Incentives to Produce Race-related Research*

Arun Advani Elliott Ash Anton Boltachka David Cai Imran Rasul

February 2025

Abstract

An established literature has studied potential biases in the economics publication process based on traits of authors. We complement such work by studying whether the subject matter of study relates to publication outcomes. We do so in the context of race-related research: work that studies economic well-being across racial/ethnic groups. We investigate the implicit career incentives economists have to work on such topics by examining paths to publication for a corpus of 22,056 NBER working papers (WPs) posted from 1974 to 2015. We use an algorithm to classify whether a given WP studies race-related issues. We then construct paths to publication from WPs to data on published articles, and compare paths for race-related WPs to various counterfactual sets of WPs. We document that unconditionally, race-related NBER WPs are less likely to be published in any journal, in an economics journal, and more likely to publish in lower tier economics journals. Once we condition on observable characteristics including field and author affiliations, differences in paths to publication largely disappear, and such work is actually slightly more likely to publish in top-tier economics journals. Consistent with unconditional differences in paths to publication being salient to researchers, we find evidence of *ex ante* selection into WPs studying race-related issues in that they are of higher readability than other WPs. To understand the interplay with selection of researchers, we compare results to paths to publications for 10,306 CEPR WPs posted from 1984 to 2015. We conclude by discussing implications for economists' incentives 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

Inequalities in well-being between racial and ethnic groups remain large and persistent. Yet their causes and solutions to address them remain highly contested areas of debate in many countries [Alesina *et al.* 2024, Platt *et al.* 2025]. Economists view themselves as having an important role to play in informing societal debates around horizontal inequalities, including those related to racial/ethnic gaps [Fourcade *et al.* 2015, Spriggs 2020, Maesse *et al.* 2022]. Our ability to do so depends on the foundation of research economists have produced relevant for studying economic well-being across racial/ethnic groups.

We refer to this body of work as race-related research. In earlier work, we studied the supply of race-related research in economics over six decades, using an algorithmic approach to classify publications in economics journals as either being race-related or not [Advani *et al.* 2025]. We documented that from nearly zero race-related publications in the early 1960s, the share of race-related publications in economics rose to a peak of 2.9% in the mid 1970s, fell to 1.6% by the mid 1990s, and has risen steadily thereafter to 2% today. This represents a cumulative body of knowledge of around 4000 race-related publications from 1960 to 2020.

We make no normative judgement on whether this volume is too high or too low, but if the publication process provides implicit incentives to economists to work on certain topics [Heckman and Moktan 2020], there can be important implications for the direction of innovation and knowledge generated from informing researchers about actual paths to publication for race-related research. Professional societies and associations in economics often spend resources to inform (junior) researchers how the wider publications process operates. In earlier work, we used data from the *Social Science Prediction Platform* to specifically document that economists overestimate the volume of race-related research published in economics journals and so might misperceive paths to publication for such work [Advani *et al.* 2024].[1]

In this paper, we examine the publications process in economics, discussing implications for the production of knowledge to inform debates on racial/ethnic gaps. Publications are the key metric along which career success is defined, carrying rewards in terms of hiring, pay, promotion and tenure. Yet, this is a process where there have long been expressed about biases against race-related work [Alexis *et al.* 2008], including those expressed by editors of leading journals [Omeokwe 2020]. We study the implicit incentives researchers have to work on race-related topics by examining paths to publication for working papers, comparing paths for race-related work relative to other work.

We consider paths to publication from two prominent working paper (WP) series: the NBER and CEPR. Our sample covers 22,056 NBER WPs posted from 1974 to 2015, and 10,306 CEPR

[1] Based on 300 responses from economists, we found they: (i) overestimate the share of race-related research in economics five-fold; (ii) overestimate the growth of race-related research in economics; (iii) incorrectly predict that the top-5 journals currently have lower shares of race-related publications than the discipline as a whole.

WPs first posted from 1984 to 2015. We use our algorithm to classify which WPs are race-related and construct paths to publication, by linking WPs to data on published articles, and compare paths for race-related working papers to various counterfactual sets of working papers.

Our main analysis focuses on paths to publication for the NBER series because this is the most prominent WP series, and produces more race-related content than the CEPR series. However, drawing inferences about the publications process for race-related research using the NBER WP series might be misleading because such work is produced by a group of large but select academics [Kleemans and Thornton 2021, Koffi and Wantchekon 2022]. We address concerns over researcher selection by considering paths to publication for CEPR WPs, produced predominantly by European scholars.

We document the following new insights. First, race-related NBER WPs are less likely to publish in any journal, in an economics journal, and publish in lower quality journals conditional on publishing in an economics journal, as measured by their *AER*-weight [Angrist *et al.* 2020]. However, once we condition on characteristics of WPs including field and author affiliations, most of these differences in paths to publication for NBER race-related WPs disappear. Moreover, conditional on observables, race-related NBER WPs experience no longer publication lags at the top-tier than other working papers, consistent with such papers not being held to a higher standard. Finally, we see no difference in citations between race-related and non race-related WPs, consistent with such papers not being held to a lower standard.

If unconditional differences in paths to publication are more salient to researchers, or they perceive such work to face greater publication risks, then a more positively selected sample of race-related working papers will be produced to begin with. To examine the issue, we follow Hengel [2022] and construct readability scores for each NBER WP. This reveals that race-related NBER WPs do indeed have significantly higher readability scores than non race-related work, consistent with *ex ante* selection into the production of race-related WPs.

Finally, the comparison of paths to publication between NBER and CEPR WPs is also informative of selection effects into the production of race-related work. CEPR WPs, produced predominantly by European scholars, are less likely to be race-related, and publication outcomes are slightly worse than for NBER WPs. Relative to non race-related WPs, we find that conditional on observables, race-related CEPR WPs are 3pp less likely to publish in an economics journal ($p = .050$), and conditional on publication, are published in significantly lower quality journals measured by their *AER*-weight ($p = .032$). These two margins might be especially prominent in researcher minds when deciding topics to research, and can potentially contribute to the lower share of race-related work produced in the CEPR WP series than the NBER WP series.

A concern over the interpretation of these findings is that the style of race-related WPs might differ from other work, for example in methodology or policy relevance. A premium to these traits in the publication process might mask penalties that race-related work otherwise faces, leading to the observed null results (with the opposite being true if there are publication penalties for

these other traits). We address this by considering alternative sets of counterfactual WPs rather than all non race-related working papers: (i) WPs that study similar topics as race-related work, such as inequality, but just not through the lens of racial/ethnic differentials; (ii) using machine learning to classify the topic of WPs and then control for these broad topics instead of *JEL* codes. In both comparisons our basic results on paths to publication continue to hold for both NBER and CEPR WP series.

Our work sheds new light on the ability of academic economists to contribute on a scientific basis to public debates on racial justice and the causes, consequences and solutions to persistent economic inequalities across racial/ethnic groups. By studying the implicit incentives provided by the publications process to work on such ideas, we complement an existing literature that has focused on how elements of the publication process might be biased against *individuals* based on their traits – such as gender or race, that ultimately feed into the under-representation of minorities [Lundberg and Stearns 2019, Bayer and Rouse 2016, Bayer *et al.* 2020]. In contrast, we focus on the issue of whether the *subject matter* of research influences its path to publication. To be the best of our knowledge, there are no comparable estimates of this in the economics discipline. Our findings provide valuable insights for helping the discipline understand how features of the publications process influence researcher choices of what to study, and hence the scientific knowledge accumulated in the discipline.[2]

2 Identifying Race-Related Research

2.1 Corpus

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 provides details on each series.[3]

2.2 Algorithm

Our intention is to identify work relevant for the study of the economic well-being of racial and ethnic groups, across countries and over time. We refer to this body of work as 'race-related' research. Our approach to doing so was developed in Advani *et al.* [2025], and we closely follow the description there. We use an algorithm to classify any working paper as being race-related or not. This uses keywords along two dimensions: (i) the racial or ethnic group being studied; and, (ii) the issue being studied. The 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

[2]Studies of minorities in economics have focused on barriers to entry, promotion and publication [Collins 2000, Price 2009, Bayer and Rouse 2016, Bayer *et al.* 2020, Logan and Myers 2020, Slater 2020].

[3]The NBER currently has around 1600 members, the CEPR has over 1700 members (with some overlap in membership). WPs with at least one co-author in the network are eligible to publish in the networks' series.

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*.[4]

The full lexicon of group keywords is in Table A1. We define alternative bands of group keywords that gradually expand the racial/ethnic groups picked up by the algorithm. 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 specific group. Band 1 adds another 19 group base keywords relating to the main minority groups in the US (*African American*, *Latino* and *Native American*). Our core results are based on combining the 35 group keywords in Band 0 and Band 1. 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. Finally, we exclude publications containing any of the eliminated phrases from 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.

Advani *et al.* [2025] discuss: (i) potential rates of misclassification errors; (ii) validation of the algorithm through a comparison with classifying publications using a large language model; (iii) validation of the algorithm using unsupervised topic modelling methods to classify study areas, including the study of issues related to race/ethnicity. The algorithm is not designed to capture the universe of race-related research and some gray areas remain (e.g. topics related to immigration). However our algorithm is easily replicable and refinable.

2.3 Aggregate Trends in Race-Related Research

To give context to the importance of the publications process shaping the flow of race-related research, we briefly present evidence on aggregate trends in the production of such work from Advani *et al.* [2025]. Panel A of Figure A1 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%. Panel B shows the time series of race-related publications in economics journals from 1960 to 2020. In the most recent decade, close to 2% of all publications were race-related. 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; (ii) the uptick in the share of race-related research in the NBER and CEPR WP series slightly predates the uptick in race-related journal publications in the mid-1990s.

[4] All keywords for classification purposes are considered in a case-insensitive manner and wildcards are used to capture different word spellings or forms (e.g. American and British English spellings).

Advani *et al.* [2025] discuss other trends in the production of race-related research, including: (i) the fields in which such work has been published; (ii) how the group and topics (such as discrimination) studied in race-related research have changed over time. In this paper our focus is on paths to publication, where aggregate trends and field differences are conditioned out. We later examine paths to publication by field, and in the Appendix we discuss how paths to publication vary by the group or topic studied.

2.4 Matching Working Papers to Publications

To study paths to publication we link WPs to publications as follows. For each WP we find all articles in the *Web of Science* or *Scopus* databases with the same coauthors published after the WP release date, and compute the string distance between this set of published papers and the WP (so retrieving titles similar to both the WP and published article). To capture all possible matches between WPs and published articles, we intentionally set the match similarity threshold to 50 score points. This helps avoid missing potentially true matches caused by spelling errors in names/titles. A single WP may have multiple matches above the set threshold. When multiple matches are found, we retain the WP-publication pair with the highest similarity. More than 90% of all matched pairs have a similarity score above 95, and approximately 80% of all matches have a perfect score of 100 points. Panel A of Figure A2 presents the distribution of string similarity among matched NBER and CEPR WPs. This evidences highly accurate matches between WPs in each series and published articles. Panel B shows how match rates vary by publication years using progressively stricter thresholds for each WP series.

