Political Polarization, Scientific Collaboration and Consequences for the Rate and Direction of Innovation

Jino Lu (WUSTL, jinolu@wustl.edu)
Florenta Teodoridis (USC, florenta.teodoridis@marshall.usc.edu)
Joe Raffiee (USC, raffiees@marshall.usc.edu)
Dan Fehder (USC, fehder@marshall.usc.edu)

ABSTRACT

This article studies the impact of political polarization on knowledge production. We construct a panel of 337,494 academic scientists linked with longitudinal voter registration records containing political party affiliation. Using this data, we first show that political polarization affects scientific collaboration. Following the divisive 2016 US presidential election, we find that Democrat and Republican scientists are 3.5 percent less likely to collaborate as measured by co-authored publications. We then explore the impact of reduced inter-party collaboration on knowledge production outcomes. We find that Democrat (Republican) scientists who collaborated with Republicans (Democrats) prior to the 2016 election subsequently experience a 5.6 decrease in publications, a 4.7 percent decrease in citation-weighted publications, and are 16.3 percent less likely to publish in new topic areas. Taken together, our results suggest that political polarization disrupts co-authorship collaboration networks and that the disruption of such networks negatively impacts the rate and direction of knowledge produced.

I. Introduction

Political polarization and ideological divide in the United States has reached historic highs. In particular, affective polarization – the tendency for members of one political party to dislike and distrust members of the opposing party (Iyengar, Sood, and Lelkes, 2012) – has risen sharply over the past several decades (Boxell, Gentzkow, and Shapiro, 2022). The consequences of rising inter-party animosity are wide-ranging and not just restricted to the realm of politics and political cooperation (McConnell et al., 2018). For example, individual preferences to associate with co-partisans and avoid counter-partisans have been documented across a variety of economic and social settings (Alford et al., 2011; Chen and Rohla, 2018; Mosleh et al., 2021). Indeed, Americans increasingly segregate along party lines residentially (Brown and Enos, 2021; Kaplan, Spenkuch, and Sullivan, 2022) and with whom they converse and trust (Gentzkow and Shapiro, 2011).

Motivated by these trends, we ask whether rising levels of affective polarization impact knowledge production. We explore this question for two reasons. First, knowledge production is a key driver of economic growth (Romer, 1990). Accordingly, there has been sustained interest in uncovering factors which influence the rate and direction of inventive activity (Nelson, 1982; Lerner and Stern, 2012) and ideological beliefs are reflected in research behavior (Solow, 1956). Second, knowledge production is increasingly a function of collaborative efforts amongst scientists (Jones, 2009, 2010; Agrawal, Goldfarb and Teodoridis, 2016), both within and across university and institutional boundaries (Jones, Wutchy, and Uzzi, 2007). In theory, increases in affective polarization has the potential to impact collaboration decisions and disrupt otherwise productive research teams, the latter of which has been shown to have adverse effects on knowledge production outcomes (Azoulay, Graff Zivin, and Wang, 2010; Jaravel, Petkova, and Bell, 2018). Yet despite these observations, our understanding of the potential impact of rising affective polarization on the production of knowledge remains quite limited, perhaps in part due to prior data limitations (Fehder

et al., 2024). As described by Bell et al. (2018: 648), "most sources of data on innovation (e.g., patent records) do not record even basic demographic information, such as an inventor's age or gender", and certainly such data sources do not contain measures of partisan affiliation.

In this study, we construct a novel large-scale dataset which longitudinally links individual scientists with voter registration records containing partisan affiliation, thereby allowing us to explore the relationship between rising levels of affective polarization and knowledge production. We collect data on individual scientists from Dimensions, a bibliography database containing millions of research publications, and source longitudinal voter registration records via the national voter file obtained from the commercial data supplier L2. We link individual scientists with voter registration records using the fastLink probabilistic record linking algorithm, an approach which results in a final panel dataset of 337,494 scientists paired with their partisan affiliation.

Using this data, we implement a difference-in-differences empirical framework in which we examine how increased levels of affective polarization following the unexpected outcome of the divisive 2016 U.S. presidential election impact collaboration between Democrat and Republican scientists. We then investigate the consequences of rising affective polarization on three important knowledge production outcomes: (1) the number of publications produced (quantity), (2) the citation-weighted number of publications produced (quality), and (3) the scope of topics explored (diversity). Importantly, our empirical strategy does not rely on the assumption that the outcome of the 2016 election is the sole and driving cause of rising affective polarization in the U.S. – it is well-documented that levels of affective polarization in the U.S. have been increasing over time (Boxell et al., 2022). Rather our strategy rests on the assumption that the result of the 2016 election unexpectedly accelerated the upward affective polarization trend within the U.S., an assumption we corroborate with polling (Katz, 2016) and political sentiment data (Tyler and Iyengar, 2024).

To explore the impact of affective polarization on research collaboration, we study the collaboration likelihood within scientist pairs, comparing the probability of co-publication(s) within pairs of scientists from opposing political parties (i.e., Democrat and Republican) to the probability of co-publication(s) within pairs of scientists from non-opposing political parties (i.e., scientists from the same party or pairs including Independents). To control for unobserved characteristics which may impact collaboration between scientists, we include scientist-pair fixed effects throughout our pair-level analysis. A key identifying assumption is that the probability of collaboration between scientists from opposing parties and collaboration between scientists from non-opposing parties will follow parallel trends prior to the outcome of the 2016 presidential election. We test this assumption to demonstrate an absence of pre-trend differences – pairs of scientists from opposing parties are no more or less likely to collaborate prior to the 2016 presidential election relative to pairs of scientists from non-opposing parties.

We find that after the 2016 election, opposing-party scientists are less likely to collaborate relative to non-opposing party scientists. Specifically, Democrat and Republican scientists are, on average, 3.5 percent less likely to collaborate relative to scientists from non-opposing parties. The magnitude of this effect is steadily increasing in the four years following 2016 and holds at both the extensive (any co-publication) and intensive (number of co-publications) margin. A further analysis of collaboration patterns reveals a shift in collaboration on the basis of partisan affiliation: after the 2016 election, Democrat and Republican scientists are less likely to collaborate with one another but increasingly likely to collaborate with scientists who share their partisan views and with scientists whose affiliation is Independent, although the overall quantity and quality of their collaboration networks declines. Together, these patterns are consistent with the intuition that increases in affective polarization disrupt existing research collaboration relationships between counter-partisans

and that search frictions limit the degree to which new collaborations with co-partisans and with Independents offset the loss resulting from severed ties with counter-partisans.

Our finding that Democrat and Republican scientists collaborate less after the 2016 election suggests that rising levels of affective polarization may have detrimental consequences for knowledge production. To explore this further, we shift our unit-of-analysis from the scientist-pair to the individual scientist, comparing the research output of Democrat and Republican scientists with a greater share of opposing-party co-authors (prior to 2016) with the output of scientists with a lower share of opposing-party co-authors. The intuition is that scientists with a greater share of opposing-party co-authors prior to 2016 are more exposed to the uptick in affective polarization following the 2016 election, and thus such scientists will experience greater intensity of treatment. A central identifying assumption in our scientist-level analysis is that the various dimensions of scientist output that we study – quantity of publications, quality of publications, and diversity of scientist portfolio – follow parallel trends prior to the 2016 election.

We find that Democrat (Republican) scientists with a greater share of Republican (Democrat) co-authors prior to the 2016 election produce 5.7 percent fewer publications after the 2016 election, compared to scientists with a lower share of opposing-party co-authors. The negative impact of affective polarization on scientist output extends to the quality of scientist output – Democrat (Republican) scientists with a greater share of Republican (Democrat) co-authors prior to the 2016 election produce approximately 4.7 percent fewer citation-weighted publications after the 2016 election, compared to scientists with a lower share of opposing-party co-authors. Lastly, we find that increases in affective polarization negatively impacts the diversity of topics within a scientist's portfolio and a scientist's tendency to publish in new topic areas. Specially, the research portfolio of Democrat (Republican) scientists with a greater share of Republican (Democrat) co-authors prior to the 2016 election becomes .2 percent less diverse after the 2016 election, compared

to scientists with a lower share of opposing-party co-authors, and such scientists are 16.3 percent less likely to explore new research topic areas. We confirm the absence of pre-trend differences across each of our outcome variables prior to 2016. Together, these patterns suggest that rising affective polarization has resulted in less collaboration amongst counter-partisans and that reduced collaboration in existing counter-partisan scientific teams has negatively affected the quantity, quality, and diversity of knowledge scientists produce.

From a normative perspective, we would want knowledge worker collaboration choices to be immune from ideological influence. However, conceptually, to the extent sorting occurs, that does not necessarily mean negative implications for innovation. Nevertheless, considering the increasing polarization trends, evaluating the potential impact of polarization on knowledge production has the potential to contribute to our understanding of factors that could impact the rate and direction of innovation. On the negative side, we find evidence of sorting following a step increase in political polarization that is associated with reduced productivity and diversity of explored knowledge trajectories. On the positive side, the overall effect we measure, while not trivial, is not excessive, suggesting that the overall impact, at least now, may be relatively small. However, the fact that we find any effect at all suggest further investigation is warranted as political polarization continues to be on the rise.

