Can Truth-Seeking Triumph Over Tribalism? Evidence from Ideological Reversals Amongst Economists[1]

Matt Knepper[2]
University of Georgia, Economics Department

Brian Wheaton[3]
UCLA, Anderson School of Management

February 7, 2025

Abstract: Research in economics often carries direct political implications, with findings supporting either right-wing or left-wing perspectives. But what happens when a researcher known for publishing right-wing findings publishes a paper with left-wing findings (or vice versa)? We refer to these instances as ideological reversals. This study explores whether such researchers face penalties – such as losing their existing audience without attracting a new one – or if they are rewarded with a broader audience and increased citations. The answers to these questions are crucial for understanding whether academia promotes the advancement of knowledge or the reinforcement of echo chambers. In order to identify ideological reversals, we begin by categorizing papers included in meta-analyses of key literatures in economics as "right" or "left" based on their findings relative to other papers in their literature (e.g., the presence or absence of disemployment effects in the minimum wage literature). We then scrape the abstracts (and other metadata) of every economics paper ever published, and we deploy machine learning in order to categorize the ideological implications of these papers. We find that reversals are associated with gaining a broader audience and more citations. This result is robust to a variety of checks, including restricting analysis to the citation trajectory of papers already published before an author's reversal. Most optimistically, authors who have left-to-right (right-to-left) reversals not only attract a new right-wing (left-wing) audience for their recent work, this new audience also engages with and cites the author's previous left-wing (right-wing) papers, thereby helping to break down echo chambers.

1 Introduction

Does science inevitably push toward greater understanding of the world, or can science go wrong because of perverse incentives within the scientific community? This crucial question was

[1] We are grateful to Paola Giuliano, Clemence Tricaud, Melanie Wasserman, and Romain Wacziarg for helpful advice and suggestions. We also thank the participants of the Northwestern Kellogg Political Economy Seminar and the UCLA Anderson GEM faculty lunch.
[2] Contact: mknepper@uga.edu
[3] Contact: brian.wheaton@anderson.ucla.edu

a matter of intense debate amongst philosophers of science throughout the 20th-century. Karl Popper embodied the most optimistic view of science. From his perspective, science is a continuous, incremental march toward truth which progresses by eliminating false theories as they are disproven and replacing them with better ones. Thomas Kuhn took a more skeptical view than that of Popper, arguing that science operates within paradigms, clinging to old theories for lengthy periods of time and rejecting or even penalizing challenges to those theories. The paradigm is maintained until the evidence against it becomes truly overwhelming, at which point a new paradigm – not always better than the old paradigm in an objective sense – emerges to replace it. Michael Polanyi's view, in many ways complementing that of Kuhn, stressed that the personal experiences and beliefs of the scientists themselves – and the structures/traditions of the scientific community – exert a substantial force on how science does (and does not) proceed.

In this paper, we examine a context relevant to this debate: economics academia. Economics, as a discipline, studies a multitude of topics that have direct political implications, with findings potentially supporting either right-wing or left-wing viewpoints. For example, a study of minimum wages may find substantial disemployment effects, or it may find negligible ones. A study of the gender wage gap may find a minimal gap or a substantial one. And a study of government spending may find a large crowd-out effect, or it may find a small one. While authors often tend to stick to publishing findings that support one side of the spectrum across their entire body of work, it is not rare for an author to eventually venture outside his/her ideological realm and publish a result on the opposite side of the political spectrum from their previous work. We term these events ideological reversals, and we study their consequences. In particular, we investigate whether such authors are penalized – abandoned by the audience who previously cited them without gaining a new audience – or rewarded – gaining an expanded audience and more citations. This has direct implications for the formation and persistence of echo chambers in academia, which could potentially serve to hobble the progress of science toward uncovering truth.

To undertake this study, we scraped Web of Science for the abstracts, author/topic/journal information, and full citation network of every economics article ever published. We obtained meta-analyses for a multitude of literatures within economics that have political implications (e.g.,

minimum wage, economics of unions, taxable income elasticity, intergenerational transfers, and dozens more) and categorized the papers in these meta-analyses into those supporting more right-wing conclusions or those supporting more left-wing conclusions. With these papers as an input, we then develop and apply a machine learning algorithm to classify the ideology of the scraped papers. This allows us to construct a balanced panel at the (author, paper, year) level with information on citations and ideology. We can further split the citations variable into citations by left-wing papers and citations by right-wing papers for auxiliary analysis. With this, we are able to begin our data exploration.

We first perform some descriptive analysis of the data, which establishes several key facts:

- Fact #1: The typical author mostly publishes results on one side of the political spectrum.
- Fact #2: Ideological reversals are not rare; they occur at least once for 40% of authors.
- Fact #3: Ideological reversals become much more common later in an author's career, with authors essentially never undergoing a reversal in the first decade of their career.
- Fact #4: Most ideological reversals do not represent a permanent defection to the other side of the political spectrum, but rather the beginning of repeatedly publishing results on both sides of the spectrum.
- Fact #5: Ideological reversals occur much more frequently amongst authors who are (initially) classified as right-wing.

We then proceed to inferential analysis, investigating the effects of ideological reversals. We restrict our sample to papers published in Top 100 journals, as the median number of citations received by papers outside this set is 0. Using an event-study difference-in-differences approach, we find no evidence of pre- trends and strong evidence of an *increase* in citations following an ideological reversal. In response to concerns that ideological reversals may correlate with a change in the quality of paper an author writes, we restrict attention to papers published before the reversal. These papers, too, begin to accumulate significantly more citations after the author's reversal paper is published. Furthermore, there is no evidence of a change in the quality of an author's papers

around the same time as the author's reversal. The average impact factor of the journals that a reversing author publishes in does not change prior or subsequent to an ideological reversal. In other words, it does not appear that the boost in citations is due to improved papers.

The result continues to be robust to the addition of various fixed effects (career length and paper age) and various sample restrictions geared toward improving the similarity of the treatment and control groups. We also note that the result goes through even if we restrict our sample solely to papers included in the meta-analyses we collected. In other words, our result does not rely on our machine learning procedure. Nor does it rely on using publication dates from Web of Science as opposed to the initial date that the first draft of the paper was written. For a subset of papers, we scrape Google Scholar to determine the date of the first publicized version of the working paper. Using these dates instead does not modify our result.

We do not find any evidence of heterogeneity in our main result by the direction of the ideological reversal. Both left-to-right and right-to-left reversals are rewarded by increased citations of roughly the same magnitude. The boost in citations received subsequent to a left-to-right reversal is mostly driven by citations from right-wing authors, and the boost in citations received subsequent to a right-to-left reversal is mostly driven by citations from left-wing authors. Encouragingly, however, the new right-wing (left-wing) audience garnered by a left-to-right (right-to-left) reversal audience also engages with and cites the author's previous left-wing (right-wing) papers. This dynamic suggests that ideological reversals help prevent the formation of echo chambers in economics academia and expose authors to opposite ideological findings. While this is far from an indisputable proof of Popper's view of science and the scientific community, we regard these findings as optimistic for the triumph of truth-seeking over tribalism.

2 Literature Review

Our work relates to a growing literature of papers studying the incentives in economics academia and how they influence what is published in journals and/or who advances in the profession. Card and DellaVigna (2013) provide a descriptive look at papers published in top economics journals from the 1970s to the 2010s, discussing how journals' practices have shifted

and how this has influenced authorship and publication in economics. Card and DellaVigna (2014) look specifically at the partial and general equilibrium consequences of page limits imposed by journals on economics articles. Card and DellaVigna (2020) study the objective function of editors, providing evidence on what editors maximize as they decide which papers to publish in their journals.

Hengel (2022) studies the writing standards applied to papers written by female economists relative to those applied to papers written by male economists. She finds that female-authored papers are 1 to 6% better-written than male-authored ones, and the gap widens during peer review, which also takes longer for female authors. Hengel then uses a subjective utility framework to argue that this is consistent with higher standards being applied to women. Koffi (2024) studies the recognition of research by women compared to research by men, finding that all-female papers are cited 10% less than all-male papers. The gap disappears, however, when accounting for past publication records. Koffi and Marx (2023) investigate the gender gap in commercialization of science, finding a substantial gap favoring men that does not appear to be explained by the quality of the science or its commercial potential. The gap vanishes in cases when authors self-commercialize and thus is driven by cases when authors must seek investment or cooperation from firms. Koffi, Pongou, and Wantchekon (2024) find evidence of a citation gap favoring white authors over non-white ones. They show that the gap does not appear to be driven by differences in research quality or author visibility but rather homophily in citation patterns – with racial clusters forming in citation networks.