3 Race-Related Research and the Publication Process

We now document paths to publications for race-related research. We focus primarily on the NBER series because this is the most prominent WP series, and it produces more race-related content than the CEPR series. Results for the CEPR series are presented afterwards, and help inform whether the same insights apply across networks of US- and Europe-based researchers.

3.1 Descriptives

Table 1 presents descriptives comparing paths to publication for race-related NBER WPs and non race-related NBER WPs. We identify 888 WPs posted between 1974 and 2015 as race-related. Panel A focuses on publication outcomes. 63% of non race-related WPs are published in an academic journal within the *Web of Science* or *Scopus* catalogs, and this likelihood is 3pp lower for race-related WPs ($p = .081$). Moreover, race-related WPs are 4pp less likely to be published in an economics journal, rather than in a journal from another discipline ($p = .028$), and this

remains true even conditional on them being published in a journal in any discipline ($p = .007$). The two types of WP however have similar publication lags, of around 2.4 years ($p = .473$). This evidence suggests that, unconditional on all other factors, on a number of margins of publication outcome, race-related research fares worse than other work.

Panel B focuses on publication quality. Given our study period has witnessed changing journal influence, we adjust for journal quality using the journal weighting scheme employed by Angrist *et al.* [2020], which produces an *AER*-equivalent weight for each journal in each year. We see that the average *AER*-weight of journals published in is significantly lower for race-related WPs ($p = .002$). Examining the tails of the distribution of journal quality, the next two rows show that: (i) race-related publications are less likely to be published in a journal with zero *AER*-weight ($p = .017$); (ii) race-related publications are no less likely to be published in the top-tier general interest journals (the top-5) ($p = .361$).[5]

To probe further how the likelihood of publication of race-related and non race-related research varies by journal quality, Figure 1 shows the unconditional difference in the likelihood that race-related and non race-related WPs are published in journals: (i) in the top-5; (ii) in the top, middle and lower terciles of journals with positive *AER*-weight; (iii) in a zero *AER*-weight journal. Panel A shows that for NBER WPs there is a bimodal distribution of outcomes in publication quality, where relative to non race-related WPs, race-related WPs are more likely to be published in low (but not zero) *AER*-weight journals, but also more likely to be published in the top-tier. Given these diverse outcomes, researchers might perceive there to be more publication risk associated with race-related research than other work.

The final row in Panel B of Table 1 consider citations as a measure of publication quality, although these also matter for reputation, and decisions related to hiring, promotion and funding [Ellison 2013, Koffi 2021]. While the *AER*-weight reflects the decisions of editors and referees, citations are determined by the discipline as a whole. Despite the bimodal distribution of journal quality where race-related NBER WPs are published, such research is not differentially cited from other WPs ($p = .635$).

[5] 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:

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

Economics journals that are not covered in Angrist *et al.* [2020] are given a zero weight.

3.2 Estimation

To establish whether these unconditional differences are robust, we estimate the following OLS specification for publication outcome y for NBER WP a first posted in year t:

$$y_{at} = \beta RR_a + \theta X_a + \alpha_t + \sum_{j \in J(a)} \alpha_j + \sum_{s \in S(a)} \alpha_s + \varepsilon_{at}, \tag{1}$$

where RR_a is a dummy for whether the WP is classified as race-related, X_a are WP characteristics, α_t are fixed effects for the year in which the WP is first posted, α_j are *JEL* codes the WP refers to (so $J(a)$ refers to the set of *JEL* codes for WP a), α_s are dummies for the institution affiliation of each author WP (the set $S(a)$ is the affiliations of all co-authors), and ε_{at} is an error term. We treat outcomes for WPs to be independent and report robust standard errors.[6]

The parameter of interest is β: the difference in publication outcome y_{at} for race-related and not race-related WPs, conditional on WP characteristics, publication time, field and author affiliations. The covariates are not necessarily all confounders, but may include controls mediating or colliding with the outcomes. Hence their inclusion is informative of robustness but do not give the estimates any causal interpretation. Counterfactual WPs are those in the same field (as measured by *JEL* classifications), so studying similar research topics but just not through the lens of race or ethnicity. We consider alternative sets of counterfactual WPs below.

4 Results

4.1 Publication Outcomes

We first consider publication outcomes. Columns 1 to 3 in Table 2 show that once we condition on date of posting, *JEL* codes, WP characteristics and author affiliation dummies, the unconditional differences in publication outcomes between race-related NBER WPs and not race-related WPs disappear: race-related NBER WPs are not differentially likely to be published in any journal

[6] X_a includes the number of pages (and its quadratic), the title length (and its quadratic), the number of authors and *JEL* codes. There are 20 unique top-level *JEL* codes, α_j. When a WP has multiple codes, we split the assignment equally across listed codes. Panel A of Table A4 shows there are significant differences in the length, titles and number of *JEL* classifications of race-related WPs relative to others. However the magnitude of these differences are small. Panel B examines the group and topic content of race-related WPs. Non race-related WPs rarely mention any of the group keywords, and rarely relate to the topics our algorithm is based on except inequality, where 24% of non race-related WPs mention some of the keywords under this heading (shown in Table A2). Information on institutional affiliation is derived from the *Scopus* database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the author affiliation in *Scopus* with an economics publication that shares the same first and last name as the NBER WP author. The selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. To account for possible measurement error, we control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER WP authors are found in two thirds of cases. We define the α_s dummies to cover the 100 most frequent institutions, a dummy for other affiliations and a dummy for no matched affiliation.

in the *Web of Science* or *Scopus* databases, differentially likely to be published in a economics journal, or to differ in their publication lag in economics journals (Columns 1 to 3). These null impacts are precisely estimated. For example, the 95% confidence interval for $\widehat{\beta}$ rules out the publication lag for such work being more than .14 years longer than for other work (relative to a baseline of 2.7 years).

Columns 4 to 6 of Table 2 focus on publication quality, conditional on the WP being published in an economics journal. Race-related NBER WPs are not published in journals of differential quality as measured by their *AER*-weight, including zero weights. At the tails of the distribution of journal quality we see that race-related WPs are significantly less likely to be published in a journal with zero *AER*-weight ($p = .007$). To reconcile this with the null impact in Column 4 it is thus the case that race-related papers are more likely to be published in low (but not zero) *AER*-weight journals. At the same time Column 6 shows they are significantly more likely to be published in a top-5 journal ($p = .086$). The magnitude of this last effect is 3.9pp, corresponding to a 16% increase over the baseline likelihood of non race-related NBER WPs being published in top-5 journals conditional on them being published in an economics journal.

Column 7 shows publication lags for WPs published in top-5 journals are not different between race-related and other WPs. This is consistent with such papers not being held to a higher standard as proxied by longer refereeing processes. Finally, we consider citations for published WPs (where citations accumulate over both WP and published versions and we note that 98% of non race-related WPs are ever cited). We see no difference in total citations between race-related and non race-related WPs – a result robust to controlling for journal fixed effects (Column 9). All this points to such work not being held to a lower standard of publication.[7]

If the publication process provides implicit incentives to scholars to work on certain topics [Heckman and Moktan 2020], and individuals are perfectly informed of these features of paths to publication for NBER WPs, then our results provide little evidence that demand-side processes should discourage NBER-affiliated researchers from working on such topics. However, if unconditional differences in paths to publication are more salient to researchers, or they consider there to be higher publication risk in pursuing race-related work because it is more likely to publish in low *AER*-weight journals, they might be disincentivized to work on such issues. We come back to this when examining evidence of *ex ante* selection into the production of race-related WPs.

[7] One concern is that our results are biased if NBER WPs are only posted once they are accepted for publication. To check for this, we repeat the analysis restricting the sample to WPs with a publication lag of at least one year. We find the differential likelihood of being published in a zero *AER*-weight journal becomes smaller (but still statistically significant at the 10% level) and the differential likelihood of being published in a top-5 journal is 3.1pp but not significantly different between race-related and non race-related WPs.

4.2 Fields, Groups and Topics

Our earlier work documented a balkanization of race-related research into a few fields, and with such work being largely absent from many others [Advani *et al.* 2025]. We use this variation to narrow in on publication outcomes and the supply of race-related work by field. To do so we consider the strongest incentive for researchers, the likelihood of a race-related WP being published in a top-5 journal. We then estimate (1) for this outcome separately for each *JEL* code, obtaining $\widehat{\beta}_j$ estimates for *JEL*-code j. Panel A of Figure 2 plots for each *JEL* code: (i) the unconditional probability of a race-related NBER WP being published in the top-5; (ii) the conditional estimate, $\widehat{\beta}_j$. We overlay this with a histogram of the share of WPs in the *JEL*-code that are race-related, ordering fields by increasing share of race-related WPs.

Three points are of note. First, across *JEL*-codes, the unconditional probability of a race-related NBER WP being published in the top-5 does not vary substantially; nor does the conditional estimate, $\widehat{\beta}_j$. Second, there is only a weak relationship between this probability and the share of race-related research actually produced under any given *JEL*-code. Hence there is little responsiveness to these publication incentives on the supply of race-related research by field. Third, the same very weak relationship exists in terms of the *AER*-weighted quality of journal publications for race-related work (Panel B).

For each WP the algorithm classifies as race-related, we use the keywords to pinpoint which minority groups and topics are studied – including for studies of discrimination [Bohren *et al.* 2020]. In the Appendix we examine whether paths to publication for race-related WPs vary depending on the group or topic studied.

4.3 Counterfactual Working Papers

We have so far compared race-related WPs to non race-related WPs, conditional on date of WP posting, *JEL* code, WP characteristics and author affiliation. A concern might be that even within such bands, the style of race-related WPs differs – for example in methodology or policy relevance. A premium to these traits in the publication process might then mask any penalties that race-related work otherwise faces, leading to the observed null results. To address this, we consider alternative approaches to identifying counterfactual WPs. First, we consider not race-related WPs that have at least one of the topic keywords (Table A2) in their title and/or abstract. As Panel B of Table A4 shows, this counterfactual mostly includes WPs studying inequality, just not through the lens of racial/ethnic differentials. Second, we use an unsupervised topic model (LDA) to classify the topic of WPs and then control for these topics instead of *JEL* codes.[8]

[8] We use a LDA model to identify topics. To determine the optimal number of topics, we analyze a combination of coherence score and perplexity measures across models with different numbers of topics, manually inspecting the word distribution for each topic and model. For our benchmark model, we choose 30 topics. Figure A3 displays word clouds for these topics.