II. Increased Political Polarization Following the 2016 Presidential Election

Donald Trump's election as President of the United States on November 8, 2016 is considered one of the largest presidential upsets in American history,[1] with virtually all pre-election polling and prediction market data suggested Trump had a slim chance of victory. For example, prior to the election CNN gave Trump 9% chance of winning the election, PredictWise, which uses both

[1] https://www.politico.com/story/2016/11/election-results-2016-clinton-trump-231070

prediction market and poling data and gave Trump a 12% chance of winning, Nate Silver's FiveThirtyEight's gave Trump a 28.6% chance of winning, and the New York Times's polling forecasted Trumps' chance of winning at 15%. While the trend of rising affective polarization in the US (and abroad) precipitated the 2016 presidential election (Boxell et al., 2022), the degree of affective polarization within the US following the unexpected result of the 2016 election accelerated, as reflected in the American National Election Studies (ANES). Figure 1 plots affective polarization using the ANES data between 2004 and 2020. As displayed in Figure 1, the degree of affective polarization (calculated as the difference between a respondent's feelings toward their own party and the respondent's feelings toward the opposing party) has been rising since 2004, however, there is a noticeable uptick in the slope after 2016, suggesting affective polarization began to increase at an increasing rate, an empirical result consistent with surveys conducted by the Pew Research Center.[2]

III. Data and Sample Construction

We use data from two primary data sources. We obtain scientist data from Dimensions (Herzog, 2020) and voter registration records from the national voter file obtained from the commercial data supplier L2.

Dimensions has been increasingly used in the innovation literature, given its high coverage of academic publications and author disambiguation algorithm (Arora et al., 2024; Yin, 2022; Fry and MacGarvie, 2023). We restrict our analysis to scientists with at least one publication record during 2014 and 2021 because the longitudinal voter registration data we obtain is available only since 2014. We further restrict our sample to scientists with at least one paper during 2014 and 2016 (i.e., before period), such that they could be matched to the L2 National Voter File. We collect

[2] https://www.pewresearch.org/politics/2019/10/10/partisan-antipathy-more-intense-more-personal/

publication records for all our sample scientists between 2012 and 2021, five years before and five years after the 2016 election event.

We obtain voter registration data from the national voter data as provided by L2, a highly trusted and widely used commercial data supplier for political science scientists, congressional offices, political campaigns, political consultants, pollsters, commercial firms, media outlets, and research institutions (Brown and Enos 2021).[3] The national voter file is constructed from voter registration records, which L2 sources directly from each state's County and State Board of Election and/or the Secretary of State. The L2 data includes longitudinal voter registration records for the years 2014 to 2021. Although voter registration forms in most states prompt individuals to self-report their partisan affiliation when registering to vote, not all states allow voters to register by political party. For the 20 states where voter registration records do not contain partisan affiliation (e.g., voters do not register by party), L2 provides a predicted partisan affiliation using a proprietary classification algorithm which utilizes primary ballot selections and their extensive commercial consumer and household data.[4]

The central challenge associated with constructing our dataset which matches Dimensions publication data with L2 voter registration record is the fact that we lack a common high-quality discriminating variable which can facilitate the linking of scientists with voter registration records using conventional string-matching techniques. To overcome this challenge, our record-linking strategy follows the approach implemented in Raffiee, Fehder and Teodoridis (2022) and Raffiee, Teodoridis and Fehder (2023) which utilizes the fastLink probabilistic record-linking model (Enamorado, Fifield and Imai 2017, 2018), a model which uses the Expectation Maximization (EM) algorithm. Our final data set is comprised of 337,494 scientists in the United States and their self-

[3] Please visit https://l2-data.com/our-data/ for more information about L2.
[4] L2's party identification modeling varies from state to state, depending on the data available in that state. For more information on the party identification models, please see: https://l2-data.com/states.

report their partisan affiliation when registering to vote during 2012 and 2021. We restrict to years after a scientist have their first ever publication and before their last publication to ensure that our results are not driven by zeros in dependent variable before a scientist starts their career[5, 6] For more information about the matching process and our benchmarking of the quality of our matches, please see Appendix C.

We identify the partisan affiliation of scientists in our sample based on the matching results in the before period (i.e., 2014-2016). Among the 337,494 scientists, more than 93% have registered with only one party during 2014-2016. For the scientists who registered with multiple parties, we assign them to the parties they registered the most with during the before period[7, 8] In summary, among the 337,494 scientists in our sample, 128,951 (47.5%) are affiliated with the Democrat party, 53,366 (19.7%) are affiliated with the Republican party, 85,242 (31.4%) are Independent, and 3,656 (1.3%) are registered as "Libertarian", "Other", "Working Family Party", or "Declined to State".

IV. Empirical Strategy

We examine how increased political polarization after the 2016 presidential election shapes collaboration between scientists affiliated with the Democrat and Republican parties.

IV.1. Changes in Collaboration after the 2016 US Presidential Election

To examine scientists' change in collaboration patterns, we follow past studies and focus our analysis at the scientist-pair level (Catalini, Fons-Rosen and Gaulé, 2020). To this end, we construct

[5] We obtain the year of a scientist's first and last publications through the Dimensions database.
[6] In Appendix A, we conduct a robustness check by restricting our sample to scientists who have at least one publication after the 2016 presidential election and show that our results remain strongly robust.
[7] If a scientist registered with two parties in equal number of years, which only account for 1% in our sample, we assign the scientist to their most recent affiliation. For example, if a scientist registered with Democrat in 2014 and Republican in 2015, we assign the scientists to the Republican party.
[8] About 2% of the scientists in our sample change their partisan affiliation after the 2016 presidential election. In Appendix A, we show that our main results remain robust when restricting our sample to scientists who have never changed their partisan affiliation during our observation period.

the corresponding sample of scientist pairs from our sample of 337,494 scientists. As we run a logit model with pair fixed effects, thus scientist pairs that never collaborate drop out of our sample.[9] Our final sample at the scientist pair-year level include 3,888,253 scientist pairs that have ever collaborated during 2012 and 2021.

Formally, we estimate a difference-in-differences logit regression for the years 2012-2021:

$$Collaboration_{ijt} = \beta 2016PresidentialElection_t \times Treated\ pair_{ij} + \mu_{ij} + \gamma_t + \varepsilon_{ijt} \quad (1)$$

where $Collaboration_{ijt}$ captures whether scientist i and scientist j have any collaborations in year t.[10] $Treated\ pair_{ij}$ equals one for pairs with one Democrat scientist and one Republican scientist, and zero for other pairs, such as pairs with one scientist from Democrat party and one Independent. $2016PresidentialElection_t$ equals one for years after 2016, and zero for years before 2016. μ_{ij} captures scientist pair fixed effects, and γ_t captures year fixed effects.

We interpret a negative estimated value of the coefficient β as suggesting that after the 2016 US presidential election, which led to increased political polarization, Democrat scientists and Republican scientists were less likely to collaborate with each other, compared with other types of scientist pairs.

IV.2. Consequence on Knowledge Production Outcomes

Next, we examine the consequences of a change in Democrat and Republican collaboration on these scientists' knowledge production outcomes. To this end, we change our analysis to our

[9] In appendix, we show that our main results are robust to including a random sample of non-collaborating pairs and running an OLS model.
[10] In Appendix A, we show that our results remain robust to using a count of co-publication.

sample of 337,494 scientists and estimate a difference-in-differences Poisson regression at the scientist-year level for the years 2012-2021:

$$Y_{it} = \beta 2016PresidentialElection_t \times ShareOppositeTopQuartile_i + \mu_i + \gamma_t + \varepsilon_{it} \quad (2)$$

where $ShareOppositeTopQuartile_i$ is a dummy treatment variable, which equals one if scientist i's share of coauthors from the opposite party (during 2012-2016) is in the top quartile (i.e., greater than 23%) among all scientists, and zero if in the bottom quartile (i.e., equals to 0).[11, 12] In particular, if a scientist is affiliated with the Democrat party, this measure captures the share of coauthors during 2012-2016 from the Republican party, and vice versa for Republicans; if a scientist is Independent or others, this measure will be zero. Y_{it} captures scientists' knowledge production outcomes, including (1) publication count in year t, (2) citation-weighted publication counts in year t,

[11] We measure the intensive margin of coauthor count rather than unique coauthors. For example, consider a Democrat scientist who only have two coauthors during 2012-2016, coauthor A from Democrat party and coauthor B from Republican party. If the scientist collaborated with coauthor A three times and B only once during 2012-2016. This $ShareOpposite_i$ measure for this scientist will be 25% (rather than 50% when doing an extensive count). We do this to take into account of a scientist's intensity of collaboration with different parties, and thus exposure to the opposite party. However, our results remain strongly robust when calculating this measure using an extensive margin of coauthor count.

[12] It is important to acknowledge the critique raised by Ai and Norton (2003) that the sign and magnitude of interaction coefficients in nonlinear models do not necessarily equal the marginal effect of the interaction terms. We took a few steps to demonstrate that this does not influence the conclusion I draw from the estimations. First, we construct the treatment variable as a dummy indicator, as the concern by Ai and Norton (2003) does not hold in nonlinear models with binary interaction terms (Kolasinski and Siegel 2010, Green 2010, Puhani, 2012). Furthermore, in Appendix A, we run additional a few robustness checks. We define the treatment variable as a dummy variable which equals to one if scientist *i* had any coauthors during 2012-2016 from an opposite party, and zero otherwise. In additional, we run an OLS model instead of the logit model. The results remain strongly consistent.

(3) diversity of concepts in year t,[13, 14] (4) number of new concepts in year t, where we define a concept as being new to scientist i if the scientist has not published papers covering that concept in the prior five years. $2016PresidentialElection_t$ equals one for years after 2016, and zero for years before 2016. μ_i captures scientist fixed effects, and γ_t captures year fixed effects.

We interpret a negative estimated value of the coefficient β as suggesting that after the 2016 US presidential election, which led to increased political polarization, scientists with more coauthors from opposite parties became less productive and less likely to explore new research topics.

V. Results

We present basic descriptive statistics for all variables in Table 1.