A small handful of recent papers study whether authors' ideology has an impact on the findings of those authors. Jelveh, Kogut, and Naidu (2024) deploy machine learning to predict the political behavior of a subset of economists (for which such data is available), show good out-of-sample fit of their model, and then predict partisanship for all economists. They find evidence of substantial sorting into fields by partisanship, and they also show that there are meaningful differences in papers' results within the same field according to the political partisanship of the writers. Ringgenberg, Shu, and Werner (2023) study a similar question, looking beyond just economics. They develop a measure of political slant – based on the frequency of citations from

liberal vs. conservative think tanks – to examine whether author ideology influences the results of research. According to this measure, research in economics and political science displays a liberal slant, whereas research in finance and accounting displays a conservative slant. They validate their measure of slant with political donations data that enables them to observe authors' donations. Javdani and Chang (2023) also study the role of ideology in economics academia, but with a much different approach. Using an RCT, the authors show that prominent economists are more likely to report agreement with identical statements when they are attributed to mainstream establishment sources as opposed to non-mainstream ones.

We build on this literature by studying a question that has not yet been investigated. Our paper is the first to analyze how it is treated when authors venture away from publishing results that one would predict from their past behavior or their observed ideology. When authors follow the science and it takes them to different ideological conclusions – conclusions that could potentially alienate their existing citation network – how is this received by the profession? Additionally, our machine learning algorithm provides another approach to measuring paper and author ideology, and our descriptive analysis provides further evidence of the relation between authors' ideology and the results they publish.

3 Analytic Framework

3.1 Data

The first main data source that we use as an input to our analysis is Web of Science. Web of Science is an online platform providing detailed citation data for academic papers across various disciplines. In addition to containing the full citation network (papers cited by *i* and papers citing *i* for every paper *i*), Web of Science provides paper abstracts, publication dates, author/topic/journal information, and other metadata. Figure 1 provides an example of a Web of Science page. Because Web of Science contains the full citation network, it is possible to use these data to create a panel dataset on citations accumulated over time. The data on Web of Science cannot readily be downloaded en masse. As such, we systematically scraped Web of Science for

its data on every economics paper ever published – a total of over 2,100,000 papers.[4] This was a monthslong undertaking that required extensive VPN use to evade IP bans and other tools deployed by Web of Science to prevent mass scraping.

Despite scraping the universe of economics papers, for our analysis we focus on papers published in Top 100 journals – a number closer to 200,000. We do this for two reasons. The first is computational: a (paper, author, year)-level panel dataset constructed from 2,100,000 papers contains approximately 150 million observations and is unwieldy to analyze. The second – and more important – reason is that the overwhelming majority of papers published outside Top 100 journals receive zero citations, whereas the overwhelming majority of papers published in Top 100 journals receive citations. Ultimately, for an author to be rewarded or penalized in response to an ideological reversal, it is first necessary that people realize a reversal has occurred. "If a tree falls in a forest, and no-one is around to hear it, does it make a sound?" We abstract from such considerations by discarding this large set of papers outside of Top 100 journals. We also restrict our focus on papers from the 1990s onward, as abstracts are missing for most papers prior to 1990 (presumably because these papers were originally published on paper, not digitally).

We also scrape Google Scholar for the date of the first available working paper version of each economics paper in a Top 20 journal from 1995 to the present. We do this because Web of Science reports the date a paper is published, not the date it is first available as a working paper. Therefore one might worry that this leads to some measurement error regarding when an ideological reversal actually occurs and is observed by others. Google Scholar – which reports the date of each version of a working paper prior to its publication – provides a way around this. The reason for scraping a narrower subset of papers (those in Top 20 rather than Top 100 journals) is because scraping Google Scholar is even more involved than scraping Web of Science. Google deploys more tools – such as reCAPTCHA – to prevent scraping, which can be costly in time and money to penetrate.

[4] We erred on the side of broadness and scraped all papers in the "Business & Economics" category, which includes papers in Finance and other business journals as well.

Our second main data source is a collection of meta-analysis papers covering dozens of different topic areas within economics. We obtained meta-analyses on the minimum wage, the economics of unions, the taxable income elasticity, the fiscal multiplier, intergenerational transfers, trade and productivity, trade and domestic employment, crowd-out, the gender wage gap, unemployment insurance, disability benefits, universal preschool, childcare and employment, immigration and wages, and more. In literatures with multiple meta-analyses, we used the one that covered the most papers. Virtually all meta-analyses contain a table or figure (typically Table 1 or Figure 1) summarizing the findings of each paper in the literature with a coefficient and/or a t-stat. For example, the meta-analysis on the minimum wage reports the elasticity of employment with respect to the minimum wage, the meta-analysis on unions and productivity reports the t-stat of the effect of unionization on productivity, and so on. We use this to categorize the papers in each meta-analysis into those with more left-wing or more right-wing findings.

For each literature, we then asked ChatGPT which conclusion was more consistent with the left-wing perspective and which was more consistent with the right-wing perspective. This produced results in line with our expectations, as the question is typically a straightfoward one. For example, lesser (greater) disemployment effects from a minimum wage increase are consistent with a left-wing (right-wing) perspective in the minimum wage literature. More positive (negative) effects of unionization on productivity are consistent with a left-wing (right-wing) perspective in the unions and productivity literature. Literatures involving trade are where our intuition sometimes deviated from ChatGPT's output. While tariffs and other trade restrictions have more recently become more popular on the right in American/Western European politics, this polarity was reversed prior to the 1990s. Because our analysis focuses on more recent decades and because such a reversal in polarity is slow-moving in any case, this is unlikely to be a major issue, and in any case, our results are robust to dropping trade-related literatures from our analysis.

3.2 Machine Learning Procedure

Taking our ideological classification of papers in the meta-analyses, we then use text feature extraction in Natural Language Processing (NLP) and supervised classification methods,

employing machine learning models to classify the rest of the papers in our dataset from Web of Science. Text Feature Extraction is the process of converting unstructured text data into a numerical format that machine learning models can process. Since models cannot directly work with raw text, feature extraction transforms text into structured numerical vectors while preserving its meaning and context. The text we use to train our models involves the full texts or only the abstracts of each paper (depending on the model).

The CountVectorizer converts a collection of text documents into a bag-of-words (BoW) representation. It creates a vocabulary of all words in the dataset and then represents each document as a vector of word counts, where each dimension is a specific word or term and each value is a word count.

TF-IDF improves upon CountVectorizer by weighing words based on their importance in a document relative to the entire dataset. It reduces the influence of commonly used words and highlights important words in specific documents. Term Frequency (TF) quantifies the importance of a word within a document by measuring how often it appears relative to the total word count. Inverse Document Frequency (IDF) assesses a word's significance across an entire collection of documents by computing the logarithm of the ratio between the total number of documents and the number of documents in which the word appears. In other word, IDF determines the rarity of a word within the corpus. The TF-IDF score, obtained by multiplying TF and IDF, emphasizes words that carry higher informational value and are more relevant to a document's subject matter. Similar to CountVectorizer, in the final vectors, each dimension represents a specific word or term, but each value is a TF-IDF score.

Since we have the vectors of words that could be processed by machine learning models and we want to determine if the paper is left-leaning or right-leaning, we use five supervised classifiers to do such a binary classification. They are Logistic Regression, Multinomial Naïve Bayes based on Bayes' Theorem, Random Forest, AdaBoost, and XGBoost. These classification models have different strengths and weaknesses based on their mathematical foundation, interpretability, and performance. For example, the Multinomial Naïve Bayes model handles high-dimensional data

well, such as texts and works well when features represent word counts, while it is less effective when features are correlated.

Our dataset from the meta-analyses has a roughly equal number of left-wing and right-wing papers (by construction, since papers left of median are classified as left-wing and papers right of median are designated right-wing). Random guessing accuracy would produce correct results around 50% of the time. We randomly divide the entire set as the training and the testing set by 80-to-20. Combining two text feature extraction models and five classification models, we have ten models and use them to train on either the full texts or just the abstracts. Based on GridSearchCV function, we optimize each model's hyper-parameters, such as max_features, min_df, max_df, and stop_words, to get the best 3-folder or 2-folder cross-validation accuracy. Then, we apply these models to test on the same testing set, which includes only the abstracts of each paper. Comparing the best cross-validation score, train accuracy, and test accuracy, we select the best model that has these three metrics relatively high and close to one another to avoid the overfitting problem (see Table 1). Our final model is the Multinomial Naive Bayes with TF-IDF trained full texts, where both the train and test accuracy is about 80%, much higher than that of random guessing.

