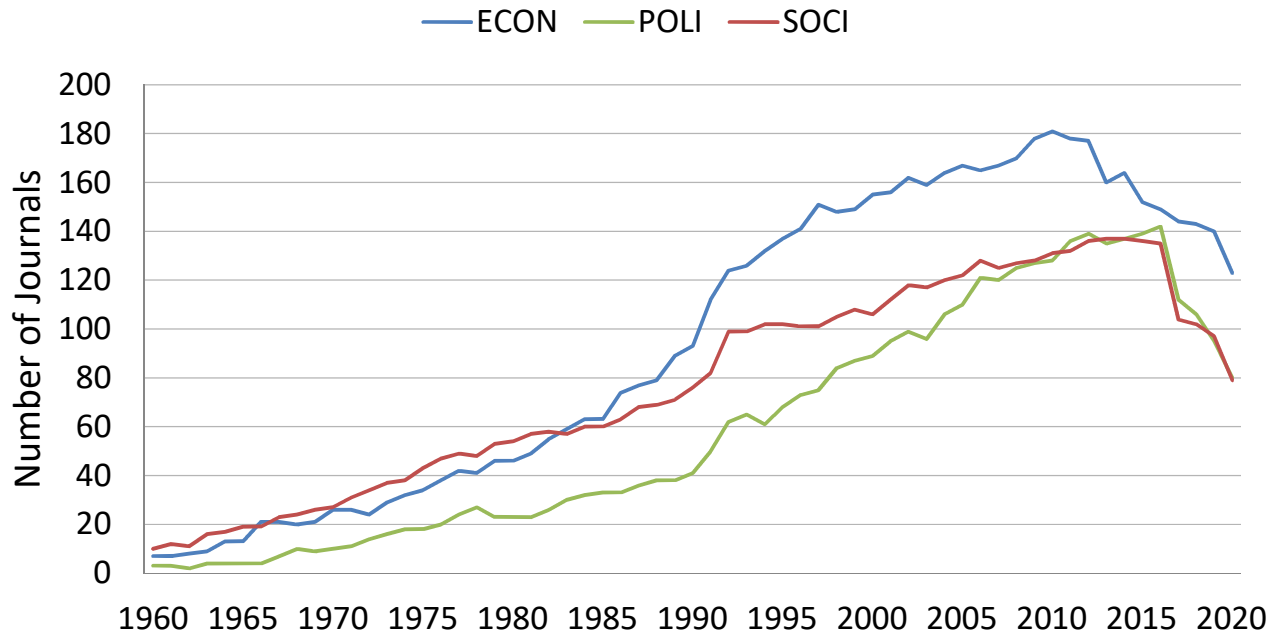
**B. Predicted Share of Race Related Research in T5 Economics Journals**