V.1. Political Polarization and Knowledge Workers' Collaboration

Table 2 presents the results from equation (1). Column (1) present the coefficient and Column (2) displays the incidence rate ratio. The results suggest that after the 2016 US presidential election, Democrat and Republican scientists were 3.8% less likely to collaborate with each other, compared with other type of scientist pairs, such as Democrat and Independent, Democrat and others, Republican and Independent, Independent and Independent, etc. This result remains robust when restricting scientist pairs to the ones with at least one Democrat or Republican scientist,

[13] Concepts are "normalized noun phrases describing the main topics of a publication" in the Dimensions database; they are derived from publication abstracts using machine learning techniques. As an example, a publication titled "Decoherence in Adiabatic Quantum Computation" is assigned with concepts such as "adiabatic quantum computation", "quantum computation", "quantum annealing", "quantum computing", "energy eigenbasis", "quantum master equation formalism", "quantum Monte Carlo algorithm", "role of decoherence", "quantum adiabatic algorithm", "open system setting", "quantum features", "closed-system setting", "master equation simulations", "decoherence time", "master equation formalism", etc. We confirmed with an expert in quantum computer that this concept list indeed captures the essence in the abstract of the publication. See https://api-lab.dimensions.ai/cookbooks/1-getting-started/7-Working-with-concepts.html#:~:text=A%20Dimensions%20API%20query%20normally,is%20a%20top%20level%20item for more details on the concept feature in the Dimensions database.

[14] We calculate $Diversity_{it} = 1 - \sum_{k=1}^{n} Conceptpercentage_{itk}^2$, where $Conceptpercentage_{itk}^2$ is the squared percentage of scientist *i*'s publications in year *t* in concept *k*.

suggesting that the result is not driven by e.g., Independent more likely to collaborate with Independent.

We examine the timing of this effect to ensure that the decline in collaboration between Democrat and Republican scientists is driven by increased political polarization, which occurred following the 2016 presidential election. To do this, we replace $2016PresidentialElection_t$ with a series of year-specific dummy variables. We plot the estimated coefficients of the interaction between the treated dummy and these year-specific dummies (Figure 2). Each point in the graph represents the estimated difference in count of publications produced by treated scientist pairs (i.e., one Democrat and one Republican scientist) and control scientist pairs (i.e., all other pairs). The graph shows that the difference between the treated and control pairs was small and stable before the 2016 presidential election. Thereafter, the treated scientist pairs became less likely to produce publications with each other relative to control pairs, as evidenced by the lower coefficients on the graph.

V.2. Consequence for Knowledge Production Outcomes

Table 3 presents the results from equation (2). The results suggest that after the 2016 US presidential election, scientists with a larger share of coauthors with opposite political ideology produced fewer (citation-weighted) publications (Columns (1) and (2), became less diverse in their research topics, and less likely to explore new research topics. To get a better sense of the magnitude of these effects, we find that scientists with share of coauthors from opposite political ideology in the top quartile (i.e., greater than 23%) produced 7.4% fewer publications and 5.9% fewer citation-weighted publications, were 0.2% less diverse in their research topics, and covered 17.6% fewer new topics, compared with scientists with share of coauthors from opposite political ideology in the bottom quartile (i.e., equal to 0).

As before, to examine the timing of the effect, we present the year-specific differences in Figure 3. We allocate a scientist to the treatment group if their share of coauthors from the opposite political ideology is in the top quartile (i.e., greater than 23%), and to the control group if it is in the bottom 25th percentile. As shown in Figure 3, the difference in knowledge production outcomes between treated and control scientists was small and stable before the 2016 US presidential election but afterward the performance began to decline among scientists with a greater share of coauthors from the opposite political ideology relative to scientists with a smaller share of coauthors from opposite political ideology.

V.3. Mechanisms

Table 4 suggests that our main results are driven only by reduced collaboration between scientists affiliated with opposite parties, rather than by Independent or other scientists. In particular, Column (1) suggests that treated scientists (i.e., scientists who had more than 23% share of coauthor from the opposite party during 2012-2016) experience a 10.2% decline number of coauthors after the 2016 US presidential election. Furthermore, this decline is driven by the scientists' lower tendency to collaborate with coauthors with opposite political ideology, as suggested by a 15.2% decline in number of coauthors with opposite political ideology (Column 3). In contrast, these scientists are 4.7% more likely to shift their collaboration towards coauthors within the same party, as suggested in Column (5).

In Table 5, we replace the DV with number of new coauthors i.e., count of coauthors who the focal scientists have never collaborated with before the focal year. The results suggest that treated scientists were 17.3% less likely to form new collaboration with other scientists from opposite parties (Column 2), and were 6.9% more likely to work with other scientist within the same

party (Column 3). However, the gain of new coauthors from the same parties does not fully make up for the loss of the opposite party coauthors.

Taken together, Table 4 and 5 suggest that after 2016, treated scientists increasingly stop collaboration with coauthors from opposite parties and work more with coauthors from the same party. However, there is friction in their search for new coauthors. In other words, their number of new coauthors from the non-opposite party does not fully make up for their loss of coauthors from opposite party. Moreover, Table 5 Column (4) suggests that the average quality of treated scientists' coauthors declined after the 2016 election. We measure the quality of a coauthor using the average citation of the coauthor's publications in the previous five years (not including the focal year) and count the citation of each publication until up to five years following its publication.

V.4. Robustness Checks

V.4.1. Robustness for the scientist pair level results

Figure A1 and Table A1 suggest that the scientist pair level results remain robust to alternative functional forms. First, we show that the results are robust to a Poisson model, where we replace the dummy dependent variable with the scientist pair count of co-publications the focal year. Next, we depart from our main analysis specification, where we run a logit model with scientist pair-fixed effect, which drops scientist pairs who have never collaborated out of the sample, to consider an OLS model which includes a random sample of non-collaborating pairs. The results in Figure A1b & A1c and Table A1 columns (2) & (3) are consistent with our main results in Table 3 and Figure 3.

In Figure A2 and Table A2 we show that the scientist pair level results remain robust to alternative sample construction approaches. First, in our main results, we restrict to scientists who have at least one paper before 2016, such that we can match scientists in Dimension database to the L2 database. However, we do observe that some scientists drop out of the sample right after 2016.

One might argue that these scientists drop out because they were affected by the increased polarization and loss of coauthors; however, one might be concerned that these scientists drop out because of other reasons unrelated to political ideology, such as retirement. Therefore, we run a robustness check by restricting our sample to scientists with at least one paper during 2017 and 2021 (i.e., after period) to ensure that our result of change in collaboration is not driven by scientists exiting scientific publication efforts after 2016. Reassuringly, the results in Figure A2a and Table A2a are consistent with our main results. Second, in Figure A2b and Table A2b, we restrict our sample to scientist pairs with at least one Democrat or Republican scientist. The results are also in line with our main results in Table 2 and Figure 2, suggesting that our main finding is indeed driven by declined collaboration between Democrat and Republican coauthors, rather than increased collaboration among control scientist pairs (e.g., not driven by increased collaboration within Independent scientists). Third, in Figure A2c and Table A2c, we restrict our sample to scientists who have never change their party affiliation during our observation period, and the results remain robust.

In Figure A3 and Table A3, we restrict our analysis to different levels of collaboration, namely collaboration with only 2 coauthors (minimum number of coauthors), less than 4 coauthors (bottom 25th percentile), less than 6 coauthors (bottom 50th percentile), and less than 9 coauthors (bottom 75th percentile). We do so because one may be concerned that the collaboration dynamics and frictions in a two-coauthored paper are different from those in scientific publications with larger number of coauthors. The results suggest that scientists are more likely to drop marginal coauthors from the opposite party. In papers with a smaller number of coauthors, the benefits of collaboration seem to outweigh the costs of ideology conflict, and thus the effects become smaller.

V.4.2. Robustness for the scientist level results

In Figure A4 and Table A4, we show that our results remain robust when restricting our sample to scientists with at least one paper during 2017 and 2021 (i.e., after period) to ensure that our result of decline in innovation is not driven by scientists exiting their publication efforts after 2016. In Figure A5 and Table A5, we restrict our sample to scientists who never changed their party affiliation during our observation period, and the results remain robust.

In Figure A6 and Table A6, we repeat our analysis with an alternative definition our treatment variable. In our main results, we define our treatment variable as a dummy indicator which equals one if a scientist's share of coauthors from the opposite party during 2012-2016 is in the top quartile (i.e., greater than 23%) and zero if the scientist is in the bottom quartile (equals to 0). This approach drops scientists whose share of coauthors from opposite party is greater than 0 but smaller than 23 percentiles. As such, we conduct a robustness test where we redefine our treatment variable as scientists who have any coauthors from the opposite party during 2012-2016. The results in Figure A6 and Table A6 are consistent with our main findings. They suggest that after 2016, scientists with least one coauthor from the opposite party (during 2012-2016) produce 7.5% less publications and 10.9% fewer citation-weighted publications, become 0.3% less diverse in research topics, and explore 21.5% fewer new research topics. In Figure A7 and Table A7, we show that our main results remain robust to an OLS model with a continuous treatment intensity measure.

V.5. Additional Analyses

In Figure B1, we plot the raw graph of changes in collaboration for changes in collaboration between different types of scientist pairs. In Figure B2 and Table B1, we repeat our main analysis using each party combination of control scientist pairs as control, once at a time. In Figure B3, we plot the change in distribution of scientists from different parties, over time. In Table B2, we calculate the average number of coauthors per paper by different fields of research. In Figure B4 and Table B3, we repeat the analyses by different fields of research: Engineering, Biological Sciences,

Chemical Sciences, Physical Sciences, Mathematical Sciences, Information and Computing Sciences, Economics, etc. Table B1 shows the average number of coauthors per paper by different fields of research. Finally, in Figure B5 and Table B4, we split the scientist pairs into different subsamples based on whether the scientists have made any donations before 2016. Donating to a political party is correlated with a higher intensity of political identification. We find that our main results are driven by scientist pairs where neither or only one have made donations before 2016 (Column (1) and (2)). Scientist pairs where both made donations before 2016 were less affected by the 2016 presidential election.