3.3 Identification

In order to ascertain the effects of ideological reversals on subsequent citations, we estimate an event-study differences-in-differences specification:

$$Y_{a,p,t} = \alpha + \sum_{m=A}^{B} \beta_m \cdot Reversal_{a,t}^{m} + \gamma_{a,p} + \eta_t + \delta' X_{a,p,t} + \epsilon_{a,p,t},$$

where $Y_{a,p,t}$ denotes our outcome of interest (typically, number of citations) for paper p by author a in year t, $Reversal^{m}_{a,t}$ is an indicator variable denoting whether author a's first reversal occurred m years ago as of year t, $X_{a,p,t}$ denotes a vector of controls, $\gamma_{a,p}$ are author-by-paper fixed-effects, η_t are year fixed-effects, and $\boldsymbol{\epsilon}_{a,p,t}$ is an error term. Standard errors are clustered at the author level,

since this is the level at which reversals occur. In some specifications we allow for a separate set of coefficients for left-to-right reversals and right-to-left reversals for heterogeneity analysis.

We focus on authors' first reversals because we think these are by far the most likely to be relevant. That is, a first reversal reveals an author's type – as someone who appears to be open to publishing on more than one side of the political spectrum. Once this is revealed, subsequent reversal would seem to have lesser information content. Nevertheless, we do some additional analysis including subsequent reversals.

Identification using this specification relies on a version of the parallel trends assumption. That is, if reversal-undergoing authors had, counterfactually, *not* undergone an ideological reversal, then their citation trajectory would have been the same as we observe for non-reversal authors. A key reason to be concerned about the validity of this assumption is plausibly-correlated changes – specifically, the conjecture that authors may begin to publish papers of different quality around the same time they undergo an ideological reversal. We deal with this concern in two ways. First, we show directly that the quality of papers doesn't seem to change post-reversal. That is, there is no (economically or statistically) significant change in the average impact factor of the journals an author publishes in prior or subsequent to their reversal. Second, we can restrict our analysis to papers published prior to an author's ideological reversal. We show that our main result is unaffected by this sample restriction to the pre-reversal subset of papers. Together, we consider these findings as encouraging for the veracity of the assumption underlying our main specification.

We also deploy a traditional static difference-in-differences specification to show that our results go through just the same under this very standard approach and are in no way driven by the particulars of our main specification:

$$Y_{a,p,t} = \alpha + \beta \cdot ReversalStatic_{a,t} + \gamma_{a,p} + \eta_t + \delta' X_{a,p,t} + \epsilon_{a,p,t},$$

where $ReversalStatic_{a,t}$ denotes whether author a's first reversal has occurred and all other variables are as above.

4 Results

4.1 Descriptive Statistics

We begin by establishing some basic facts through descriptive analysis of the data.

Fact #1: The typical author mostly publishes results on one side of the political spectrum.

We define an author as left-wing if they published more left-wing papers than right-wing ones over the past five years, and we analogously define right-wing authors. We observe that, on average, left-wing authors publish 6.5 left-wing papers for every 1 right-wing paper. And right-wing authors publish 4.9 right-wing papers for every 1 left-wing paper. In other words, there is substantial polarization at the author level. These findings also suggest that our categorization of papers by ideology is not a meaningless one driven by noise. If it was, it would be essentially random whether an author ends up defined as right- or left-wing, and therefore both ratios would be close to 1-for-1.

Fact #2: Ideological reversals are not uncommon.

39.6% of the authors in our dataset undergo at least one ideological reversal over the course of their career. The average number of ideological reversals for an author in our dataset is 1.2. In other words, reconciling Fact #1 with Fact #2, even though authors mostly stick to publishing results on one side of the political spectrum, many authors will (occasionally) publish a result on the other side.

Fact #3: Ideological reversals become much more common later in an authors' career.

For authors in the first 10 years of their career, 4.4% have undergone at least one reversal (average number: 0.08). For authors between their 10th and 20th year of their career, 39.4% have undergone at least one reversal (average number: 1.17). For authors between their 20th and 30th

year of their career, 41.6% have undergone at least one reversal (average number: 1.02). This non-linear pattern suggests that career-advancement concerns such as tenure may play a role. Authors may feel more comfortable deviating from their past results once they are tenured or once they are elevated to other important positions in the profession (e.g., journal editor/conference organizer/etc.).

Fact #4: Most ideological reversals are not permanent defections to the other side of the political spectrum.

The ratio of own-ideology to opposing-ideology papers published by an author substantially declines after a reversal, but the ratio does not invert. In other words, an ideological reversal typically does not suggest a permanent switch to the other side. Rather, it suggests more openness to publishing results on both ideological sides. In this vein, conditional on having at least one ideological reversal, the mean number is 3.06. We feel that this fact is particularly important for interpreting the reaction to an ideological reversal. A plausible interpretation of an author's first reversal is that they have revealed themselves as someone who will likely be publishing results on both sides of the political spectrum – seemingly following the science where it takes them – rather than exclusively publishing results of one political valence.

Fact #5: Ideological reversals occur much more frequently amongst authors who are (initially) categorized as right-wing.

61.6% of authors initially categorized as right-wing eventually undergo a reversal at some point in their career, whereas 35.1% of authors initially categorized as left-wing authors undergo one. In other words, authors who are initially left-wing (i.e., authors who start their career publishing papers with left-wing results) are more likely to continue publishing only left-wing results throughout their career.

4.2 Inferential Statistics

Equipped with these facts as context, we now turn to our regression results. We begin with Figure 2, which can essentially be thought of as a bridge between our descriptive and inferential analysis. Figure 2 shows the share of papers categorized as left-wing written by an author in the years before and after his/her reversal. Unsurprisingly (and essentially definitionally), a right-to-left reversal is associated with a substantial increase in the left paper share, whereas a left-to-right reversal is associate with a substantial decline in the left paper share. No meaningful pre- trends are apparent. It is noteworthy, however, that the magnitude of the coefficient is roughly 0.5 in the first few years after the reversal, declining to 0.2 twenty or more years after the reversal. If every reversal was associated with a total defection to the opposite side of the ideological spectrum (only publishing papers on one's new side), the coefficient would be a steady 1.0. The smaller coefficients we observe here again underline the fact that an author who has a reversal tends to publish papers on both sides of the ideological spectrum in subsequent years, which is useful for interpreting the way reversals are perceived.

Figure 3 displays the results of our baseline specification with all papers published in Top 100 journals from 1990 to the present in our sample. There is a sharp clear increase in citation count following an ideological reversal with essentially no evidence of pre- trends. In other words, papers by authors that have ideological reversals do not appear to be on different trajectories pre-reversal, but a clear divergence occurs post-reversal, with authors who undergo reversals obtaining additional citations. The citation boost accumulates to approximately 9 over a one-decade period and 30 over a two-decade period. Note that standard errors are extremely tight thanks to the considerable size of our dataset.

Figure 4 displays the results of several robustness checks on our baseline specification. Panel 1 restricts our sample to authors who had a reversal at some point in their career. Therefore the control group is authors who have not yet had an ideological reversal (but will do so at some point in the future) – a group plausibly more comparable to other authors undergoing reversals. Results are very close to the baseline finding both qualitatively and quantitatively. Panel 2 restricts to the analysis of within-topic reversals. Most events measured as ideological reversals occur across

topic areas – for example, an author publishes a tax paper with a right-wing conclusion followed by a minimum wage paper with a left-wing conclusion. However, this might be less noticeable from an audience perspective – or less striking – than a within-topic reversal, so one might ask whether the latter triggers a different response. This does not appear to be the case; again, the result is similar to baseline. Panel 3 adds career length and paper age fixed-effects to the baseline specification, again yielding similar conclusions. Panel 4 restricts the sample solely to papers included in the meta-analyses we collected. In other words, it does not rely on machine learning at all. The result is qualitatively the same as our main result and remains statistically significant. The only differences are wider standard error bars (due to a much smaller sample size) and a larger effect size (due to a higher average number of citations amongst papers in these meta-analyses compared to the typical Top-100 paper).

As noted in the Identification section of the paper, identification in this context relies on a version of the parallel trends assumption: that if reversal-undergoing authors had, counterfactually, not undergone an ideological reversal, then their citation trajectory would have been the same as we observe for non-reversal authors. This assumption is violated if authors begin writing papers of an improved quality around the same time as they have a reversal. In this case, one might worry that the boost in citations is due not to the reversal but to the improved quality of the papers being written by the author. We confront this important concern in two ways, both displayed in Figure 5. First, in Panel 1, we show directly that there is no evidence of a change in average paper quality coincident with an ideological reversal – the average impact factor of the journals in which the author is publishing does not change. Second, in Panel 2, we restrict our sample to papers the author published *before* his/her reversal. These papers, too, begin to accumulate significantly more citations after the author's reversal paper is published – and the magnitude of the effect isn't much different from baseline.