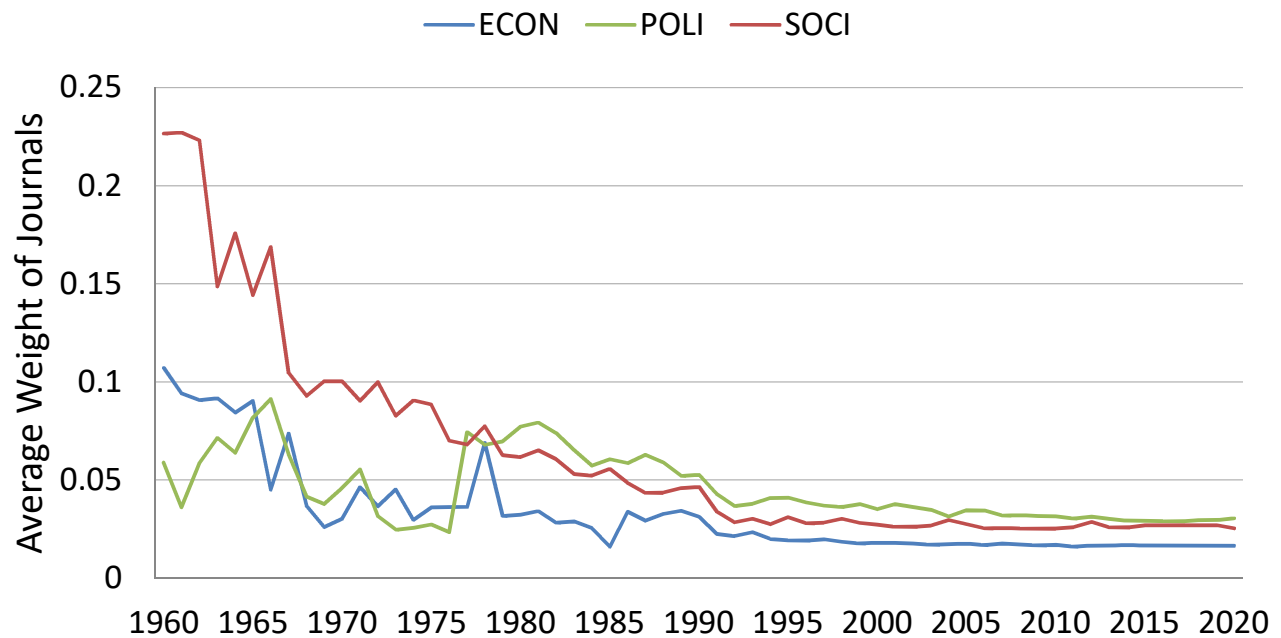
**Notes:** We summarize results from a survey we ran on the Social Science Prediction Platform. Questions were of the form "What share of articles published in [discipline/journal type] between [dates] do you think were race-related?", and were asked after an explanation of our definition of a race-related publication. Panel A compares responses across disciplines, covering all publications 2000-20 in those disciplines. Panel B looks within economics to compare over time and between all journals and Top 5 journals. The box-whisker plots show the 10th, 25th, 50th, 75th and 90th percentiles of responses. The black horizontal lines indicate true values as measured by our algorithm.

## Figure A1: Journal Coverage

### A: Number of Journals Covered, by Discipline



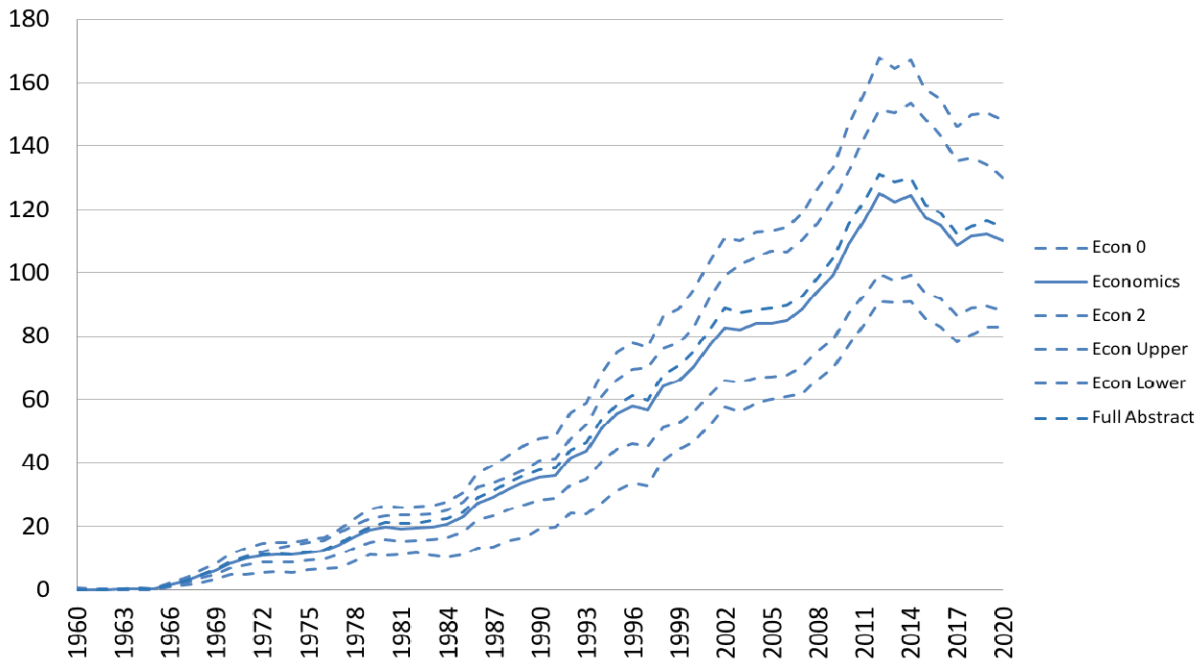
### B: Quality Weighted Journals Covered, by Discipline