VI. Conclusion

Political polarization in the United States continues to be on the rise after reaching historic highs. We take a first step in investigating the impact of this trend for the rate and direction of innovation.

The consequences of polarization for innovation are not clear ex-ante. While, from a normative perspective, we would want scientists' collaboration choices to be immune from ideological influence, sorting by ideology does not necessarily imply negative consequences for knowledge production. Our analysis is motivated by the increasing reliance on collaboration in knowledge production that combined diverse areas of expertise (Jones, 2009, 2010; Agrawal, Goldfarb and Teodoridis, 2016). To the extent polarization is correlated with a taste for certain domains of expertise, it is possible that ideological sorting reduces collaboration beneficial for the rate and direction of innovation.

Testing the impact of polarization on knowledge production is challenging due to data limitations and the endogenous nature of polarization and collaboration choices. We overcome the first obstacle by marching scientists' data with voter registration records, using a probabilistic matching algorithm (Enamorado, Fifield and Imai 2017, 2018) adapted for our needs (Fehder et al., 2024). Our focus on the 2016 election provides an opportunity to evaluate the impact of ideology on knowledge production following an unexpected step increase in polarization. From this perspective, our analysis is mainly directional, as the magnitude of our estimated effects are restricted to the 2016 step increase in polarization.

We find that after the 2016 election, opposing-party scientists are 3.5 percent less likely to collaborate relative to non-opposing party scientists, with the magnitude of the effect steadily increasing in the four years following 2016. The consequence of the decrease in collaboration is a 5.7 percent reduction in publications and 4.7 percent reduction in citation-weighted publications of scientists with a high pre-election opposing party co-authors, compared to scientists with a lower share of opposing-party co-authors. We also find that the impacted scientists exhibit a reduced tendency to publish in new topic areas.

While, directionally, our effects suggest negative consequence of polarization for innovation outcomes, the relatively small size of the effects we measure are somewhat reassuring. The concern is that we capture the impact magnitude of the 2016 step change in polarization which might be different from its cumulative effect over time.

References:

Agrawal, A, Goldfarb, A., Teodoridis, F. 2016. "Understanding the Changing Structure of Scientific Inquiry." American Economic Journal: Applied Economics, 8(1): 100-128.

Alford, J.R., P.K. Hatemi, J.R. Hibbing, N.G. Martin, L.J. Eaves. 2011. "The politics of mate choice." The Journal of Politics 73(2) 362-379.

Arora, A., Belenzon, S., Cioaca, L. C., Sheer, L., & Zhang, H. 2023. "The effect of public science on corporate R&D" (No. w31899). National Bureau of Economic Research.

Azoulay, P., Joshua Graff Zivin and Jialan Wang. 2010. "Superstar Extinction." Quarterly Journal of Economics, 125(2), pp. 549-589.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, J. Van Reenen. 2018. "Who becomes an inventor in America? The importance of exposure to innovation." The Quarterly Journal of Economics 134(2) 647-713.

Boxell, L., Gentzkow, M., Shapiro, J.M. (2022) "Cross-country trends in affective polarization." The Review of Economics and Statistics. 106(2):557-565.

Brown, J.R., R.D. Enos. 2021. "The measurement of partisan sorting for 180 million voters." Nature Human Behaviour 1-11.

Catalini, C., Fons-Rosen, C., & Gaulé, P. 2020. "How do travel costs shape collaboration?." Management Science, 66(8), 3340-3360.

Chen, M. K., & Rohla, R. 2018. "The effect of partisanship and political advertising on close family ties." Science, 360(6392), 1020-1024.

Enamorado, T. 2021. "A primer on probabilistic record linkage." Handbook of Computational Social Science, Volume 2: Data Science, Statistical Modelling, and Machine Learning Methods.

Enamorado, T., B. Fifield, K. Imai. 2017. fastLink: Fast Probabilistic Record Linkage with Missing Data. Version 0.6.(https://github.com/kosukeimai/fastLink).

Enamorado, T., B. Fifield, K. Imai. 2019. "Using a probabilistic model to assist merging of large-scale administrative records." American Political Science Review 113(2) 353-371.

Fehder, D., Teodoridis, F., Raffiee, J., Lu, J. 2024. "The Partisanship of American Inventors", Research Policy, 53 (7).

Fry, C., & MacGarvie, M. 2024. "Author country of origin and attention on open science platforms: Evidence from COVID-19 preprints." Management Science, 70(8), 5426-5444.

Gentzkow, M., Shapiro, J. 2011. "Ideological segregation online and offline." The Quarterly Journal of Economics 126(4): 1799-1839.

Herzog, C., Hook, D. & Konkiel, S. "Dimensions: bringing down barriers between scientometricians and data." Quant. Sci. Stud. 1, 387–395 (2020)

Iyengar, S., Sood, G., & Lelkes, Y. 2012. "Affect, not ideology: A social identity perspective on polarization." Public Opinion Quarterly, 76(3), 405-431.

Jaravel, X., Petkova, N. and Bell, A., 2018. "Team-specific capital and innovation." American Economic Review, 108(4-5), 1034-1073.

Jones, B. 2009. "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" Review of Economic Studies, 76(1): 253-281.

Jones, B. 2010. "As Science Evolves, How Can Science Policy?" NBER Innovation Policy and the Economy, 11: 103-131.

Kaplan, E., Spenkuch, J. L., & Sullivan, R. 2022. "Partisan spatial sorting in the United States: A theoretical and empirical overview." Journal of Public Economics, 211, 104668.

Katz, J. 2016. Who Will be President? The New York Times. https://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html

Lerner, Josh and Scott Stern, eds. 2012. "The Rate and Direction of Inventive Activity" Revisited, Chicago, IL: University of Chicago Press.

McConnell, C., Margalit, Y., Malhotra, N. and Levendusky, M., 2018. "The economic consequences of partisanship in a polarized era." American Journal of Political Science, 62(1), pp.5-18.

Mosleh, M., Martel, C., Eckles, D., & Rand, D. G. 2021. "Shared partisanship dramatically increases social tie formation in a Twitter field experiment." Proceedings of the National Academy of Sciences, 118(7), e2022761118.

Nelson, R. R. 1982. "The Role of Knowledge in R&D Efficiency." Quarterly Journal of Economics, 97(3): 453-470.

Raffiee, J., Teodoridis, F., Fehder, D. 2023 "Partisan Patent Examiners? Exploring the Link Between the Political Ideology of Patent Examiners and Patent Office Outcomes", Research Policy 52(9).

Romer, P. M. 1990 "Endogenous Technological Change." Journal of Political Economy, 98(5):71-102.

Solow, R. M. 1956 "A Contribution to the Theory of Economic Growth." Quarterly Journal of Economics, 70: 65-94.

Tyler, M. and Iyengar, S., 2024. "Testing the robustness of the ANES feeling thermometer indicators of affective polarization." American Political Science Review, 118(3), 1570-1576.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. 2013 "Atypical combinations and scientific impact" Science, 342(6157), 468-472.

Wuchty, S., Jones, B., Uzzi, B. 2007. "The Increasing Dominance of Teams in the Production of Knowledge." Science, 316: 1036-1039.

Yin, Y., Dong, Y., Wang, K., Wang, D., & Jones, B. F. 2022. "Public use and public funding of science." Nature Human Behaviour, 6(10), 1344-1350.