Because Web of Science reports the date a paper is published, not the date it is first available as a working paper, one might worry that this leads to some measurement error regarding when an ideological reversal actually occurs and is observed by others. To deal with this issue, for every economics paper in a Top 20 journal, we scraped Google Scholar to obtain the date that a working

paper version first appeared. In Figure 6, we use these dates instead of the publication dates from Web of Science to measure reversals. Panel 1 runs the baseline specification using these alternative dates. Panel 2 runs the version that restricts to papers whose first version appeared prior to the reversal paper (i.e., the analogue of the robustness check we performed in Figure 5). Qualitatively, the pattern is similar – a boost in citations following an ideological reversal. In Panel 1, there is some evidence of a pre- trend – the author's citations are on a mildly declining trend prior to the reversal. However, the magnitude of this pre- trend is small compared to the post-treatment effect size and, perhaps more importantly, the pre- trend vanishes in Panel 2, after performing the sample restriction to a more exogenous setting. The increased magnitude of the effect in both panels relates to the fact that our sample here is papers in Top 20 journals, which receive more citations on average than papers in Top 100 journals.

In Figure 7, we study whether there are meaningful differences in the consequences of a right-to-left reversal versus a left-to-right reversal. We find no such differences – both are rewarded to roughly the same extent. Figure 8 studies the effects of an author's reversal on the share of citations the author's papers receive from left-wing authors. Panel 1 focuses on all papers written by the author – pre- and post-reversal. Left-to-right reversals are associated with a fall in this share, whereas right-to-left reversals are associated with an increase – both in the neighborhood of 0.2 to 0.25. In other words, the citation boost received by authors who undergo a left-to-right (right-to-left) reversal is driven by right-wing (left-wing) authors. Panel 2 focuses on only papers written pre-reversal, and the result is the same. This means the new right-wing (left-wing) audience garnered by a left-to-right (right-to-left) reversal audience also engages with and cites the author's previous left-wing (right-wing) papers. In other words, ideological reversals ultimately end up exposing readers to ideas from the opposite side of the political spectrum, which can help undermine echo chambers.

In Table 1, we repeat the key analyses from throughout this section, instead applying the static difference-in-differences regression approach. Results are qualitatively similar; that is, reversals are associated with increased subsequent citations (Column 1), even when one restricts analysis to the non-machine-learning sample of papers found in our meta-analyses (Column 2) or to papers

published before the author's reversal (Column 3). There is, at most, a very small *decline* in paper quality – measured by impact factor of the journal in which the paper is published – after a reversal (Column 4). There is no substantial heterogeneity in the effects by direction of reversal (Column 5). And ideological reversals expose readers to ideas from the opposite side of the political spectrum as the new audience acquired by a reversing author cites their older pre-reversal papers as well.

5 Conclusion

In order to explore a microcosm of the debate over whether political ideology and personal preferences of scientists hobble the progress of science, we study what happens when a researcher in economics academia publishes a paper with an ideological polarity opposite the author's previous work. What we find is that such ideological reversals are not penalized but rather rewarded, with the reversal leading to a sharp and persistent boost in citations accumulated by the author. This result holds not only for new papers published by the author, which could plausibly change in other ways correlated with the reversal, but even for old papers published prior to the reversal. Furthermore, authors who undergo left-to-right (right-to-left) reversals not only attract a new right-wing (left-wing) audience for their recent work, this new audience also engages with and cites the author's previous left-wing (right-wing) papers. We assess this result as encouraging for the state of economics academia, implying that researchers are not incentivized to stay loyal to one side of the political spectrum and eschew publishing results of the opposite political polarity. If anything, they appear to be incentivized to follow the science where it takes them.

We hope that our work inspires further study of the effects of incentives within academia. The implications of the tenure system – and the impacts of varying specific facets of its design – appear to be understudied, particularly given recent policy changes surrounding the tenure system in states such as Texas, Georgia, and Florida. The consequences of advising networks for the propagation and development of scientific progress strikes us as another area likely to yield fruit, particularly given increased interest in analysis of networks over the past decade. The relevance of different topic areas for industry and the private sector – and how this relates to scientific

progress in academia – could be of interest as well. There are a multitude of topics in this broader area that are feasible to study and enhance our understanding of the way science and academia function. While Popper might be encouraged by the results of this paper, there are plenty of other contexts where Kuhn might yet win the day.

References

Card, D., & DellaVigna, S. (2013). Nine facts about top journals in economics. Journal of Economic literature, 51(1), 144-161.

Card, D., & DellaVigna, S. (2014). Page limits on economics articles: Evidence from two journals. Journal of Economic Perspectives, 28(3), 149-168.

Card, D., & DellaVigna, S. (2020). What do editors maximize? Evidence from four economics journals. Review of Economics and Statistics, 102(1), 195-217.

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

Javdani, M., & Chang, H. J. (2023). Who said or what said? Estimating ideological bias in views among economists. Cambridge Journal of Economics, 47(2), 309-339.

Jelveh, Z., Kogut, B., & Naidu, S. (2024). Political language in economics. The Economic Journal, 134(662), 2439-2469.

Koffi, M. (2024). Innovative ideas and gender inequality. University of Toronto Working Paper.

Koffi, M., & Marx, M. (2023). Cassatts in the Attic. NBER Working Paper No. 31316.

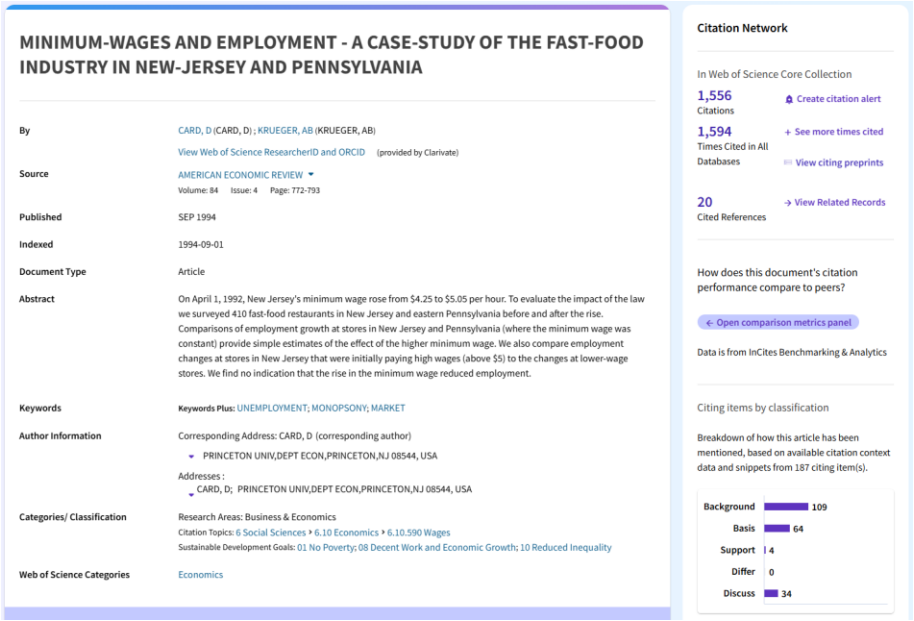
Koffi, M., Pongou, R., & Wantchekon, L. (2024). The Color of Ideas: Racial Dynamics and Citations in Economics. NBER Working Paper No. 33150.

Ringgenberg, M. C., Shu, C., & Werner, I. M. (2023). The politics of academic research. Fisher College of Business Working Paper.