Table 3 shows the results for our key outcomes with these alternative non race-related comparison groups. We generally find null results throughout: on no margin do we find significantly worse outcomes for race-related papers. We continue to find evidence that such work is subject to more publication risk in the sense that it is more likely to be published in low (but not zero) *AER*-weight journals – despite such work not being differentially cited.

4.4 Selection into Working Papers

The null relationship between race-related research and publication outcomes might reflect differential selection in an earlier stage of the process: from ideas to then writing working papers. If researchers believe race-related work is less likely to be published in an economics journal or faces greater publication risks, then a more positively selected sample of race-related working papers will be produced to begin with. Our earlier findings that race-related WPs are not subject to differential publication lags and citation rates may then reflect an equilibrium outcome, caused by *ex ante* positive selection combined with race-related research being held to a higher standard *during* the publications process.

To examine the issue of *ex ante* selection into race-related WPs more directly, we consider the readability scores of produced working papers. To do so, we follow Hengel [2022] and construct three readability scores for each NBER WP abstract: the Flesch Reading Ease, Gunning Fog and Simple Measure of Gobbledygook (SMOG) metrics. We code each metric so that higher values correspond to material that is easier to read. All scores are standardized so coefficients can be interpreted in effect sizes. We then consider these as outcomes in equation (1).[9]

The results are in Table 4. Columns 1 to 3 show that for all three measures, readability scores of race-related NBER WPs are significantly higher than for other WPs. This is consistent with there being *ex ante* selection into the production of race-related NBER WPs.

To examine how readability scores change through the path to publication, Columns 4 to 6 repeat the analysis but based on the readability score of the publication (not the original WP). Here we find a weaker pattern of results, but it remains the case that on the Flesch Reading Ease measure, race-related publications have a significantly higher readability score than other publications, with the magnitude of impact being $.15\sigma$. Columns 7 to 9 check whether readability scores of race-related NBER WPs change moving from original working papers to published articles. The estimates on each change are noisy and never statistically significant, although each point estimate is negative suggesting that non race-related WPs catch up in quality of readability to race-related work during the publications process.

Given the *ex ante* selection of race-related NBER WPs, in Table A5 we examine how readability

[9]We use the `Textatistic` library, a Python package for estimating readability metrics [Hengel 2022]. `Textatistic` employs a combination of algorithms to first count the number of sentences, characters, syllables, words, words with three or more syllables, and words not on a predefined list of easy words. It then calculates readability indices using this information.

scores correlate to publication outcomes. On some margins these do matter differentially for race-related work, suggesting it can be held to differential standards in paths to publication: race-related WPs of higher than average readability are more likely to publish in any journal ($p = .080$) and have shorter publication lags at economics journals ($p = .057$).

4.5 Selection of Researchers

Drawing inferences about the publications process for research using the NBER WP series might be misleading because such work is produced by a group of large but non-randomly selected academics [Kleemans and Thornton 2021, Koffi and Wantchekon 2022]. We address concerns over researcher selection by considering paths to publication for CEPR WPs, produced predominantly by European scholars. These are less likely to be race-related, and their publication outcomes are generally slightly worse than for NBER WPs.

Panel B of Figure 1 shows the unconditional difference between the likelihood that race-related and non race-related CEPR WPs being published in journals at various tiers. They are more likely to publish in low or zero *AER*-weight journals, and are less likely to publish in higher weight journals or in the top-5. The comparison for paths to publication of CEPR and NBER WPs is therefore informative of how race-related research from different tiers of the discipline fares in the publication process.

Our baseline results on paths to publication for race-related CEPR WPs are in Table A6. Conditional on observables, race-related WPs are 3pp less likely to publish in an economics journal ($p = .050$), and publish in significantly lower quality journals measured by their *AER*-weight ($p = .032$). These two margins might be especially prominent in researcher minds when deciding topics to research, and can potentially contribute to the lower share of race-related work produced in the CEPR WP series than the NBER WP series. On other margins – such as publication lags or citations, we find a similar pattern of null results for CEPR WPs as for NBER WPs.

Table A7 confirms these conclusions to be robust to alternative counterfactual sets of CEPR WPs. Finally, we consider whether there is evidence of *ex ante* selection into the production of CEPR WPs using readability scores. The results in Table A8 differ from those for NBER WPs: we find no difference in readability between race-related and other CEPR WPs. We continue to find no evidence that the change in readability scores between published and WP versions of research differs between race-related and other work.

Overall, for WPs predominantly produced by European economists, paths to publication for race-related research are slightly worse than for those selected into the NBER network. This might be due to there being less *ex ante* selection of such WPs based on their quality.

5 Discussion

Economists have an important role to play in informing societal debates, include understanding gaps in economic well-being across racial and ethnic groups. Our ability to do so depends on the scientific foundation of race-related research that economists have collectively produced. We shed light on career incentives to produce such work as measured by paths to publication for race-related research from working paper to academic journals.

We present a nuanced set of results: unconditional on observables, race-related NBER WPs are less likely to publish in an economics journal, publish in lower *AER*-weight journals conditional on publishing in an economics journal, and have more varied publication outcomes – being both more likely to publish in low (but not zero) *AER*-weight journals, but also more likely to be published in the top-tier. However, once we condition on observables, we find most of these differences in paths to publication for NBER race-related WPs largely disappear.

If unconditional differences are salient to researchers, or they perceive race-related research to face more publication risk, they might be deterred from working on such issues. In line with this, we document evidence of: (i) higher *ex ante* selection into the production of race-related NBER WPs measured by such WPs having higher readability scores than non race-related NBER WPs; (ii) readability scores mattering differentially for race-related work for some publications outcomes. The fact the all else equal race-related work from the CEPR WP series have worse publication outcomes than counterfactual WPs might especially deter European-based researchers working on such topics.

Our findings raise the issue of whether there is scope to increase the number of economics journals specialized in race-related research. The most prominent such journal is the *Review of Black Political Economy*, which was launched in response to concerns that economics journals were not open to publishing research on the political economy of race [Alexis *et al.* 2008]. However, a key issue that remains is understanding whether such specialization would lead to a balkanization of race-related research, as publications in other journals may be less likely to cite such work, and the broader ideas from race-related work then do not filter through to other parts of the discipline.

Finally, paths to publication might reflect differential selection at an earlier stage of the publications process beyond that captured in readability scores. This could arise from two sources. First, in the step from ideas to working paper, the provision of funding can be key. Cruz-Castro *et al.* [2022] document that US minorities are less likely to succeed in funding applications relative to White individuals. Whether this impacts the production of race-related work remains unknown. Second, if there is a link between researcher's identity and what they study, then the entry of minorities into the economics academy might be important. For example, Mason *et al.* [2005] show that papers with at least one Black author are more likely to report a finding of racial discrimination than papers with no Black authors. Links between identity and the direction of research have been documented for inventors [Einiö *et al.* 2023] and medical scientists [Dossi

2024]. Our findings lead naturally back to the pipeline problem, a margin along which initiatives of economic associations, such as the *AER*, *EEA* and *RES*, are heavily directed [Bayer and Rouse 2016, Bayer *et al.* 2020], raising the possibility that such initiatives can have long-run effects on the direction of economics research.

A Appendix

A.1 Working Papers Data Sources

For the NBER series, we construct a corpus starting from 28,206 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. 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 Groups and Topics Studied

To examine how paths to publication for race-related WPs vary by the group or topic studied, we estimate heterogeneous effects of race-related NBER WPs for the outcomes considered above. Table A9 presents the results on groups, where the omitted category is non race-related WPs. On the whole, there are relatively few differences in publication outcomes for race-related NBER WPs focused on different groups. However, two notable results emerge.

First, NBER WPs focused on Black groups are more likely to be published in a journal relative to non race-related WPs ($p = .033$), and relative to race-related research on non-specific groups ($p = .072$) and all other groups ($p = .016$). Second, the likelihood of being published in a top-5 journal (conditional on being published in an economics journal) differs depending on the minority group being studied. As Column 6 shows, WPs studying Black groups are 20pp significantly more likely to be published in the top-5 than those studying all Other groups ($p = .071$).

Table A10 conducts a similar analysis for race-related NBER WPs based on their topic of study, using the five-way classification: discrimination, inequality, diversity, identity, and historical issues. Overall, topics of study matter more for publication outcomes (Columns 1 to 3) than for publication quality (Columns 4 onwards).

More specifically, Column 1 shows how the likelihood to be published in a journal varies by topic. Comparing race-related research to non race-related work, no race-related topic is significantly less likely to be published than non race-related WPs. Race-related studies on discrimination and identity are more likely to be published than non race-related WPs. Within race-related topics, studies on inequality are less likely to be published than studies on discrimination ($p = .046$). As the majority of race-related papers study inequality, this can reinforce a misperception among researchers that race-related research is generally less likely to publish well. In contrast, studies of discrimination are 11.4pp more likely to be published than non race-related WPs, although these constitute a steadily smaller share of race-related research over time [Advani *et al.* 2025].

Column 2 shows the likelihood of being published in an economics journal does not differ much over topics with the exception of race-related studies of identity: these are 8.4pp more likely to be published than non race-related WPs ($p = .014$) and significantly more likely to be published in an economics journal than studies on discrimination ($p = .011$), inequality ($p = .013$), or diversity ($p = .069$). Race-related studies on inequality and identity also have significantly shorter publication lags than non race-related WPs (Column 3).

Narrowing in on race-related research on discrimination, we see little evidence of differential paths to publication with other race-related topics. The evidence does not suggest such studies are less likely to be published in an economics journals, or be held to systematically higher or lower standards, as proxied by publication lags and citations, in the pathway from working paper to publication. This is important given the potential concern that studies of discrimination in economics are hard to publish because of the conventional null of there being no discrimination, and hence the onus being to show the existence of discrimination.

References

[1] ADVANI.A, E.ASH, D.CAI AND I.RASUL (2025) Race-related Research in Economics, CEPR DP19898.

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

[3] ALESINA.A, M.FERRONI AND S.STANTCHEVA (2024) Perceptions of Racial Gaps, Their Causes, and Ways to Reduce Them, NBER WP29245.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] BOHREN.A, K.HAGGAG, A.IMAS AND D.POPE (2020) "Inaccurate Statistical Discrimination," *Review of Economics and Statistics*, forthcoming.

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

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

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

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

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

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

[15] 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.

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

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

[18] KOFFI.M (2021) Innovative Ideas and Gender Inequality, mimeo, Toronto.