Figure 1. ANES data on Affective Polarization in the US

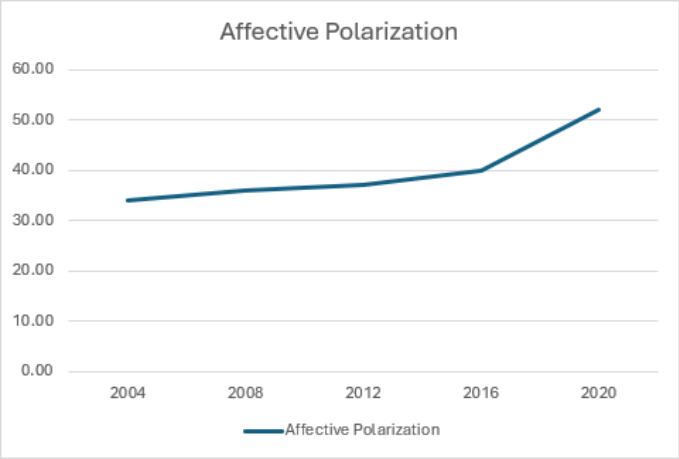
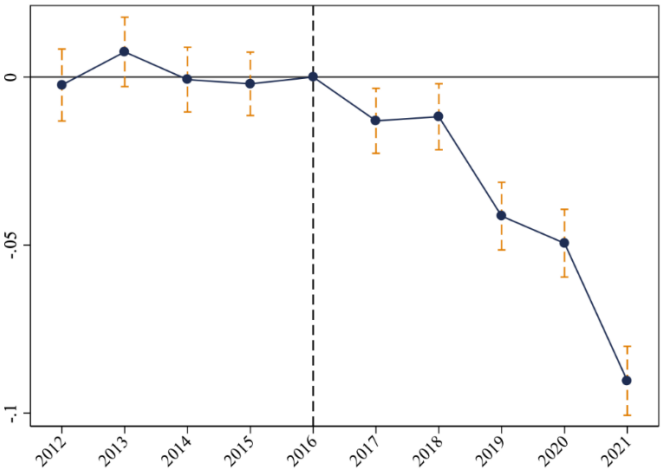
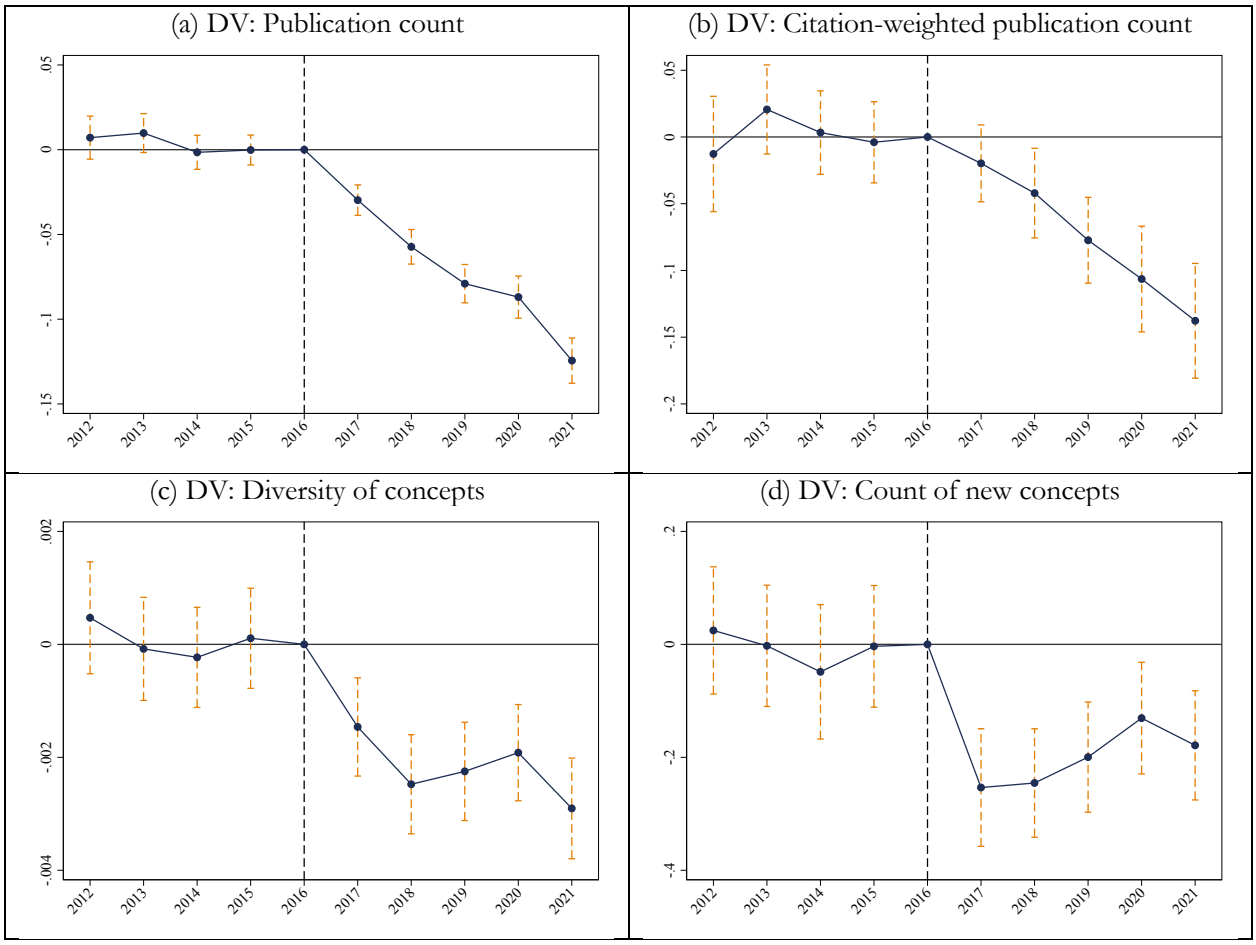


Figure 2. Democrat and Republican scientists less likely to collaborate after 2016 (Table 2)



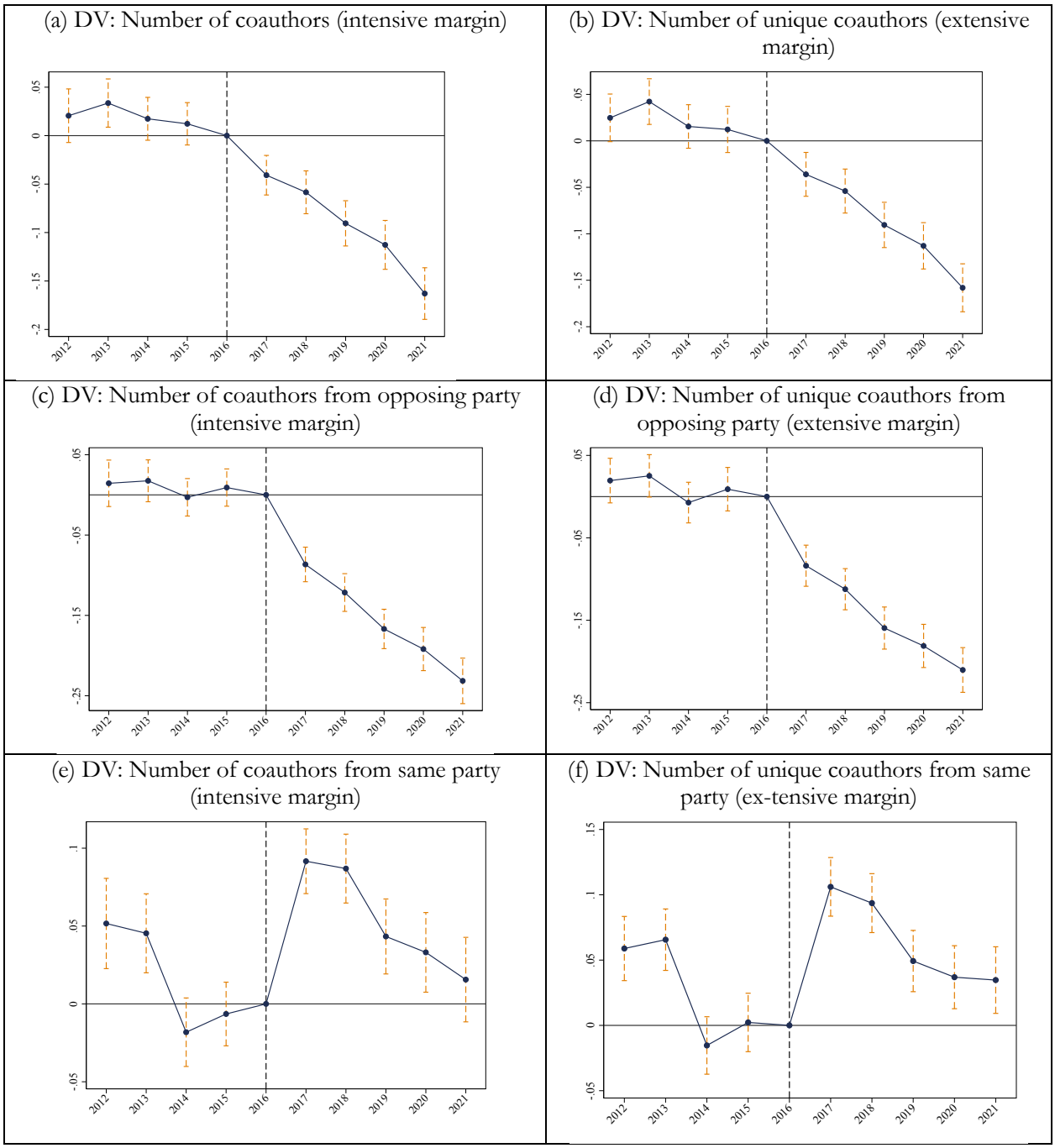
*Note*s. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist pair*×*Year* and thus describes the relative difference in likelihood of collaboration between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure 3. Scientists with a larger share of collaborators from opposite party become less productive and less novel (Table 3 Columns 1, 2, 3 and 4)



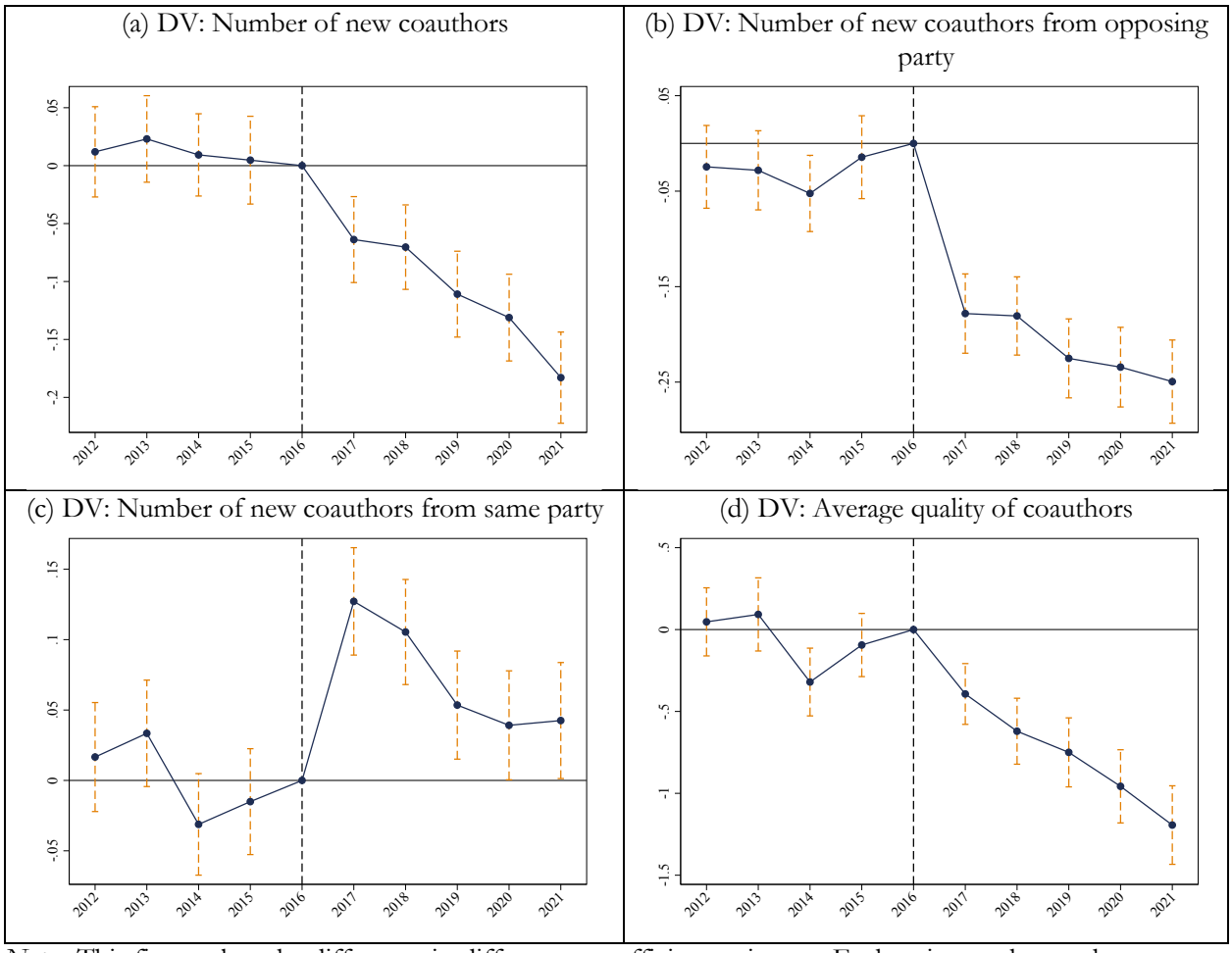
Notes. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a) publication count (b) citation-weighted publication count, (c) diversity of concepts, and (d) count of new concepts between treated and control scientists in that year. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure 4. Decline in collaboration is driven only by reduced collaboration with opposite party (Table 4).