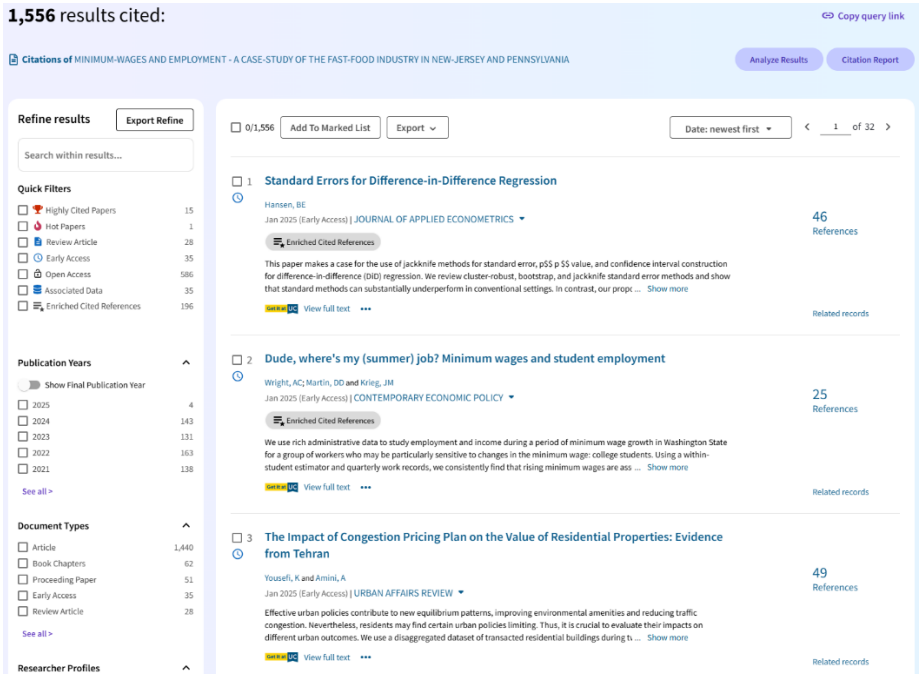
Tables and Figures

Figure 1: Web of Science Example Page

Panel 1: Main Page with Abstract and Metadata for Card and Krueger (1994)

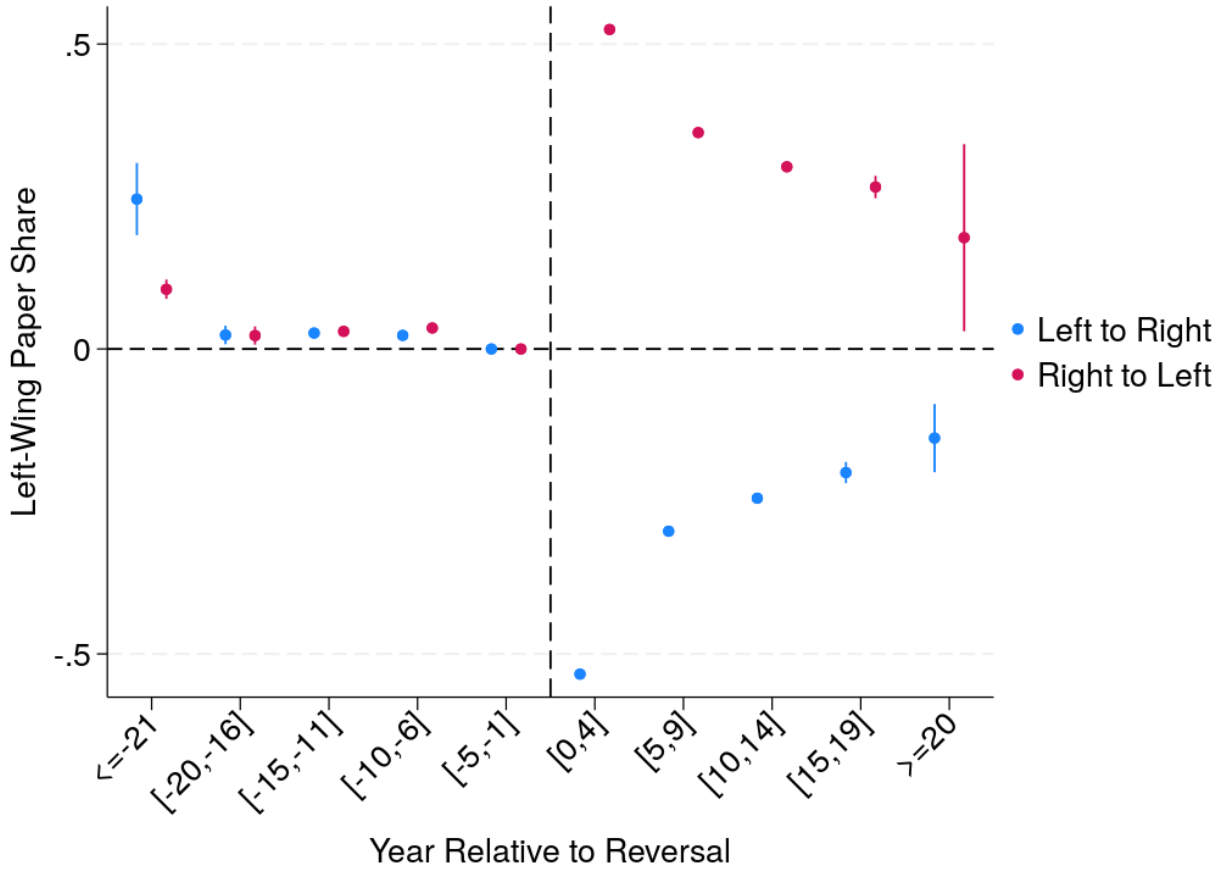


Panel 2: Papers Citing Card and Krueger (1994)



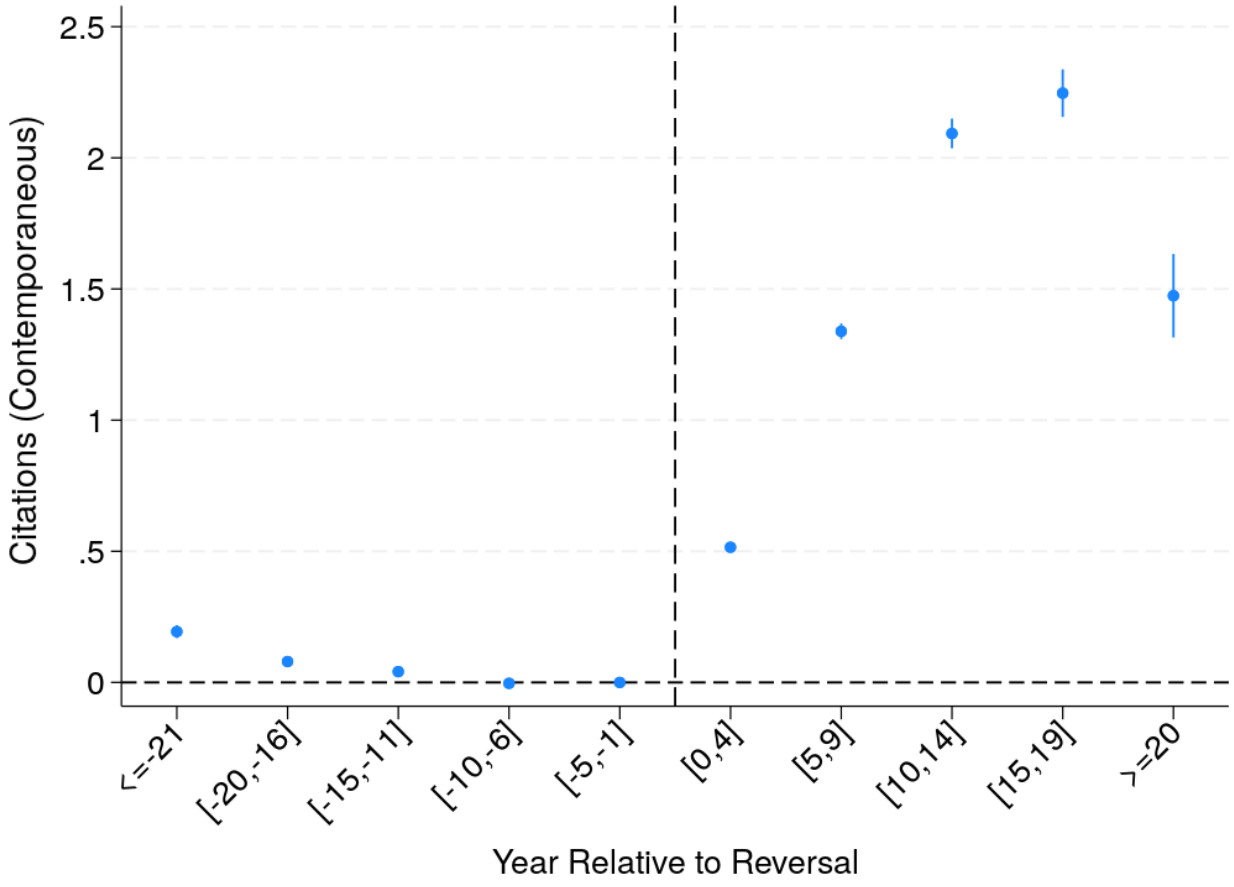
Note: This figure displays an example of a typical Web of Science page. Panel 1 shows the main page, which contains the title, abstract, author, journal, and topic information. Panel 2 shows the citations page, which contains a list of every paper that has cited this paper.