[19] KOFFI.M AND L.WANTCHEKON (2022) "Racial Justice from Within? Diversity and Inclusion in Economics," *Econometric Society Monograph*, forthcoming.

[20] 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.

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

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

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

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

[25] PLATT.L, T.PRATYUSH AND I.RASUL (2025) Majority Perceptions of Minority Groups: Economic Inequalities, Their Causes, and Policy Solutions, mimeo UCL.

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

[27] 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.

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

Table 1: Race-related NBER Working Papers

Means, standard deviation in parentheses, p-values in brackets

	(1) Race-related	(2) Not Race-related	Test of Equality [p-value]
A. Publication Outcomes			
Published in any journal	.603	.634	[.081]
Published in an economics journal	.587	.628	[.028]
Published in an economics journal | published in any journal	.973	.988	[.007]
Publication lag (years)	2.36	2.44	[.473]
	(2.42)	(2.95)	
Publication lag (years) | published in an economics journal	2.54	2.65	[.486]
	(2.67)	(3.11)	
B. Publication Quality			
Journal quality (*AER*-weight) | published in an economics journal	.050	.067	[.002]
Published in *AER* weight zero journal | published in an economics journal	.136	.182	[.017]
Published in Top-5 | published in an economics journal	.264	.244	[.361]
Total citations | published in an economics journal	96.2	102.3	[.635]
	(154)	(256)	

Notes: The sample is based on NBER working papers first posted between 1974 and 2015. Columns 1 and 2 show means and standard deviations in parentheses for working papers classified as race-related and not race-related respectively. Column 3 shows the p-values from a t-test of equality of means between race-related and not race-related articles, based on a regression with robust standard errors. In Panel A, published working papers are published in any outlet, published in any journal is if the working paper can be matched to a journal article in *Web of Science* or *Scopus*. The publication lag is the number of years between when the NBER working paper is first posted and its year of publication. In Panel B, journal quality is based on the weighting scheme used in Angrist *et al.* [2020]. Total citations are the number of citations received by an article in either the *Web of Science* or *Scopus*.

Table 2: Paths to Publication for NBER Working Papers

OLS estimates, robust standard errors in parentheses

	Published in any Journal	Published in an Economics Journal	Publication lag (years)	Journal Quality (AER-Weighted)	Published in AER zero weight journal	Published in Top-5	Publication Lag (years) | Published in Top-5	Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race-related	-.001	-.011	-.145	.004	-.049***	.039*	-.026	.098	.029
	(.020)	(.008)	(.142)	(.004)	(.018)	(.022)	(.191)	(.063)	(.057)
Outcome mean (sd)	.734	.988	2.65 (3.10)	.066 (.104)	.181	.244	2.28 (2.24)	3.64 (1.46)	3.65 (1.46)
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
WP Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Journal FE									X
Sample	All Published NBER WPs (N=19,070)	All NBER WPs Published in Any Journal (N=13,995)	NBER WPs Published in Econ Journal (N=13,822)	NBER WPs Published in Econ Journal (N=13,822)			NBER WPs Published in Top 5 (N=3,377)	NBER WPs Published in Econ Journal (N=13,511)	NBER WPs Published in Econ Journal (N=13,416)

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on NBER working papers first posted between 1974 and 2015. In Column 1 the outcome is a dummy for whether the working paper is published in any journal. In Column 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Column 3 the publication lag is the number of years between when the NBER working paper is first posted and its year of publication. In Column 4 the outcome is the AER-weighted measure of journal quality constructed in Angrist et al. [2020] (including zeroes). In Column 5 the outcome is whether the AER-weighted measure of journal quality constructed in Angrist *et al.* [2020] is zero. In Column 6 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Column 7 the outcome is the publication lag is the number of years between when the NBER working paper is first posted and its year of publication in a top-5 economics journal. In Columns 8 and 9 the outcome is the total number of citations received by an article since publication, as measured from the *Web of Science* or *Scopus*. All specifications include fixed effects for the year in which the working paper is first posted, and its JEL codes. Working paper characteristics include a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes. Author affiliation fixed effects are derived from *Scopus*. Information on institutional affiliation is derived from the *Scopus* database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the *Scopus* database with an economics publication who shares the same first and last name as the author in the NBER WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. In Column 10 we additionally control for journal of publication fixed effects. Robust standard errors are reported throughout.

Table 3: Paths to Publication for Race-related NBER Working Papers
Alternative Counterfactuals

OLS estimates, robust standard errors in parentheses

	Published in an Economics Journal		Publication lag (years)		Journal Quality (AER-Weighted)		Published in AER zero weight journal		Published in Top-5		Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Race-related	-.013	-.008	-.199	-.122	.006	.005	-.064***	-.029	.030	.013	-.033	-.051
	(.009)	(.009)	(.151)	(.145)	(.004)	(.004)	(.020)	(.019)	(.024)	(.023)	(.064)	(.059)
Outcome mean (sd)	.986 (.114)	.987 (.110)	2.65 (3.09)	2.69 (2.91)	.054 (.084)	.066 (.104)	.181 (.385)	.189 (.392)	.268 (.443)	.244 (.429)	3.66 (1.47)	3.65 (1.46)
Year FE	X	X	X	X	X	X	X	X	X	X	X	X
JEL Code FE	X		X		X		X		X		X	
Article Characteristics	X	X	X	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X	X	X	X
Journal FE											X	X
LDA Topics		X		X		X		X		X		X
Counterfactual papers:	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls
Sample	All NBER WPs Published in Any Journal	All NBER WPs Published in Any Journal	NBER WPs Published in Econ Journal									
Sample Size	3,935	13,995	3,833	13,822	3,833	13,822	3,833	13,822	3,883	13,822	3,726	13,416

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on NBER working papers first posted between 1974 and 2015. In Columns 1 and 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Columns 3 and 4 the publication lag is the number of years between when the NBER working paper is first posted and its year of publication. In Columns 5 and 6 the outcome is the *AER* -we measure of journal quality constructed in Angrist et al. [2020] (including zeroes). In Columns 7 and 8 the outcome is whether the AER-weighted measure of journal quality constructed in Angrist et al. [2020] is zero. In Columns 9 and 10 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Columns 11 and 12 the outcome is the total number of citations received by an article since publication, as measured from the *Web of Science* or *Scopus*. In Columns 1, 3, 5, 7, 9 and 11, we restrict the sample of not race-related WPs to those which have at least one of the topic keywords in their title and/or abstract. In Columns 2, 4, 6, 8, 10 and 12 we use machine learning to classify the topic of WPs and then control for these broad topics instead of controlling for JEL codes. All specifications include fixed effects for the year in which the working paper is first posted, and its JEL/topic model codes. Working paper characteristics include a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes. Author affiliation fixed effects are derived from *Scopus*. Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the NBER WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. Robust standard errors are reported throughout.

Table 4: Readability Scores, NBER Working Papers

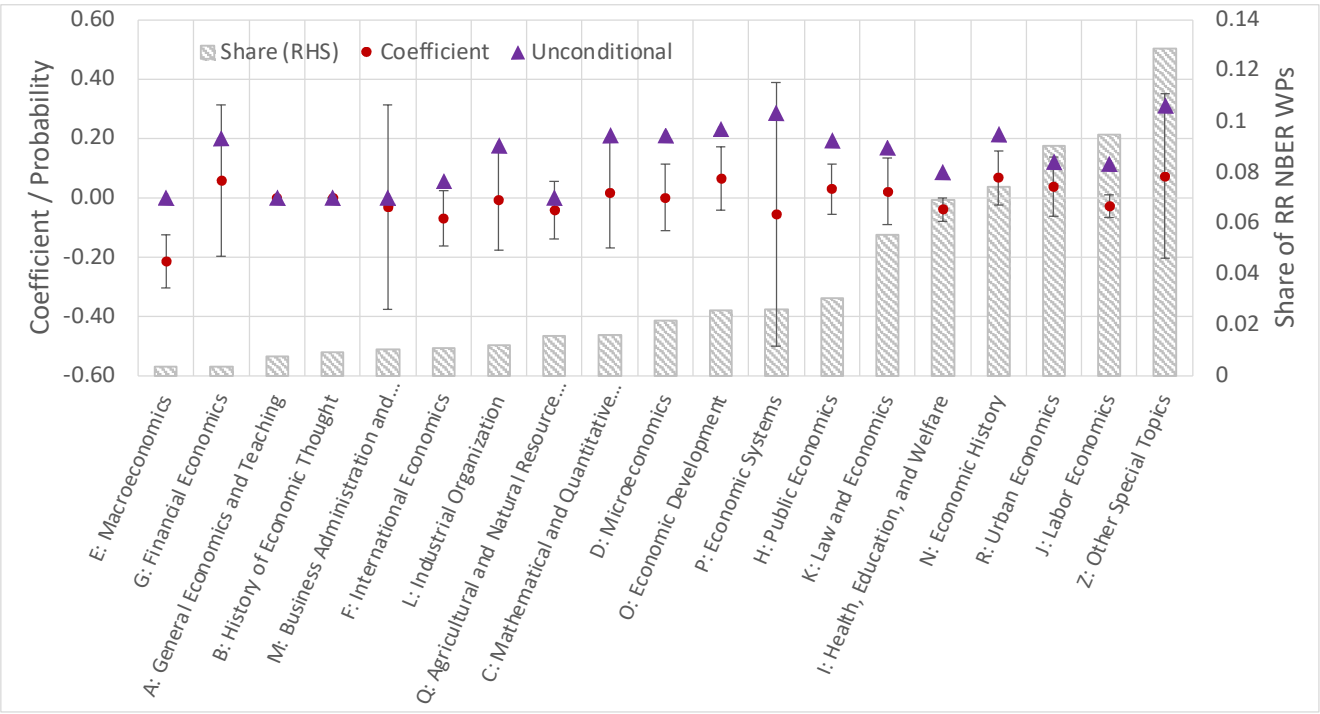
OLS estimates, robust standard errors in parentheses

	Working Paper Version			Published Version			Change Across Versions		
	Flesch Reading Ease Score	Gunning Fog Score	Simple Measure of Gobbledygook Score	Flesch Reading Ease Score	Gunning Fog Score	Simple Measure of Gobbledygook Score	Flesch Reading Ease Score	Gunning Fog Score	Simple Measure of Gobbledygook Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race-related	.214***	.150***	.145***	.148***	.093	.094	-.066	-.062	-.058
	(.059)	(.060)	(.056)	(.064)	(.058)	(.059)	(.052)	(.053)	(.053)
Outcome mean (sd)	.004	.021	.007	.011	.020	.016	.008	.001	.011
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
Article Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Sample	All NBER WPs Published in Any Journal with available SCOPUS/JSTOR abstracts (N=9,287)								