*Note*s. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a & b) number of coauthors (c & d) number of coauthors from opposing party, (e & f) number of coauthors from same party. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure 5. Treated scientists have friction in finding new coauthors to replace their previous coauthors from opposing party. (Table 5)



*Note*s. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a) number of new coauthors (b) number of new coauthors from opposing party, (c) number of new coauthors from same party, and (d) average quality of coauthors in that year. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Table 1a. Descriptive statistics at the scientist pair-year level

	Before 2016 US Presidential Election					**After** 2016 US Presidential Election				
Variable	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max	Obs
	Treated scientist pairs (i.e., one Democrat and one Republican scientist)									
Any collaboration	0.202	0.401	0	1	3,079,516	0.200	0.400	0	1	2,830,180
Number of co-publications	0.296	0.877	0	111	3,079,516	0.304	0.965	0	101	2,830,180
Age	16.328	8.130	0	46	3,079,516	20.848	8.377	1	51	2,830,180
	Control scientist pairs (i.e., other combination of party affiliations)									
Any collaboration	0.197	0.398	0	1	14,218,639	0.202	0.401	0	1	13,304,510
Number of co-publications	0.285	0.851	0	113	14,218,639	0.303	0.971	0	225	13,304,510
Age	15.269	7.959	0	46	14,218,639	19.693	8.233	1	51	13,304,510

Table 1b. Descriptive statistics at the scientist pair level – split control scientist pairs into each party affiliation type

	Before 2016 US Presidential Election					**After** 2016 US Presidential Election				
Variable	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max	Obs
	Both Democrat scientists									
Any collaboration	0.196	0.397	0	1	4,488,580	0.199	0.399	0	1	4,213,133
Number of co-publications	0.280	0.839	0	92	4,488,580	0.295	0.913	0	145	4,213,133
Age	16.490	8.263	0	46	4,488,580	20.917	8.515	1	51	4,213,133
	Both Republican scientists									
Any collaboration	0.221	0.415	0	1	846,900	0.211	0.408	0	1	756,088

Number of co-publications	0.351	1.056	0	113	846,900	0.346	1.249	0	225	756,088
Age	15.953	7.903	0	46	846,900	20.564	8.133	1	51	756,088
	Both Independent Scientist									
Any collaboration	0.198	0.399	0	1	1,649,447	0.208	0.406	0	1	1,558,349
Number of co-publications	0.286	0.830	0	87	1,649,447	0.315	0.975	0	225	1,558,349
Age	13.259	7.359	0	46	1,649,447	17.615	7.683	1	51	1,558,349
	One Democrat and one Independent scientist									
Any collaboration	0.192	0.394	0	1	4,896,246	0.199	0.399	0	1	4,617,278
Number of co-publications	0.274	0.836	0	101	4,896,246	0.295	0.919	0	146	4,617,278
Age	14.971	7.813	0	46	4,896,246	19.371	8.092	1	51	4,617,278
	One Republican and one Independent scientist									
Any collaboration	0.201	0.401	0	1	1,944,630	0.204	0.403	0	1	1,794,289
Number of co-publications	0.293	0.836	0	53	1,944,630	0.311	1.096	0	225	1,794,289
Age	14.606	7.649	0	46	1,944,630	19.078	7.936	1	51	1,794,289
	One Democrat scientist and one scientist from other parties									
Any collaboration	0.198	0.398	0	1	204,633	0.201	0.401	0	1	190,481
Number of co-publications	0.288	0.854	0	42	204,633	0.298	0.888	0	45	190,481
Age	15.714	7.927	0	46	204,633	20.207	8.210	1	51	190,481
	One Republican scientist and one scientist from other parties									
Any collaboration	0.208	0.406	0	1	73,022	0.203	0.402	0	1	66,239

Number of co-publications	0.302	0.833	0	36	73,022	0.306	0.932	0	31	66,239
Age	15.666	7.878	0	46	73,022	20.223	8.176	1	51	66,239
	One Independent scientist and one scientist from other parties									
Any collaboration between	0.186	0.389	0	1	104,127	0.201	0.400	0	1	98,825
Number of co-publications	0.266	0.813	0	37	104,127	0.304	0.966	0	39	98,825
Age	14.197	7.600	0	44	104,127	18.573	7.887	1	48.5	98,825
	Both scientists from other parties (i.e., both non-Democrat, non-Republican, and non-Independent)									
Any collaboration	0.248	0.432	0	1	11,054	0.227	0.419	0	1	9,828
Number of co-publications	0.384	0.967	0	30	11,054	0.351	0.878	0	17	9,828
Age	14.858	7.741	0	45.5	11,054	19.356	8.036	1	50.5	9,828

Table 1c. Descriptive statistics at the scientist-year level

	Before 2016 US Presidential Election					**After** 2016 US Presidential Election				
Variable	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max	Obs
	Treated scientist (share of coauthors during 2012-2016 from opposite party in the top quartile i.e., 23%)									
Publication count	2.844	4.761	0	150	352,984	3.481	5.999	0	275	285,479
Citation-weighted publication count	62.912	205.481	0	15207	352,984	59.191	204.341	0	15968	285,479
Diversity of concepts	0.975	0.065	0	1	270,799	0.98	0.056	0	1	221,784
Count of new concepts	2.348	15.989	0	1,935	243,874	1.855	9.692	0	1,243	253,114
Share of coauthors during 2012-2016 from opposite party	0.467	0.205	0.231	1	352,984	0.458	0.199	0.231	1	285,479
Count of coauthors	9.044	24.505	0	1,174	352,984	12.104	31.278	0	1,226	285,479

Count of unique coauthors	6.084	13.782	0	723	352,984	7.944	15.207	0	367	285,479
Count of new coauthors	3.981	10.879	0	673	352,984	4.505	9.707	0	299	285,479
Average quality of coauthors	13.683	19.125	0	1,124	352,984	14.291	18.198	0	1,133	285,479
Age	12.707	11.231	0	46	352,984	17.202	11.620	1	51	285,479
	Control scientist (share of coauthors during 2012-2016 from opposite party in the bottom quartile i.e., 0%)									
Publication count	2.12	3.55	0	199	723,591	2.703	4.680	0.00	356	607,394
Citation-weighted publication count	48.48	188.486	0	24,663	723,591	47.051	186.655	0	49,752	607,394
Diversity of concepts	0.971	0.072	0	1	523,241	0.978	0.059	0	1	450,703
Count of new concepts	2.243	17.982	0.00	6,531	441,738	2.12	14.211	0.00	3,585	505,251
Share of coauthors during 2012-2016 from opposite party	0	0	0	0	723,591	0	0	0	0	607,394
Count of coauthors	5.53	17.865	0	1,346	723,591	7.939	22.346	0	1,575	607,394
Count of unique coauthors	3.911	10.696	0	642	723,591	5.479	11.791	0	470	607,394
Count of new coauthors	2.723	8.972	0.00	673	723,591	3.381	8.308	0	763	607,394
Average quality of coauthors	12.739	21.297	0	2,122	723,591	13.716	20.678	0	2,426	607,394
Age	10.332	10.309	0	46	723,591	14.209	10.726	1	51	607,394

Table 2. Democrat and Republican scientists less likely to collaborate after 2016 (See Figure 2 for the equivalent graph)

DV: Any collaboration	(1)	(2)
	Coefficient	Odds ratio
Treated scientist pair × After 2016 US Presidential Election	**-0.039*** (0.002)**	**0.962*** (0.002)**
Log likelihood (LL)	-10,296,610	-10,296,610
Number of pairs	3,800,343	3,800,343
Number of observations	33,169,607	33,169,607
Scientist pair FE	Yes	Yes
Year FE	Yes	Yes

Notes. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Logit model with estimations presented as coefficient (SE) in Column (1) and odds ratio in Column (2).

*p < 0.10; **p < 0.05; ***p < 0.01.