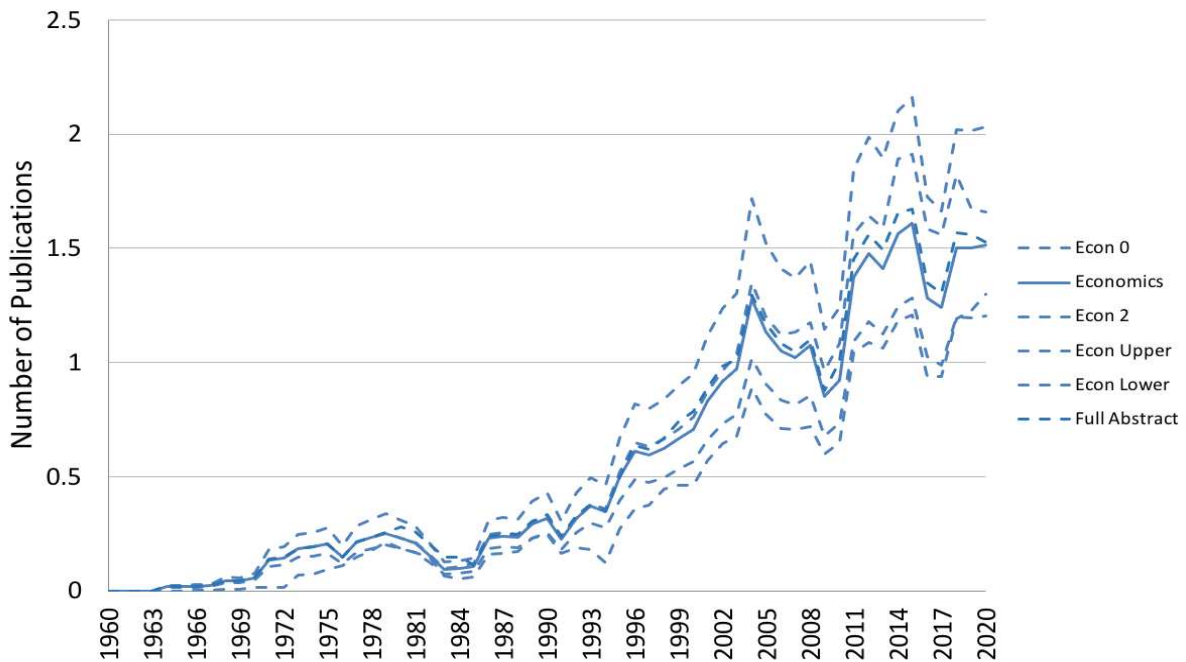
**Notes:** Panel A shows the number of journals observed in our corpus of publications each year for economics, political science, and sociology. Panel B plots the average journal weight across years and disciplines, derived from Angrist et al. 2020.