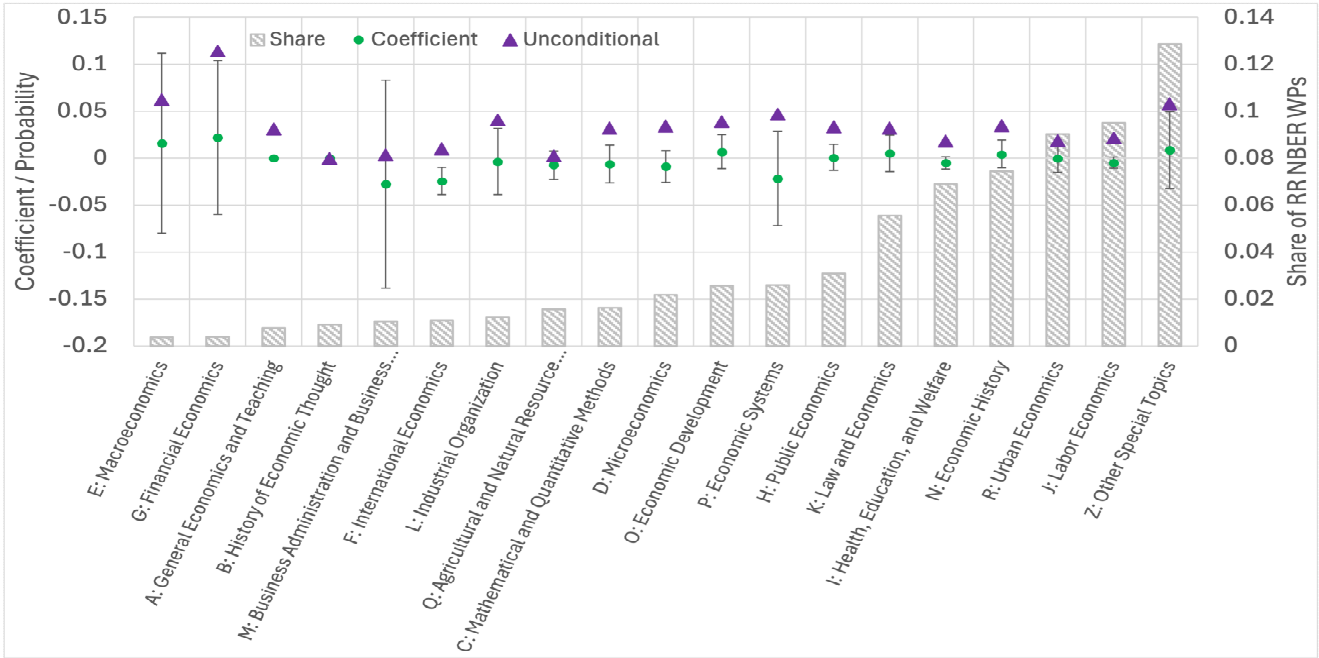
Notes:*** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on NBER working papers first posted between 1974 and 2015 that are matched to publication abstracts from *JSTOR* or *Scopus* . We consider three readability indices. The Flesch Reading Ease test assigns higher scores to materials that are easier to read and lower scores to passages that are more challenging to comprehend. The Gunning Fog Index also estimates the reading level required to understand a piece of writing. It measures the complexity of sentences and words in the text, with higher scores indicating more complex and challenging content. The Simple Measure of Gobbledygook (SMOG) measures the complexity of written content by analyzing the number of words with three or more syllables in a sample text. The higher the SMOG score, the more advanced the reading level required to understand the text. As the scores have different scales and signs, to compare scores, the Gunning Fog and SMOG measures are inverted. For all measures a higher score thus corresponds to easier to read and comprehend texts. We standardize each measure to have mean zero and standard deviation one. In Columns 1 to 3, the outcomes are the readability scores of the NBER WP. In Columns 4 to 6, the outcomes are the readability scores of the published papers. In Columns 7 to 9 the outcomes are the change in readability scores between the published papers and their NBER WP version. Working paper characteristics include a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes. Author affiliation fixed effects are derived from Scopus. Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the NBER WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. Robust standard errors are reported throughout.

Figure 1: Paths to Publication by Field

A. Publications in a Top-5 Journal



B. *AER*-Weighted Quality of Journal Publication



Notes: The sample is based on NBER working papers first released between 1974 and 2015. Panel A plots for each JEL code: (i) the unconditional probability of a race-related NBER WP being published in the top-5; (ii) the conditional estimate, that include fixed effects for the year in which the working paper is first posted, working paper characteristics (a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes). Author affiliation fixed effects are derived from Scopus. We also control for the average number of matches found for each author of an article (and its square). We overlay this with a histogram showing the share of WPs in the JEL-code that are race-related, where we order the fields in increasing share of race-related WPs. Panel B repeats the analysis where the outcomes is the *AER*-weight of the journal publication. In both Panels results for JEL codes A and B are omitted due to multicollinearity issues. 95% confidence intervals are shown throughout.

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

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.

Table A4: Race-Related NBER Working Papers

Means, standard deviation in parentheses, p-values in brackets

	(1) Race-related	(2) Not Race-related	Test of Equality [p-value]
A. Working Paper Characteristics			
Number of authors	2.12	2.11	[.733]
	(.931)	(.928)	
Number of pages	47.5	44.7	[.000]
	(16.3)	(16.7)	
Title length (letter count)	69.3	64.7	[.000]
	(27.3)	(26.2)	
Number of JEL codes	1.58	1.42	[.000]
	(1.03)	(1.07)	
B. Groups and Topics Studied			
Group: Black	.538	.007	[.000]
Group: All Other Groups	.145	.002	[.000]
Group: Non-Specified	.717	.009	[.000]
Topic: Discrimination	.176	.019	[.000]
Topic: Inequality	.765	.238	[.000]
Topic: Diversity	.215	.009	[.000]
Topic: Identity	.027	.003	[.000]
Topic: Historic	.058	.003	[.000]

Notes: The sample is based on NBER working papers first posted between 1974 and 2015. Columns 1 and 2 show means and standard deviations in parentheses for working papers classified as race-related and not race-related respectively. Column 3 shows the p-values from a t-test of equality of means between race-related and not race-related articles, based on a regression with robust standard errors. In Panel B all other groups refers to Latinx, Asian, Native American and Other groups.

Table A5: Paths to Publication and Readability for NBER Working Papers

OLS estimates, robust standard errors in parentheses

	Published in any Journal	Published in an Economics Journal	Publication lag (years)	Journal Quality (AER-Weighted)	Published in AER zero weight journal	Published in Top-5	Publication Lag (years) | Published in Top-5	Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Flesch Reading Ease Score x Race-related	.035*	.005	-.221*	-.004	.015	-.015	-.213	-.027	.018
	(.020)	(.011)	(.116)	(.004)	(.016)	(.023)	(.198)	(.067)	(.058)
Flesch Reading Ease Score	.001	.002*	.004	.002***	.001	.003	-.053	-.012	-.011
	(.003)	(.000)	(.029)	(.001)	(.003)	(.003)	(.052)	(.011)	(.010)
Race-related	-.005	-.012	-.105	.004	-.052**	.041*	.121	.105	.028
	-0.019	(.008)	(.145)	(.003)	(.018)	(.022)	(.208)	(.064)	(.057)
Outcome mean (sd)	.734	.988	2.65 (3.10)	.066 (.104)	.181	.244	2.28 (2.24)	3.64 (1.46)	3.65 (1.46)
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
WP Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Journal FE									X
Sample	All Published NBER WPs (N=19,020)	All NBER WPs Published in Any Journal (N=13,945)	NBER WPs Published in Econ Journal (N=13,773)	NBER WPs Published in Econ Journal (N=13,773)			NBER WPs Published in Top 5 (N=3,377)	NBER WPs Published in Econ Journal (N=13,511)	NBER WPs Published in Econ Journal (N=13,416)

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on NBER working papers first posted between 1974 and 2015. In Column 1 the outcome is a dummy for whether the working paper is published in any journal. In Column 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Column 3 the publication lag is the number of years between when the NBER working paper is first posted and its year of publication. In Column 4 the outcome is whether the AER-weighted measure of journal quality constructed in Angrist et al. [2020] is zero. In Column 5 the outcome is the AER-weighted measure of journal quality constructed in Angrist et al. [2020] (including zeroes). In Column 6 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Column 7 the outcome is the publication lag is the number of years between when the NBER working paper is first posted and its year of publication in a top-5 economics journal. In Columns 8 and 9 the outcome is the total number of citations received by an article since publication, as measured from the Web of Science or Scopus. All specifications include fixed effects for the year in which the working paper is first posted, and its JEL codes. Working paper characteristics include a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes. Author affiliation fixed effects are derived from Scopus. Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the NBER WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. In Column 10 we additionally control for journal of publication fixed effects. Robust standard errors are reported throughout.

Table A6: Paths to Publication for Race-related CEPR Working Papers

OLS estimates, robust standard errors in parentheses

	Published in any Journal	Published in an Economics Journal	Publication lag (years)	Journal Quality (AER-Weighted)	Published in AER Zero weight journal	Published in Top-5	Publication Lag (years) | Published in Top-5	Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race-related	.002	-.039*	.134	-.011**	.009	-.036	-.269	-.008	.154
	(.038)	(.020)	(.218)	(.005)	(.047)	(.030)	(.431)	(.132)	(.115)
Outcome mean (sd)	.627	.993	2.75 (2.39)	.039 (.082)	.276 (.448)	.131	2.56 (1.79)	3.20 (1.43)	3.21 (1.43)
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
Article Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Journal FE									X
Sample	All published CEPR WPs	All CEPR WPs Published in Any Journal	All CEPR WPs Published in Econ Journal	All CEPR WPs Published in Econ Journal			CEPR WPs Published in Top 5		
Sample Size	10,303	6,459	6,419	6,419			831	6,233	6,158

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on CEPR working papers first posted between 1984 and 2015. In Column 1 the outcome is a dummy for whether the working paper is published in any journal. In Column 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Column 3 the publication lag is the number of years between when the CEPR working paper is first posted and its year of publication. In Column 4 the outcome is whether the *AER*-weighted measure of journal quality constructed in Angrist *et al.* [2020] is zero. In Column 5 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Column 6 the outcome is the AER-weighted measure of journal quality constructed in Angrist et al. [2020] (including zeroes). In Column 7 the outcome is the publication lag is the number of years between when the CEPR working paper is first posted and its year of publication in a top-5 economics journal. In Columns 8 and 9 the outcome is the total number of citations received by an article since publication, as measured from the *Web of Science* or *Scopus*. All specifications include fixed effects for the year in which the working paper is first posted, and its JEL codes. Working paper characteristics include a linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes (unlike for NBER WPs, page counts are unavailable for CEPR WPs). Author affiliation fixed effects are derived from Scopus. Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the CEPR data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the CEPR WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of CEPR working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. In Column 10 we additionally control for journal of publication fixed effects. Robust standard errors are reported throughout.