Figure 2: Impact of Ideological Reversal on Share of Author's Papers that are Left-Wing



Note: This figure displays the effects of an ideological reversal on the share of an author's papers that are left-wing using a variant of our event-study specification that allows for separate coefficients for left-to-right and right-to-left reversals. As one would expect, a right-to-left (left-to-right) reversal is associated with an increase (decrease) in the left-wing paper share. However, the magnitude of the coefficients (i.e., far below 1) indicates that a reversal is best interpreted as signaling a propensity to write papers with results on both sides of the spectrum rather than a total defection to the opposite side.

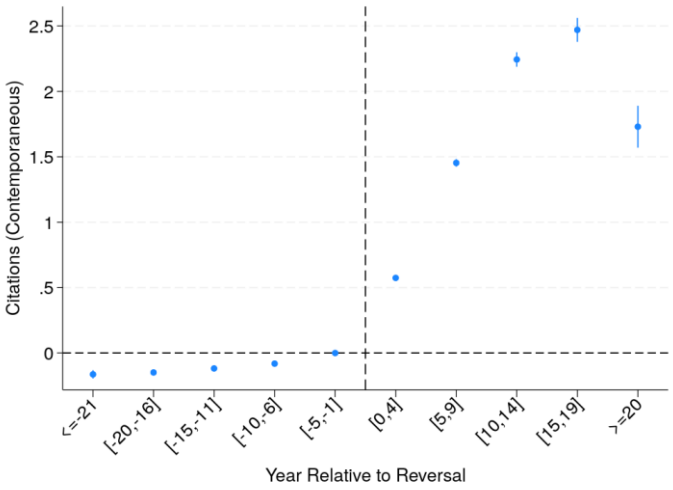
Figure 3: Impact of Ideological Reversal on Citations



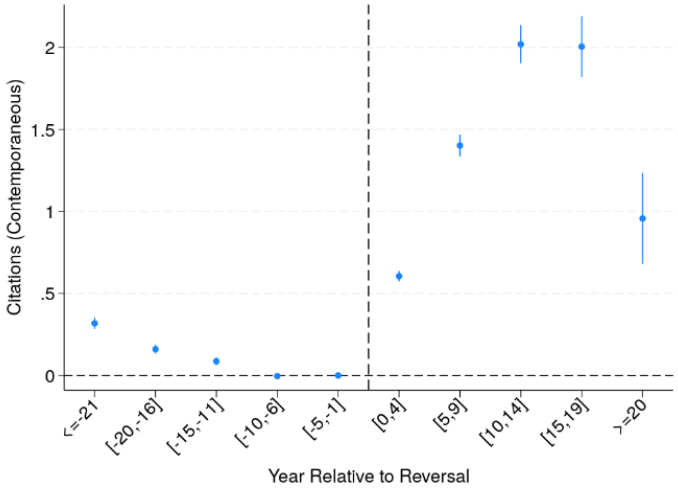
Note: This figure displays the effects of an ideological reversal on the number of citations received by an author's papers. The regression is at the paper level and therefore the coefficients represent the effect on the average paper by the author. In particular, in the first half-decade after publication the average paper by the reversal-undergoing author accumulates an extra 0.5 citations per year. Then an additional 1.3 per year in the next half-decade. And so on.

Figure 4: Impact of Ideological Reversals on Citations – Robustness

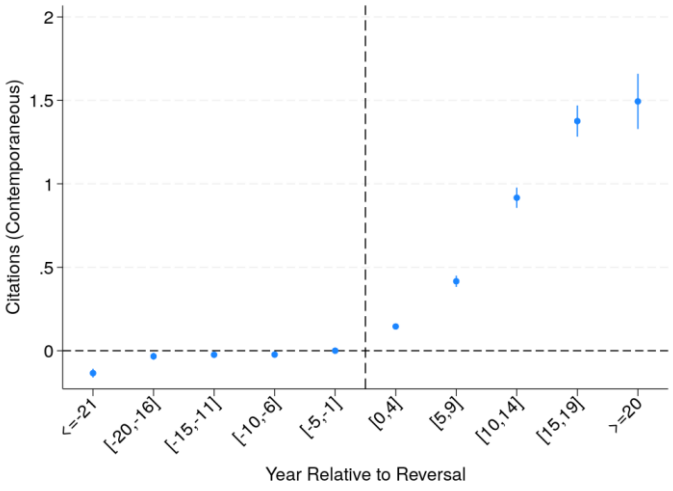
Panel 1: No Never-Reversers



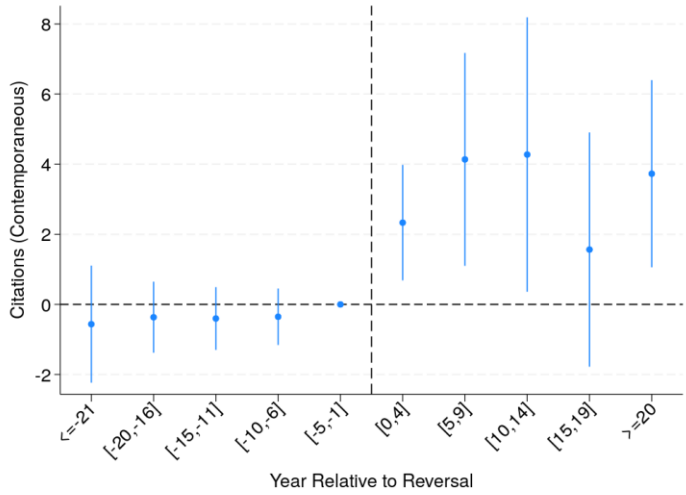
Panel 2: Within-Topic Reversals Only



Panel 3: Career Length & Paper Age FEs



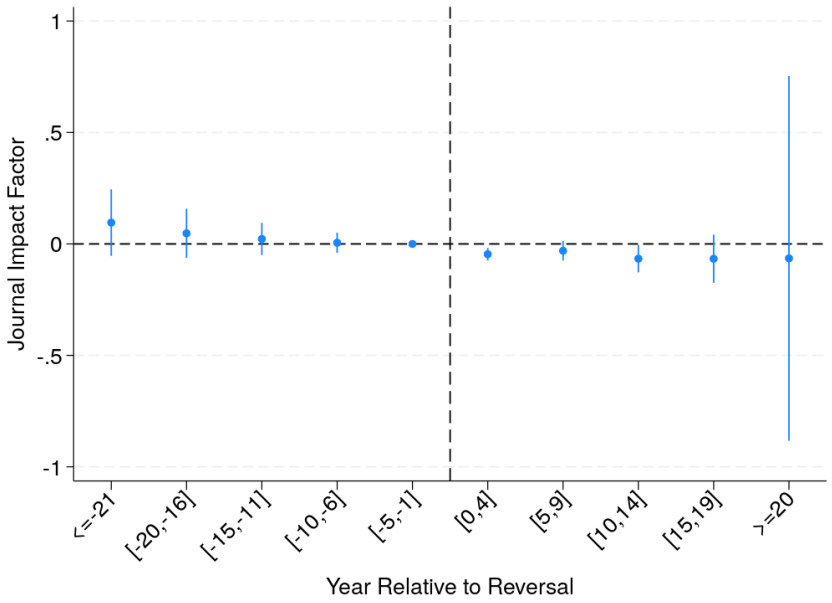
Panel 4: No Machine Learning



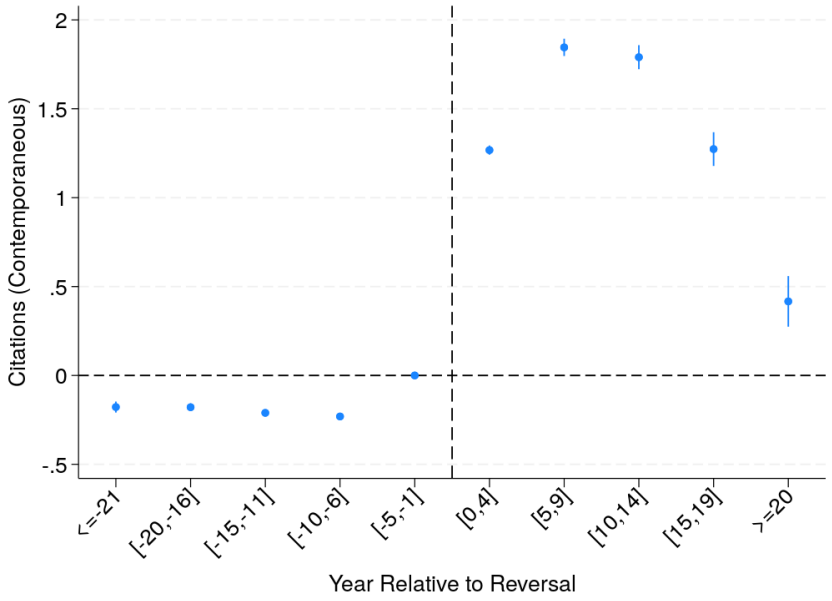
Note: This figure displays the effects of an ideological reversal on the number of citations received by an author's papers – with several robustness checks applied. Panel 1 restricts the sample to authors who have an ideological reversal at some point during their career. Panel 2 restricts the sample to authors between the 10th and 20th years of their career – the period where the propensity to have a reversal is highest. Panel 3 adds career length and paper age fixed effects. Panel 4 uses only the sample of papers included in the meta-analyses we collected; therefore it does not rely at all on our machine learning algorithm. In all cases, a boost in citations is evident post-reversal.