**Figure A2: Bounds on Number of Race Related Publications in the Social Sciences**

**A. Unweighted Number of Race-Related Publications**



**B. Weighted Number of Race-Related Publications**



**Notes:** We create upper and lower bounds according to the number of false positives and false negatives of our algorithm on all the articles from Top 5 economics journals. The thick dotted lines are lower/upper bounds while the thin dotted lines are results from the varying definitions of groups covered in algorithms. In Panel A, series Econ 0, Economics, and Econ 1 show how the number of articles we count as being race-related vary as we vary the number of group keywords included in the algorithm. The three series include groups in Band 0, 1, and 2 progressively. See Table A1 for the definitions of different bands. Series Econ Upper/Lower in Panel A show the bounds constructed using the frequencies of false positives and false negatives. Series Full Abstracts in Panel A relaxes the restriction that the group keywords must appear before the last sentence of an abstract. Panel B reports a journal-weighted version of Panel A using the journal quality weights from Angrist et al. [2020]. All series shown here are 5-year moving averages

# Table A1: Race/Ethnicity Group Keywords with Regular Expression Patterns

## Non-Specific - Band 0

rac[a-zA-Z]{0,3}  
 person[a-zA-Z]{0,1} of color[a-zA-Z]{0,1}  
 ethnic[a-zA-Z]{0,4}  
 underrepresented minorit[a-zA-Z]{0,3}  
 non-western[a-zA-Z]{0,1}

non-white[a-zA-Z]{0,1}  
 colored[a-zA-Z]{0,1}  
 caste[a-zA-Z]{0,1}  
 disadvantaged minor[a-zA-Z]{0,5}  
 people of colo[a-zA-Z]{0,1}r

advantaged-group[a-zA-Z]{0,1}  
 advantaged group[a-zA-Z]{0,1}  
 dominant group[a-zA-Z]{0,1}  
 dominant-group[a-zA-Z]{0,1}  
 people-of-colo[a-zA-Z]{0,1}r

## Main Ethnic Groups - Band 1

African-American

african-american[a-zA-Z]{0,1}  
 african american[a-zA-Z]{0,1}  
 black[a-zA-Z]{0,1}  
 negro[a-zA-Z]{0,2}  
 black-american[a-zA-Z]{0,1}  
 black american[a-zA-Z]{0,1}

Latinx

hispanic[a-zA-Z]{0,1}  
 latino[a-zA-Z]{0,1}  
 mexican-american[a-zA-Z]{0,1}  
 hispanic-american[a-zA-Z]{0,1}  
 latino-american[a-zA-Z]{0,1}  
 mexican american[a-zA-Z]{0,1}  
 hispanic american[a-zA-Z]{0,1}  
 latino american[a-zA-Z]{0,1}

Native American

native-american[a-zA-Z]{0,1}  
 american-indian[a-zA-Z]{0,1}  
 native american[a-zA-Z]{0,1}  
 american indian[a-zA-Z]{0,1}  
 native[a-zA-Z]{0,1}

## Less Prominent Ethnic and Religious Groups - Band 2

Asian

south asian[a-zA-Z]{0,1}  
 japanese-american[a-zA-Z]{0,1}  
 chinese-american[a-zA-Z]{0,1}  
 korean-american[a-zA-Z]{0,1}  
 vietnamese-american[a-zA-Z]{0,1}  
 indian american[a-zA-Z]{0,1}  
 indian-american[a-zA-Z]{0,1}  
 japanese american[a-zA-Z]{0,1}  
 chinese american[a-zA-Z]{0,1}  
 orientals

White

white[a-zA-Z]{0,1}  
 caucasian[a-zA-Z]{0,1}

Religious

jew[a-zA-Z]{0,3}  
 hebrew[a-zA-Z]{0,1}  
 yiddish  
 arab  
 muslim[a-zA-Z]  
 islam[a-zA-Z]

**Note:** [a-zA-Z]{0,k} indicates that we allow any 0 to k lower or characters to be matched.

**Table A2: Race-Related Issue Keywords with Regular Expression Patterns**

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

**Note:** [a-zA-Z]{0,k} indicates that we allow any 0 to k lower or characters to be matched.