Table A7: Paths to Publication for Race-related CEPR Working Papers
Alternative Counterfactuals

OLS estimates, robust standard errors in parentheses

	Published in an Economics Journal		Publication lag (years)		Journal Quality (AER-Weighted)		Published in AER Zero weight journal		Published in Top-5		Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Race-related	-.042*	-.037*	-.076	.199	-.001	-.010*	-.049	-.004	-.026	-.038	.153	.123
	(.023)	(.021)	(.233)	(.219)	(.006)	(.006)	(.052)	(.046)	(.039)	(.032)	(.134)	(.109)
Outcome mean (sd)	.993 (.081)	.994 (.078)	2.84 (2.31)	2.76 (2.39)	.032 (.062)	.039 (.082)	.281 (.449)	.276 (.448)	.159 (.366)	.131 (.338)	3.21 (1.45)	3.21 (1.43)
Year FE	X	X	X	X	X	X	X	X	X	X	X	X
JEL Code FE	X		X		X		X		X		X	
Article Characteristics	X	X	X	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X	X	X	X
Journal FE											X	X
LDA Topics		X		X		X		X		X		X
Counterfactual papers:	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls	all not race-related, selected topic keywords	all not race-related, LDA topics as controls
Sample	CEPR WPs Published in Any Journal	CEPR WPs Published in Any Journal	CPER WPs Published in Econ Journal									
Sample Size	1,797	6,459	1,785	6,419	1,785	6,419	1,785	6,419	1,785	6,419	1,666	6,158

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on CEPR working papers first posted between 1974 and 2015. In Columns 1 and 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Columns 3 and 4 the publication lag is the number of years between when the CEPR working paper is first posted and its year of publication. In Columns 5 and 6 the outcome is the *AER* -we measure of journal quality constructed in Angrist *et al.* [2020] (including zeroes). In Columns 7 and 8 the outcome is whether the AER-weighted measure of journal quality constructed in Angrist et al. [2020] is zero. In Columns 9 and 10 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Columns 11 and 12 the outcome is the total number of citations received by an article since publication, as measured from the Web of Science or Scopus. In Columns 1, 3, 5, 7, 9 and 11, we restrict the sample of not race-related WPs to those which have at least one of the topic keywords in their title and/or abstract. In Columns 2, 4, 6, 8, 10 and 12 we use machine learning to classify the topic of WPs and then control for these broad topics instead of controlling for JEL codes. All specifications include fixed effects for the year in which the working paper is first posted, and its JEL/topic model codes. Working paper characteristics include a linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes (unlike for NBER WPs, page counts are unavailable for CEPR WPs). Author affiliation fixed effects are derived from *Scopus* . Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the CEPR WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of CEPR working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. Robust standard errors are reported throughout.

Table A8: Readability Scores, CEPR Working Papers

OLS estimates, robust standard errors in parentheses

	Working Paper Version			Published Version			Change Across Versions		
	Flesch Reading Ease Score	Gunning Fog Score	Simple Measure of Gobbledygook Score	Flesch Reading Ease Score	Gunning Fog Score	Simple Measure of Gobbledygook Score	Flesch Reading Ease Score	Gunning Fog Score	Simple Measure of Gobbledygook Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race-related	.001	.069	.056	-.028	.040	.052	-.045	-.041	-.016
	(.100)	(.082)	(.089)	(.104)	(.093)	(.094)	(.073)	(.079)	(.079)
Outcome mean (sd)	.018	.026	.033	.000	.000	.000	-.018	-.025	-.022
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
Article Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Sample	All CEPR WPs Published in Any Journal with available SCOPUS/JSTOR abstracts (N=6,097)								

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on CEPR working papers first posted between 1984 and 2015 that are matched to publication abstracts from *JSTOR* or *Scopus*. We consider three readability indices. The Flesch Reading Ease test assigns higher scores to materials that are easier to read and lower scores to passages that are more challenging to comprehend. The Gunning Fog Index also estimates the reading level required to understand a piece of writing. It measures the complexity of sentences and words in the text, with higher scores indicating more complex and challenging content. The Simple Measure of Gobbledygook (SMOG) measures the complexity of written content by analyzing the number of words with three or more syllables in a sample text. The higher the SMOG score, the more advanced the reading level required to understand the text. As the scores have different scales and signs, to compare scores, the Gunning Fog and SMOG measures are inverted. For all measures a higher score thus corresponds to easier to read and comprehend texts. We standardize each measure to have mean zero and standard deviation one. In Columns 1 to 3, the outcomes are the readability scores of the CEPR WP. In Columns 4 to 6, the outcomes are the readability scores of the published papers. In Columns 7 to 9 the outcomes are the change in readability scores between the published papers and their CEPR WP version. Working paper characteristics include a linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes (unlike for NBER WPs, page counts are unavailable for CEPR WPs). Author affiliation fixed effects are derived from Scopus. Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the CEPR WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of CEPR working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. Robust standard errors are reported throughout.

Table A9: Paths to Publication for NBER Working Papers, Group Studied

OLS estimates, robust standard errors in parentheses

	Published in any Journal	Published in an Economics Journal	Publication lag (years)	Journal Quality (AER-Weighted)	Published in AER Zero weight journal	Published in Top-5	Publication Lag (years) | Published in Top-5	Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race-related x Black	.134**	-.008	.004	-.008	-.026	.029	.028	.135	.203
	(.058)	(.013)	(.362)	(.014)	(.053)	(.064)	(.642)	(.181)	(.163)
Race-related x All other groups	-.132	.043	.676	-.014	.122	-.186*	.474	-.257	-.119
	(.091)	(.041)	(.47)	(.018)	(.092)	(.096)	(.703)	(.279)	(.273)
Race-related x Non-specific	.007	.034	-.712	.003	.026	-.006	-.593	-.166	-.061
	(.057)	(.022)	(.473)	(.012)	(.053)	(.066)	(.565)	(.187)	(.175)
Outcome mean (sd)	.734	.988	2.65 (3.10)	.066 (.104)	.181 (.385)	.244	2.28 (2.24)	3.64 (1.46)	3.65 (1.45)
p-values, within race-related research									
Black = All other groups	[.016]	[.251]	[.269]	[.812]	[.166]	[.071]	[.671]	[.270]	[.345]
Black = Non-specific	[.072]	[.098]	[.220]	[.507]	[.432]	[.647]	[.366]	[.152]	[.169]
All other groups = Non-specific	[.188]	[.844]	[.019]	[.448]	[.346]	[.121]	[.246]	[.792]	[.861]
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
WP Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Journal FE									X
Sample	All Published NBER WPs (N=19,070)	All NBER WPs Published in Any Journal (N=13,995)	NBER WPs Published in Econ Journal (N=13,822)	NBER WPs Published in Econ Journal (N=13,822)			NBER WPs Published in Top 5 (N=3,377)	NBER WPs Published in Econ Journal (N=13,511)	NBER WPs Published in Econ Journal (N=13,416)

Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on NBER working papers first posted between 1974 and 2015. In Column 1 the outcome is a dummy for whether the working paper is published in any journal. In Column 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Column 3 the publication lag is the number of years between when the NBER working paper is first posted and its year of publication. In Column 4 the outcome is whether the *AER* -weighted measure of journal quality constructed in Angrist et al. [2020] is zero. In Column 5 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Column 6 the outcome is the AER-weighted measure of journal quality constructed in Angrist et al. [2020] (including zeroes). In Column 7 the outcome is the publication lag is the number of years between when the NBER working paper is first posted and its year of publication in a top-5 economics journal. In Columns 8 and 9 the outcome is the total number of citations received by an article since publication, as measured from the *Web of Science* or *Scopus* . All specifications include fixed effects for the year in which the working paper is first posted, and its JEL codes. Working paper characteristics include a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes. Author affiliation fixed effects are derived from *Scopus* . Information on institutional affiliation is derived from the *Scopus* database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the *Scopus* database with an economics publication who shares the same first and last name as the author in the NBER WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. In Column 10 we additionally control for journal of publication fixed effects. At the foot of each Column we report the p-value on the null that the interactions of race-related articles with the group under study (black, all other groups, non-specific groups) are equal. Robust standard errors are reported throughout.

Table A10: Paths to Publication for NBER Working Papers, Topic Studied

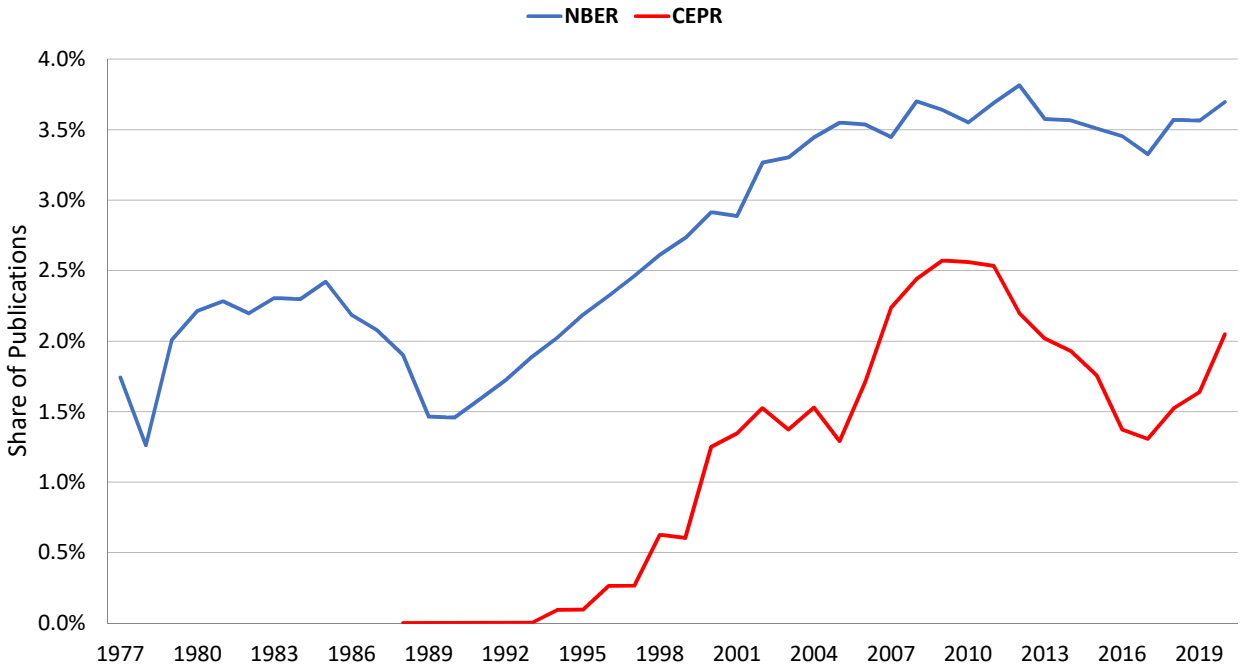
OLS estimates, robust standard errors in parentheses