Table 3. Scientists with a larger share of collaborators from opposite party become less productive and novel after 2016 (See Figure 3 for the equivalent graph)

	(1)	(2)	(3)	(4)
	DV: Publication count	DV: Citation-weighted publication count	DV: Diversity of concepts	DV: Count of new concepts
Treated Scientist × After 2016 US Presidential Election	**-0.077*** (0.005)**	**-0.061*** (0.010)**	**-0.002*** (0.000)**	**-0.194*** (0.026)**
LL/R^2	-2,705,231	-46,936,439	0.002	-4,143,041
Observations	1,960,047	1,960,047	1,466,527	1,354,829
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). Column (1), (2) and (4) present results from Poisson model. Column (3) presents result from OLS model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE).

*p < 0.10; **p < 0.05; ***p < 0.01.

Table 4. Decline in collaboration is driven only by reduced collaboration with opposite party. (See Figure 4 for the equivalent graph)

	(1)	(2)	(3)	(4)	(5)	(6)
	DV: Number of coauthors (intensive margin)	DV: Number of **unique** coauthors (extensive margin)	DV: Number of coauthors from opposing party (intensive margin)	DV: Number of unique coauthors from opposing party (extensive margin)	DV: Number of coauthors from same party (intensive margin)	DV: Number of unique coauthors from same party (extensive margin)
Treated Scientist × After 2016 US Presidential Election	**-0.107*** (0.008)**	**-0.106*** (0.006)**	**-0.165*** (0.008)**	**-0.156*** (0.006)**	**0.046*** (0.008)**	**0.049*** (0.006)**
LL	-6,822,496	-5,082,299	-997,255	-834,071	-3,178,114	-2,447,631
Observations	1,960,047	1,960,047	635,825	635,825	1,765,602	1,765,602
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). Different columns correspond to different dependent variables. Column (1)-(6) present results from Poisson model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE).

*p < 0.10; **p < 0.05; ***p < 0.01.

Table 5. Treated scientists have friction in finding new coauthors to replace their previous coauthors from opposing party (See Figure 5 for equivalent graph)

	(1)	(2)	(3)	(4)
	DV: Number of new coauthors	DV: Number of new coauthors from opposing party	DV: Number of new coauthors from same party	DV: Average quality of coauthors
Treated Scientist × After 2016 US Presidential Election	**-0.118*** (0.008)**	**-0.190*** (0.009)**	**0.067*** (0.008)**	**-0.678*** (0.064)**
LL/R^2	-4,973,172	-3,934,016	-1,951,415	0.002
Observations	1,927,432	1,903,854	1,705,325	1,969,448
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). Different columns correspond to different dependent variables. Column (1)-(3) present results from Poisson model. Column (4) presents result from OLS model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE).

*p < 0.10; **p < 0.05; ***p < 0.01.

Appendix A

Robustness checks

A.1. Scientist pair level (decline in collaboration), robust to:

Figure & Table A1. Poisson model with count of co-publication, xtreg with collab and count of co-publication, including a random sample of non-collaborating teams.

Figure & Table A2a. Restrict to scientists with at least one publication after 2016.

Figure & Table A2b. Restrict to scientist pairs with at least one Democrat or one Republican scientists.

Figure & Table A2c. Restrict to scientist pairs who have not never changed affiliation during 2014-2021.

Figure & Table A3. Restrict to papers with less than 2, 4, 6, 9 coauthors.

A.2. Author level (consequence for innovation outcome), robust to:

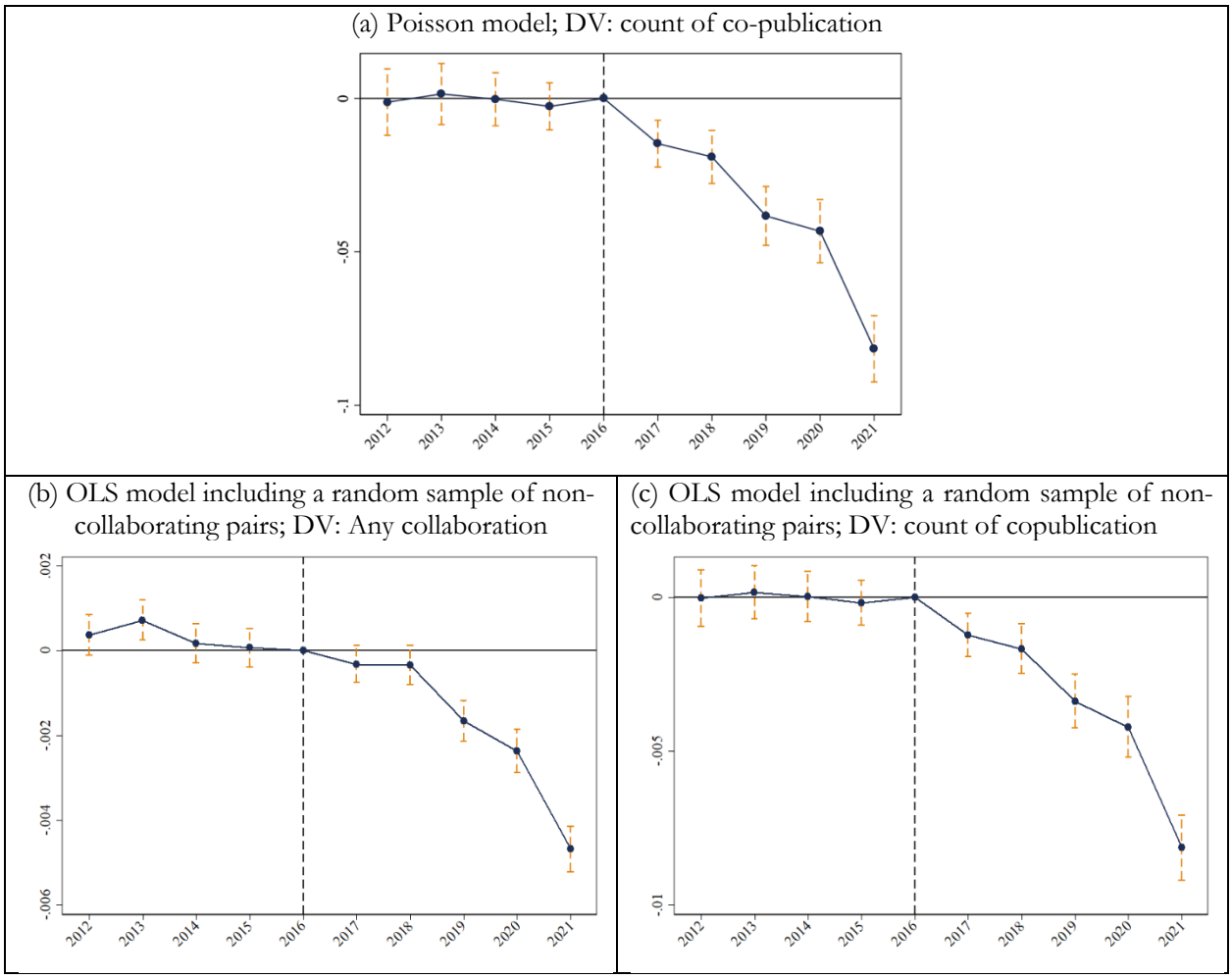
Figure & Table A4. Restrict to scientists with at least one publication after 2016.

Figure & Table A5. Restrict to scientist pairs who have not never changed affiliation during 2014-2021.

Figure & Table A6. Poisson model with treated scientists defined as those with any coauthors (rather than top quartile in share) from opposing party before 2016.

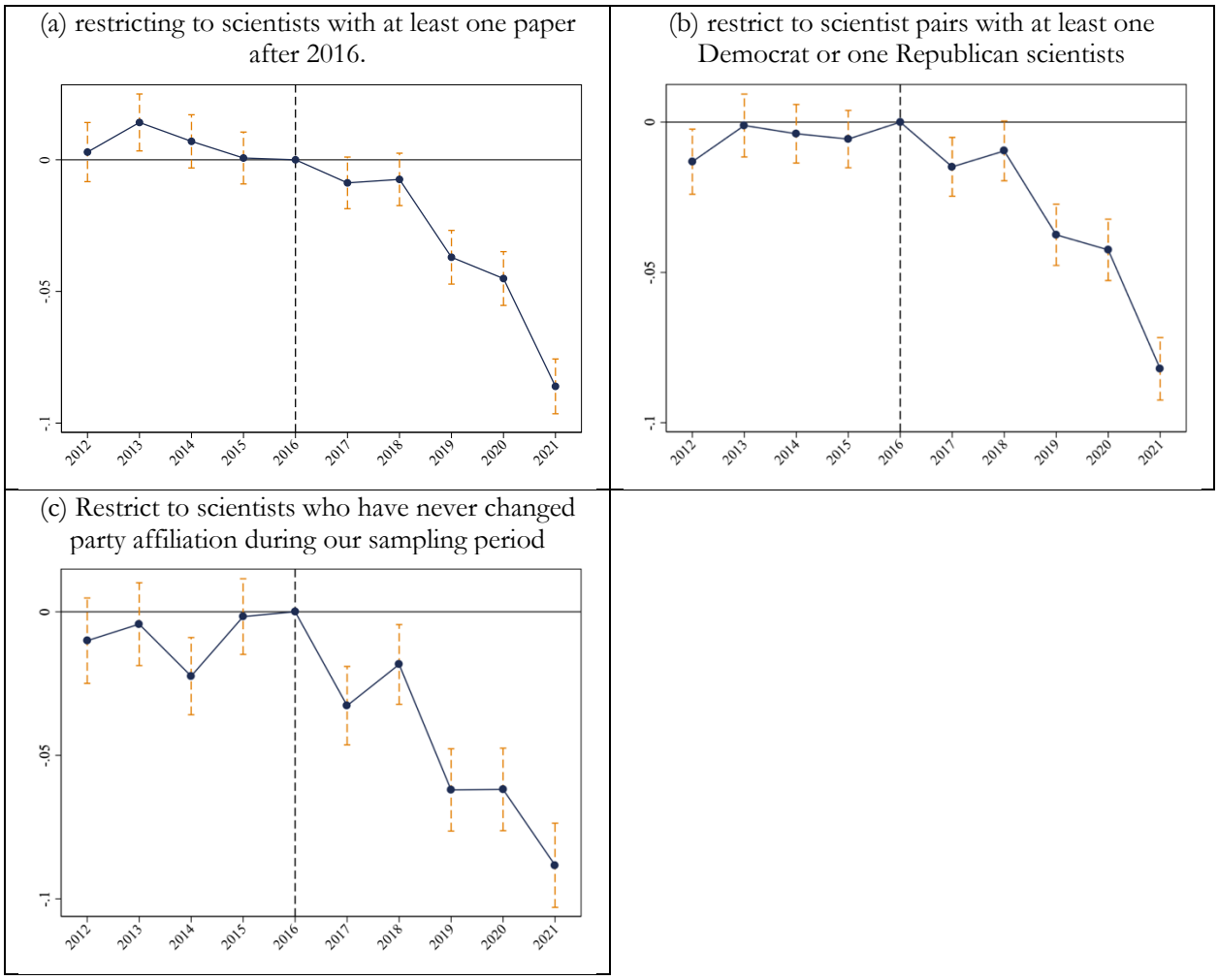
Figure & Table A7. OLS model with continuous treatment.