Figure 5: Impact of Ideological Reversals – Accounting for Paper Quality Changes

Panel 1: Effects on Average Paper Quality (Impact Factor of Journal)



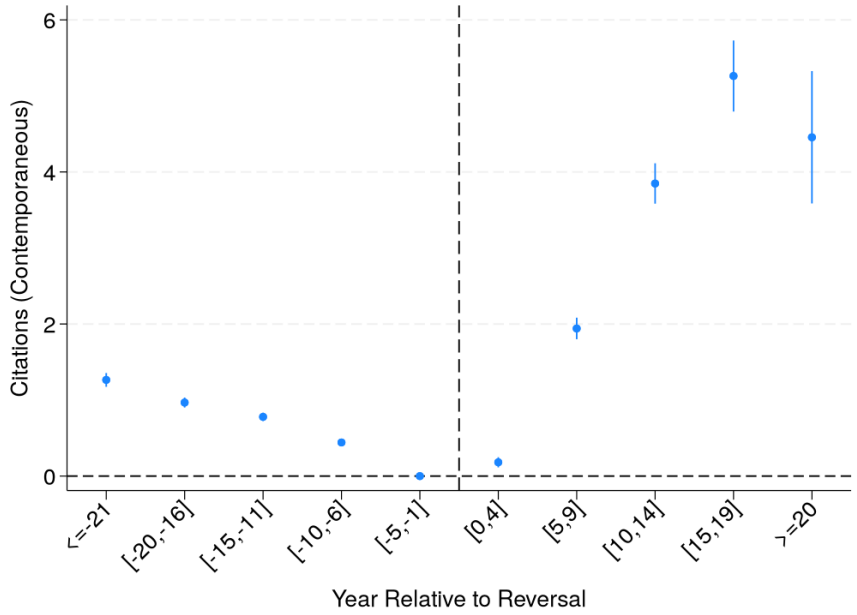
Panel 2: Effects on Citations (Restricting Sample to Pre-Reversal Papers)



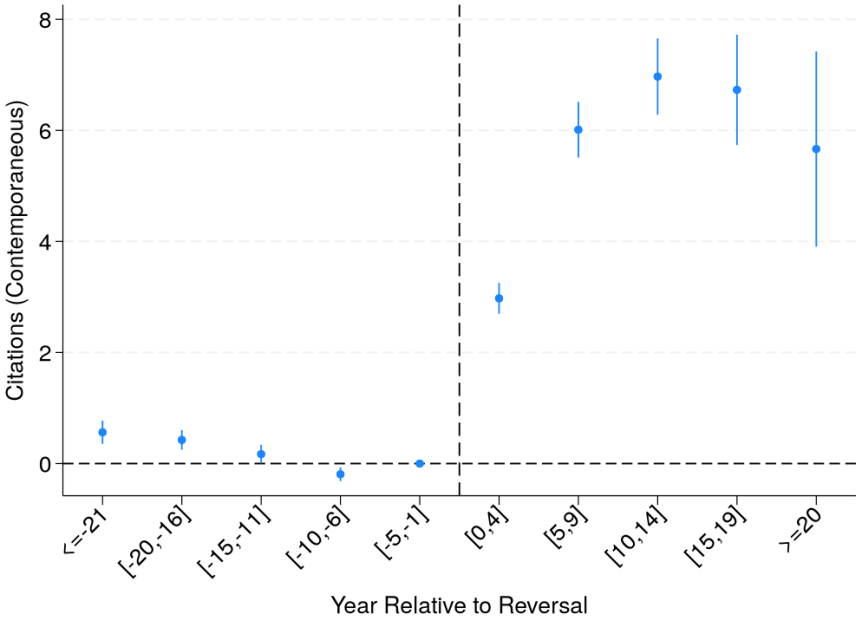
Note: Panel 1 of this figure displays the effects of an ideological reversal on the average impact factor of journals the author publishes in. There does not appear to be any meaningful increase (if anything, a miniscule decline) in the quality of papers around the time of an author's reversal. Panel 2 displays the effect on citations, as in our baseline specification, but the sample is restricted to papers already published prior to an author's reversal. Again the pattern is similar to baseline.

Figure 6: Impact of Ideological Reversals on Citations – Using Google Scholar Dates

Panel 1: Baseline

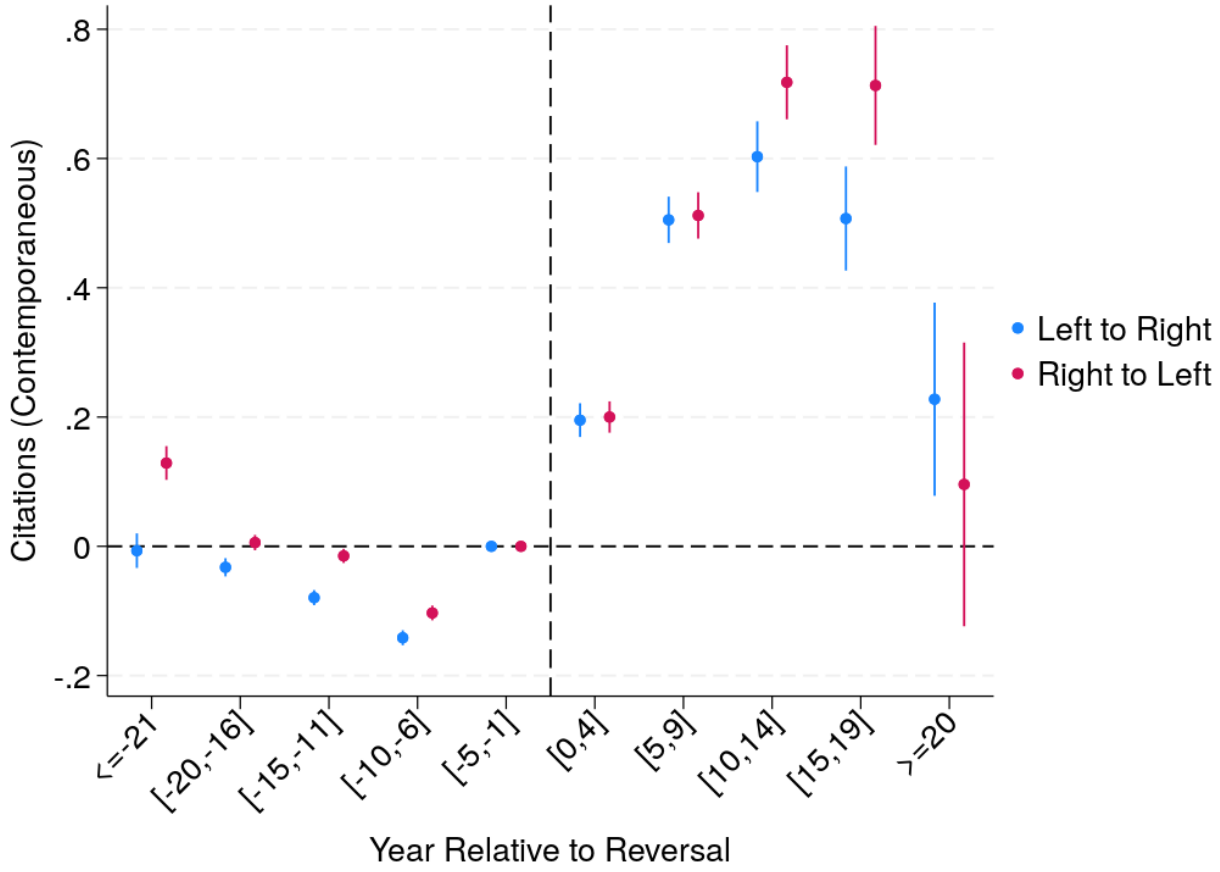


Panel 2: Restricting Sample to Pre-Reversal Papers



Note: This figure uses the date of the first available working paper version of each paper, scraped from Google Scholar, as opposed to the publication date from Web of Science. This is responsive to concerns that papers are typically made public – and therefore observed by other economists – prior to publication, so the publication dates may measure with some error the actual ideological reversal. Panel 1 runs our baseline specification (corresponding to Figure 3) with these Google Scholar dates. Panel 2 runs a specification that restricts to papers available prior to the first available date of the reversal paper (corresponding to Figure 5, Panel 2). Both restrict to Top 20 journals since scraping Google Scholar is even more involved than scraping Web of Science. Results are qualitatively similar to baseline.