	Published in any Journal	Published in an Economics Journal	Publication lag (years)	Journal Quality (AER-Weighted)	Published in AER Zero weight journal	Published in Top-5	Publication Lag (years) | Published in Top-5	Log (citations)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Race-related x Discrimination	.114**	-.038	.225	.016	-.038	.085	.688	.252	.197
	(.053)	(.028)	(.448)	(.011)	(.052)	(.069)	(.508)	(.174)	(.161)
Race-related x Inequality	-.007	-.012	-.748*	.005	.019	-.013	.182	.046	.028
	(.047)	(.018)	(.402)	(.009)	(.044)	(.057)	(.458)	(.141)	(.126)
Race-related x Diversity	.065	.020*	.270	.026**	-.058	.098	-.466	-.149	-.243
	(.059)	(.012)	(.421)	(.011)	(.057)	(.068)	(.429)	(.181)	(.162)
Race-related x Identity	.158*	.084**	-1.61**	-.011	.050	-.000	-1.07	-.163	-.046
	(.093)	(.036)	(.691)	(.023)	(.111)	(.138)	(.826)	(.347)	(.325)
Race-related x Historic	.016	-.063	.047	.011	-.126	-.004	-1.03	-.106	-.399
	(.087)	(.075)	(1.14)	(.020)	(.095)	(.106)	(1.37)	(.333)	(.303)
Outcome mean (sd)	.734	.988	2.65 (3.10)	.066 (.104)	.181 (.385)	.244	2.28 (2.24)	3.64 (1.46)	3.65 (1.45)
p-values, within race-related research									
Discrimination=Inequality	[.046]	[.336]	[.119]	[.382]	[.248]	[.196]	[.489]	[.287]	[.338]
Discrimination=Diversity	[.516]	[.070]	[.930]	[.536]	[.783]	[.898]	[.086]	[.113]	[.056]
Discrimination=Identity	[.679]	[.011]	[.040]	[.291]	[.462]	[.587]	[.111]	[.317]	[.526]
Discrimination=Historical	[.336]	[.759]	[.864]	[.809]	[.412]	[.488]	[.252]	[.349]	[.078]
Inequality=Diversity	[.288]	[.103]	[.059]	[.093]	[.295]	[.153]	[.249]	[.372]	[.161]
Inequality=Identity	[.070]	[.013]	[.193]	[.492]	[.782]	[.927]	[.137]	[.576]	[.832]
Inequality=Historic	[.805]	[.546]	[.533]	[.787]	[.172]	[.940]	[.375]	[.668]	[.174]
Diversity=Identity	[.362]	[.069]	[.020]	[.144]	[.392]	[.517]	[.472]	[.971]	[.588]
Diversity=Historic	[.642]	[.301]	[.823]	[.507]	[.531]	[.420]	[.677]	[.910]	[.650]
Identity=Historic	[.259]	[.099]	[.250]	[.475]	[.238]	[.981]	[.982]	[.905]	[.434]
Year FE	X	X	X	X	X	X	X	X	X
JEL Code FE	X	X	X	X	X	X	X	X	X
WP Characteristics	X	X	X	X	X	X	X	X	X
Author Affiliation FE	X	X	X	X	X	X	X	X	X
Journal FE									X
Sample	All Published NBER WPs (N=19,070)	All NBER WPs Published in Any Journal (N=13,995)	NBER WPs Published in Econ Journal (N=13,822)	NBER WPs Published in Econ Journal (N=13,822)			NBER WPs Published in Top 5 (N=3,377)	NBER WPs Published in Econ Journal (N=13,511)	NBER WPs Published in Econ Journal (N=13,416)

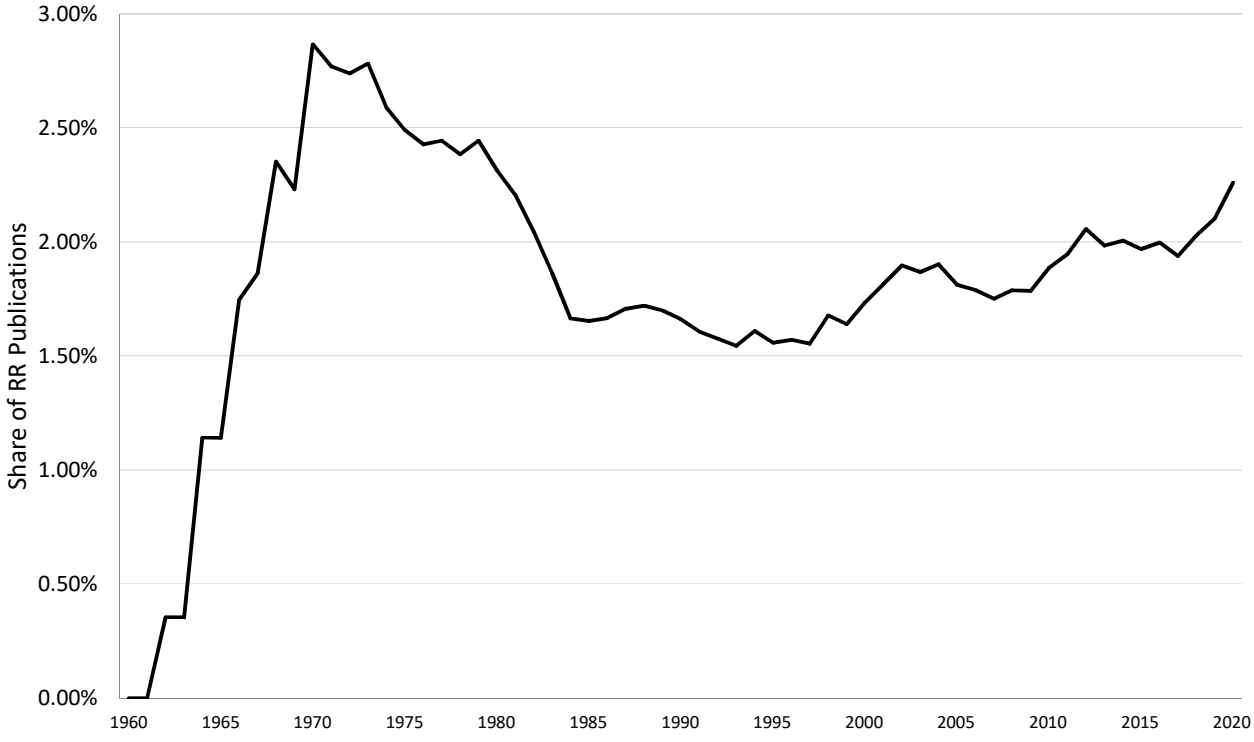
Notes: *** denotes significance at the 1%, ** at the 5%, * at the 10% level. The sample is based on NBER working papers first posted between 1974 and 2015. In Column 1 the outcome is a dummy for whether the working paper is published in any journal. In Column 2 the outcome is a dummy for whether the working paper is published in an economics journal. In Column 3 the publication lag is the number of years between when the NBER working paper is first posted and its year of publication. In Column 4 the outcome is whether the *AER*-weighted measure of journal quality constructed in Angrist *et al.* [2020] is zero. In Column 5 the outcome is a dummy for whether the working paper is published in a top-5 economics journal. In Column 6 the outcome is the AER-weighted measure of journal quality constructed in Angrist et al. [2020] (including zeroes). In Column 7 the outcome is the publication lag is the number of years between when the NBER working paper is first posted and its year of publication in a top-5 economics journal. In Columns 8 and 9 the outcome is the total number of citations received by an article since publication, as measured from the Web of Science or Scopus. All specifications include fixed effects for the year in which the working paper is first posted, and its JEL codes. Working paper characteristics include a linear and quadratic in page counts, linear and quadratic terms for the title length, dummies for the number of authors and for the number of unique JEL codes. Author affiliation fixed effects are derived from Scopus. Information on institutional affiliation is derived from the Scopus database, using first and last names. For each author-year combination we observe in the NBER data, we retrieve the affiliation of the author in the Scopus database with an economics publication who shares the same first and last name as the author in the NBER WP dataset. Moreover, the selected author should have a publication that is closest in time to the author being analyzed. When we identify multiple matches, we break ties randomly. We also control for the average number of matches found for each author of an article (and its quadratic). Affiliations of NBER working paper authors are found in two thirds of cases. The author affiliation dummies to cover the 100 most frequent institutions in our data set, a dummy for other affiliations and a dummy for no matched affiliation. In Column 10 we additionally control for journal of publication fixed effects. At the foot of each Column we report the p-value on the null that the interactions of race related articles with their topic under study (discrimination, inequality, diversity, identity historic) are equal across pairs of topics. Robust standard errors are reported throughout.

Figure A1: Time Trends in Race Related Research [Advani et al. 2025a]

A. Share of Race-related Working Papers



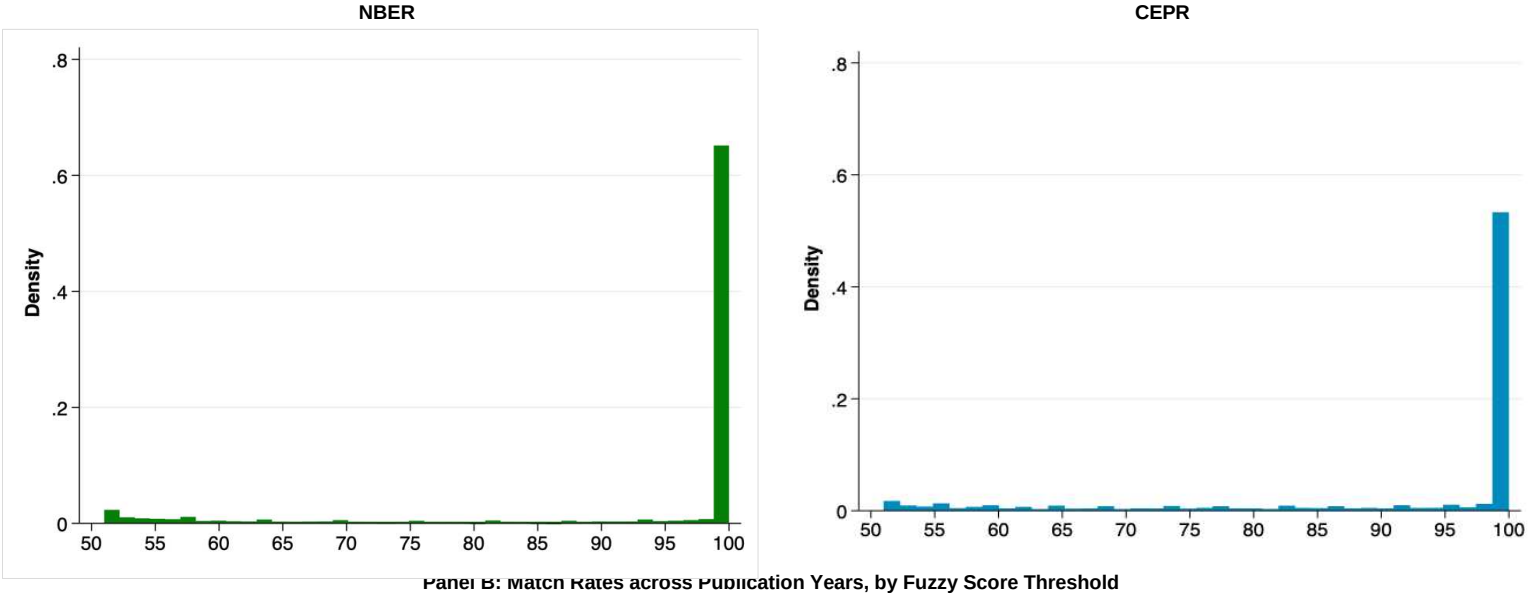
B. Share of Race-related Publications in Economics Journals



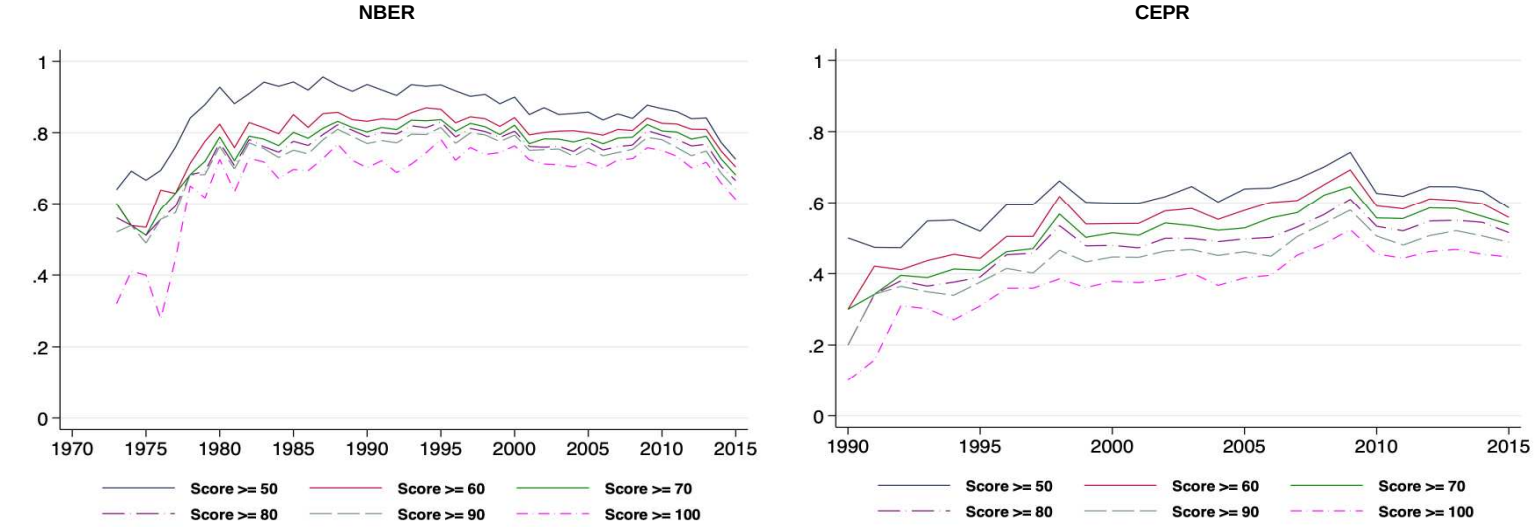
Notes: In Panel A, the sample is based on NBER working papers first released between 1974 and 2019, and CEPR working papers released between 1984 and 2019. We show the shares of working papers identified to be race-related in each series, in five-year moving averages - with the NBER (CEPR) series starting in 1977 (1987). In Panel B, we use a corpus of publications in economics journals. We use JSTOR as our baseline 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. To fill gaps in the JSTOR publication series, we utilize data from the Web of Science (WOS) and Scopus publications series. We report five-year moving averages throughout. We report the share of total publications identified to be race-related by year of publication.

Figure A2: Matching Working Papers to Publications

Panel A: Distribution of String Similarity Measure Among Matched Working Papers

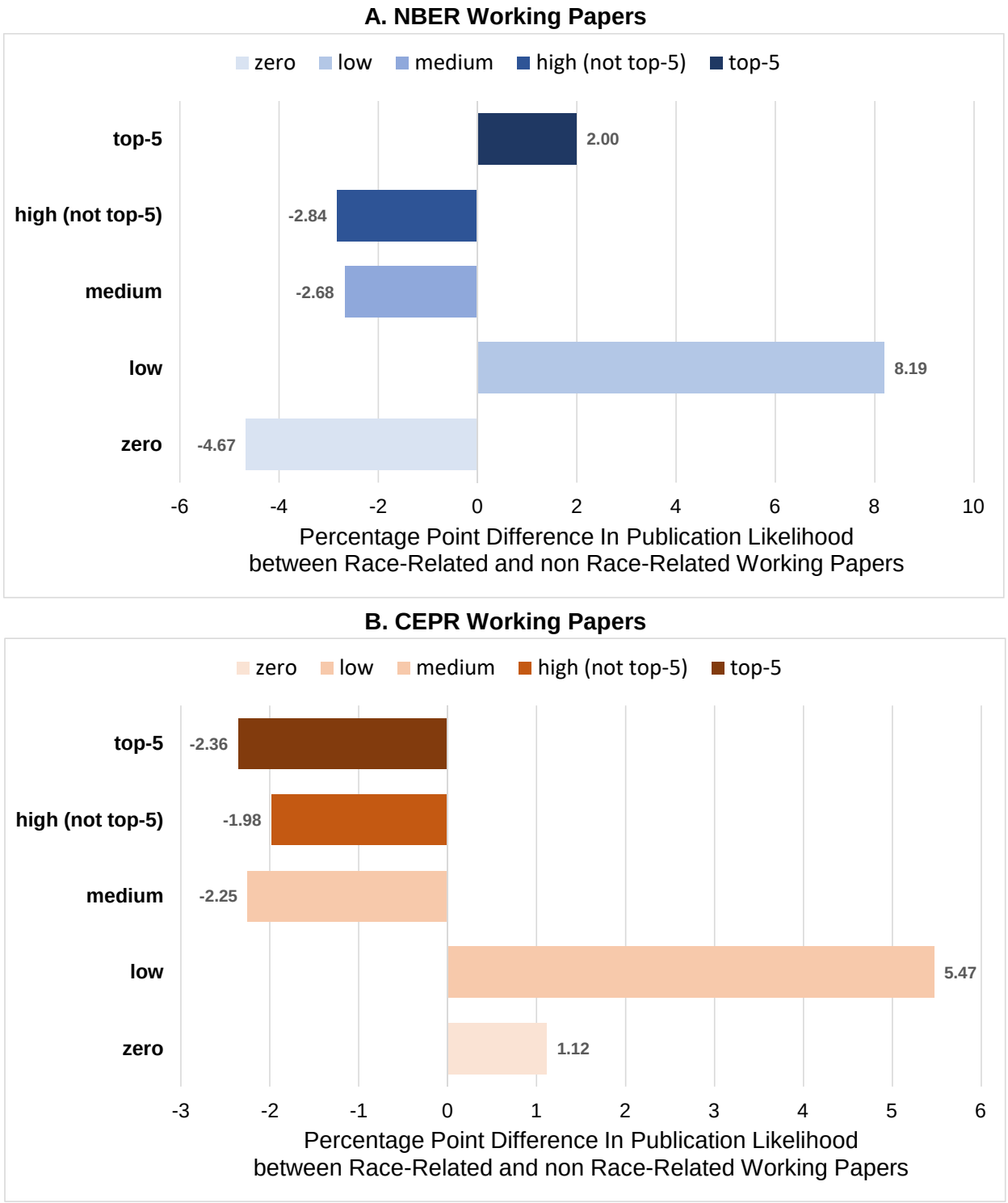


Panel B: Match Rates across Publication Years, by Fuzzy Score Threshold



Notes: Panels A and B report summary statistics from our matching process from NBER working papers to the *Web of Science* and *Scopus* databases. Panel A reports the distribution of the String Similarity measure between NBER working paper titles and matched titles from *Web of Science* or *Scopus*, conditional on finding a match. The similarity measure used is based on Levenshtein distance. A similarity of a 100 indicates a perfect match. The sample is based on all NBER (CEPR) working papers posted from 1974 to 2015 (1983 to 2015) and matched with a journal publication. Panel B reports the publication probability measured for NBER (CEPR) working papers first posted each year, and how it varies with our inclusion criteria. The six series shown correspond to gradually increasing the required threshold for matching on the String Similarity measure.

Figure A3: *AER*-Weighted Publication Outcomes of Working Papers



Notes: The sample in Panel A is based on NBER working papers first released between 1974 and 2019, and in Panel B it is based on CEPR working papers released between 1984 and 2019. We then consider the set of working papers that are published in an economics journal. The outcome is the *AER*-weighted measure of journal quality constructed in Angrist et al. [2020], where we consider whether the working paper is published in a top-5 economics journal, in the top tercile of *AER*-weights (excluding the top-5 journals), in the middle tercile of *AER*-weights, in the bottom tercile of *AER*-weights, or in a journal with zero *AER*-weight. The figure shows the unconditional differences in outcomes between race-related and not race-related working papers in the NBER and CEPR series in each of these outcomes.

Figure A4: LDA Topics in NBER Working Papers

Notes: We use Latent Dirichlet Allocation (LDA) modeling to identify topics in the corpus of NBER working papers posted from 1974 to 2019. Our benchmark model identifies 30 topics. The Figure displays word clouds for the topics generated.