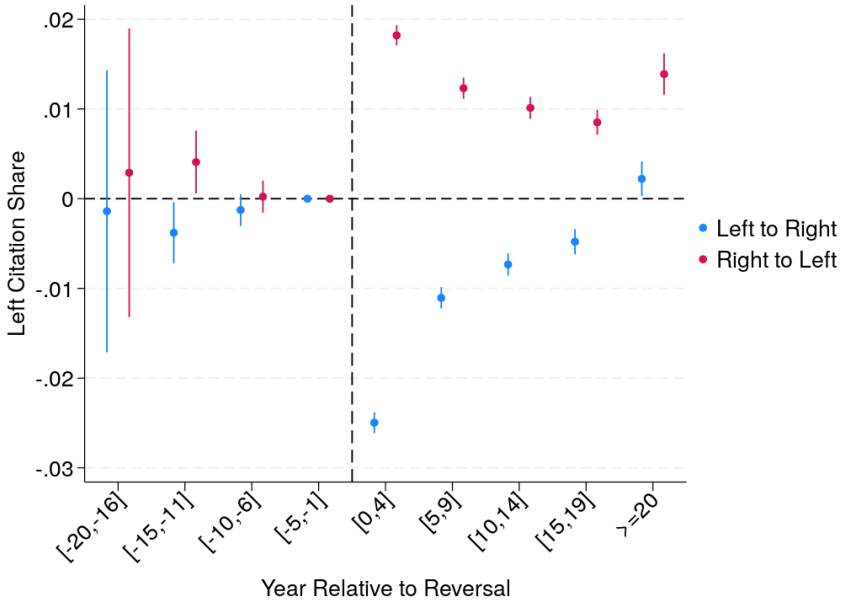
Figure 7: Impact of Ideological Reversals on Citations – Heterogeneity by Ideology



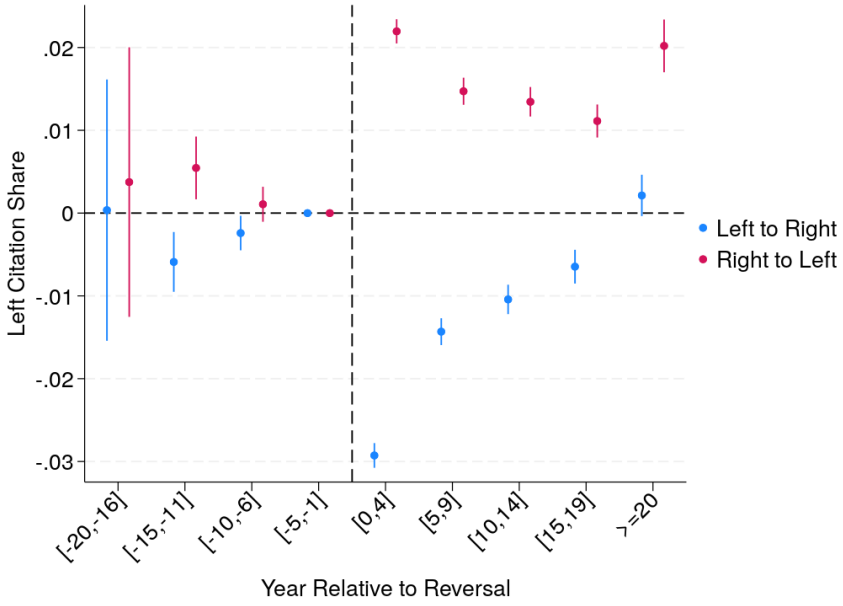
Note: This figure displays the effects of an ideological reversal on the number of citations received by an author's papers – allowing for heterogeneity between left-to-right and right-to-left reversals. There is no meaningful heterogeneity.

Figure 8: Impact of Ideological Reversals on Share of Citations Received from Left-Wing Authors

Panel 1: Entire Sample



Panel 2: Pre-Reversal Papers



Note: This figure displays the effects of an ideological reversal on share of citations received by a paper that comes from left-wing authors. Left-to-right (right-to-left) reversals are associated with a boost in citations driven by right-wing (left-wing) authors. This pattern is apparent even when restricting to papers published before the author's ideological reversal, as in Panel 2. This means the new right-wing (left-wing) audience goes back and reads/cites the author's pre-reversal left-wing (right-wing) papers as well.

Table 1: Model Performance and Selection

Best CV Score	Train Accuracy	Test Accuracy	Model
Models Trained on Abstracts			
0.6625	1.0000	0.6977	Logistic Regression with CountVectorizer
0.6935	0.9571	0.7209	Logistic Regression with TF-IDF
0.7177	0.9509	0.7674	Multinomial Naive Bayes with CountVectorizer
0.7177	0.9141	0.7674	Multinomial Naive Bayes with TF-IDF
0.6622	0.9632	0.5814	Random Forest with CountVectorizer
0.6563	0.9939	0.5814	Random Forest with TFIDF
0.6560	1.0000	0.6279	Adaboost with CountVectorizer
0.6749	1.0000	0.4884	AdaBoost with TFIDF
0.5582	1.0000	0.6047	XGBoost with CountVectorizer
0.6013	1.0000	0.5814	XGBoost with TF-IDF
Models Trained on Full Texts			
0.7172	0.5399	0.4884	Logistic Regression with CountVectorizer
0.7356	0.9018	0.7442	Logistic Regression with TF-IDF
0.7297	0.9325	0.6977	Multinomial Naive Bayes with CountVectorizer
0.7529	**0.8282**	**0.7907**	**Multinomial Naive Bayes with TF-IDF**
0.6940	0.5153	0.5581	Random Forest with CountVectorizer
0.7118	0.4724	0.5116	Random Forest with TFIDF
0.7408	0.5644	0.5581	Adaboost with CountVectorizer
0.6767	0.5460	0.5349	AdaBoost with TFIDF
0.7065	0.5706	0.5349	XGBoost with CountVectorizer
0.6824	0.5644	0.5814	XGBoost with TF-IDF

Note: This table displays the results of the various machine learning models we tested. We ultimately Multinomial Naïve Bayes with TF-IDF, as it produces the best results without an overfitting problem

Table 2: Results from Static Difference-in-Differences Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	Citations	Citations	Citations	Journal Impact Factor	Citations	Share of Citations from Left-Wing Papers
Sample Restriction:	None	Meta-Analysis Papers	Pre-Reversal Papers	None	None	Pre-Reversal Papers
Ideological Reversal	6.539***	30.79*	15.20***	-0.0423**		
	(0.183)	(16.53)	(0.318)	(0.0166)		
Left-to-Right Reversal					6.792***	-0.0316***
					(0.263)	(0.00112)
Right-to-Left Reversal					6.147***	0.0438***
					(0.381)	(0.00279)
R-Squared	0.153	0.352	0.172	0.530	0.153	0.169
Observations	29,041,200	38,880	19,195,596	324,719	29,031,704	3,738,072

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 1% level. Standard errors clustered at the author level are displayed in parentheses. This table displays the results of the static difference-in-differences regression. Column (1) shows the results on the whole sample with citations as the outcome. Column (2) shows the results on the sample restricted to papers published prior to the author's reversal. Column (3) shows the results restricted to meta-analysis papers (i.e., our no-machine-learning sample). Column (4) shows the results on the whole sample with the impact factor of the journals in which the author publishes as the outcome. Column (5) shows the results on the whole sample with citations as the oucome, allowing for heterogeneity in the effects of left-to-right versus right-to-left reversals. Column (6) shows the results on the pre-reversal sample with the share of citations received from left-wing papers as the outcome. Results are qualitatively similar to those from the event-study specifications shown in the figures above.