Figure A1. Figure 2 robust to alternative models and using a count of co-publication as DV.



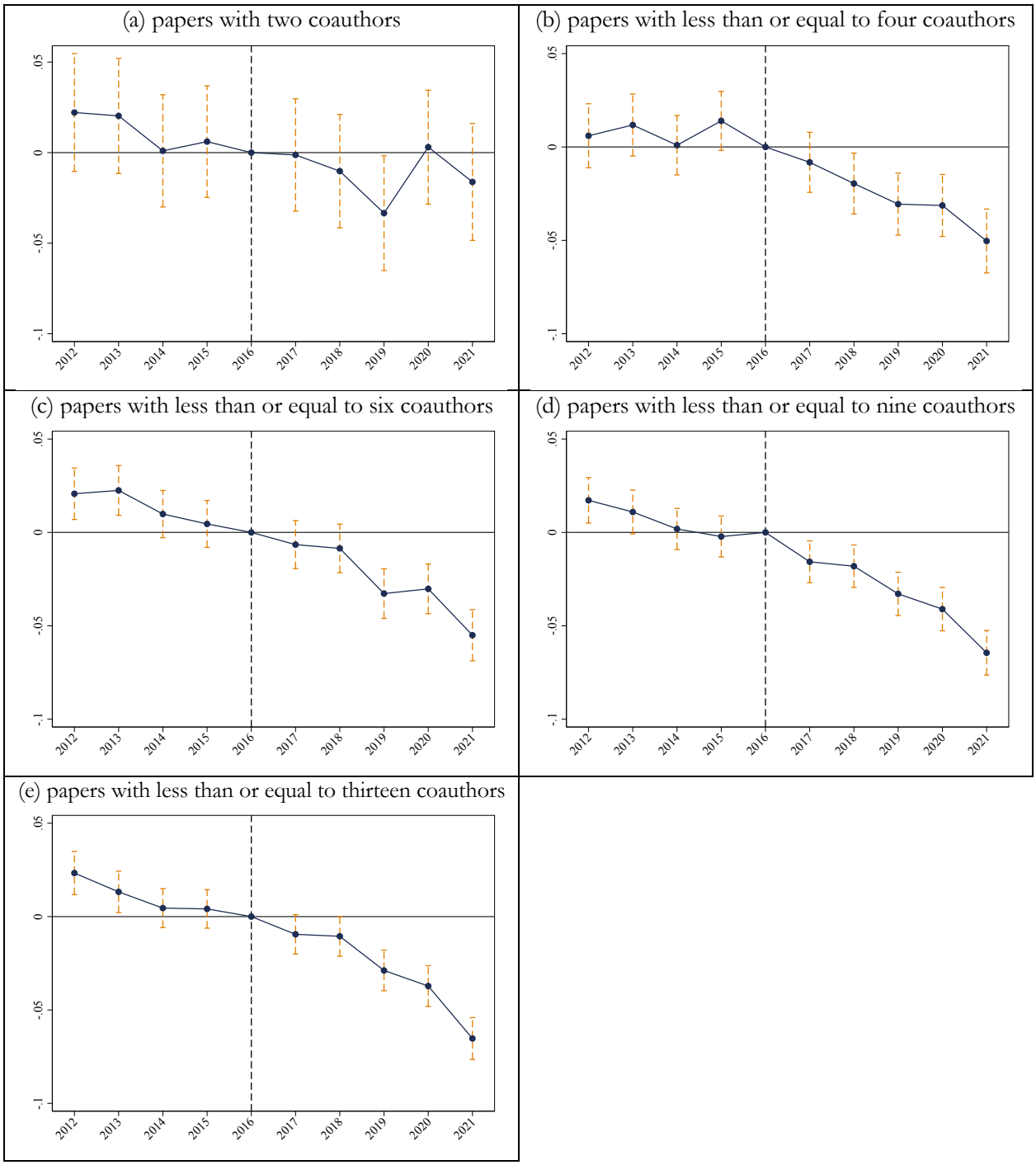
*Note*s. This figure plots the difference-in-differences coefficient estimates. It replicates Figure 2 but runs on (a) Poisson model, and (b & c) OLS model, where we replace the original dependent variable with a count of copublications between a research pair in a year. Each point on the graph represents the coefficient value on the covariate. *Treated scientist pair*×*Year* and thus describes the relative difference in (a & b) likelihood of collaboration and (c) number of collaborations between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A2. Figure 2 robust to alternative sample construction approaches



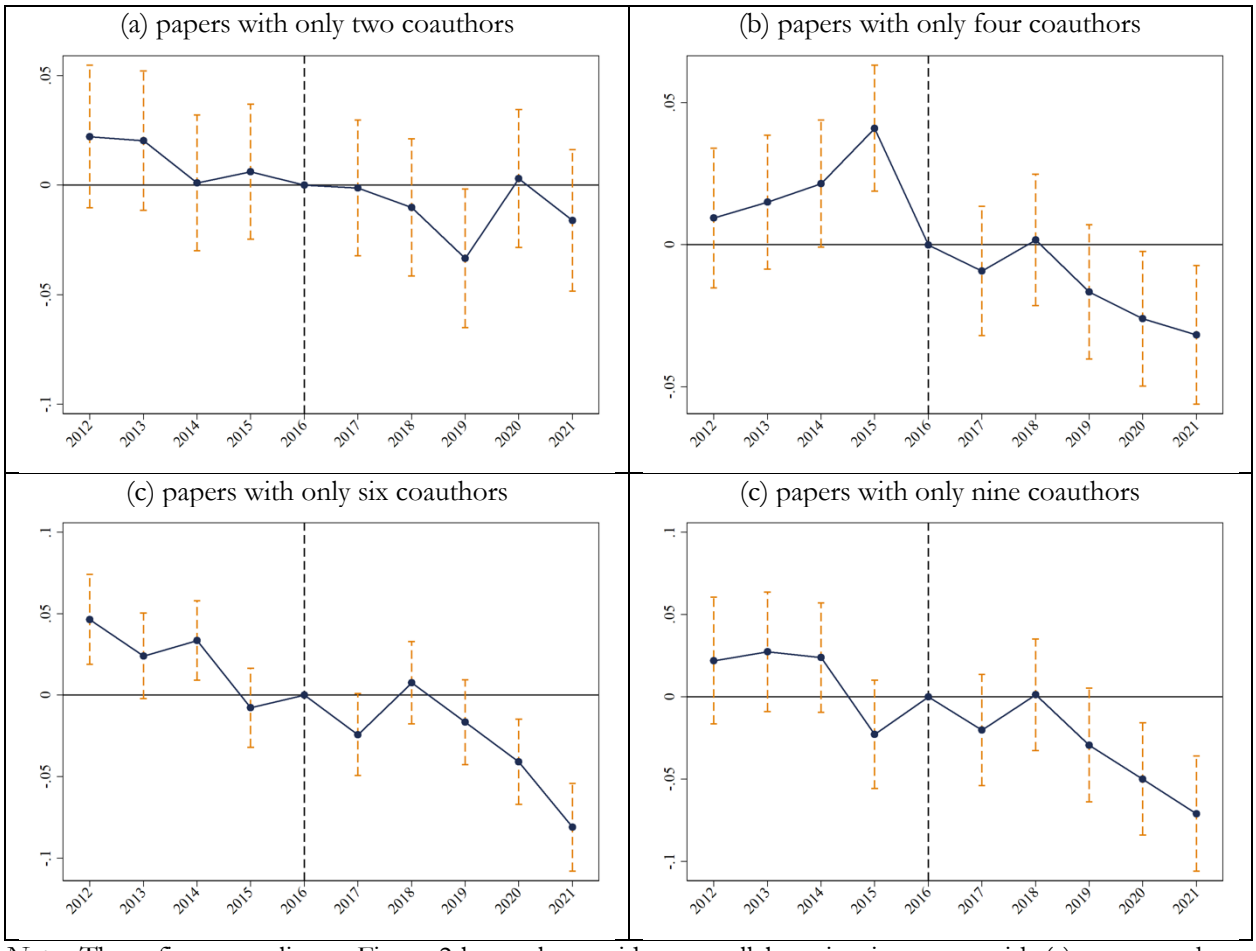
*Note*s. These figures replicates Figure 2 but employs various alternative sample construction approaches. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist pair*×*Year* and thus describes the relative difference in likelihood of collaboration between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A3a. Only considers as collaboration in papers with (a) two coauthors, (b) less than or equal to four coauthors, (c) less than or equal to six coauthors, and (d) less than or equal to nine coauthors



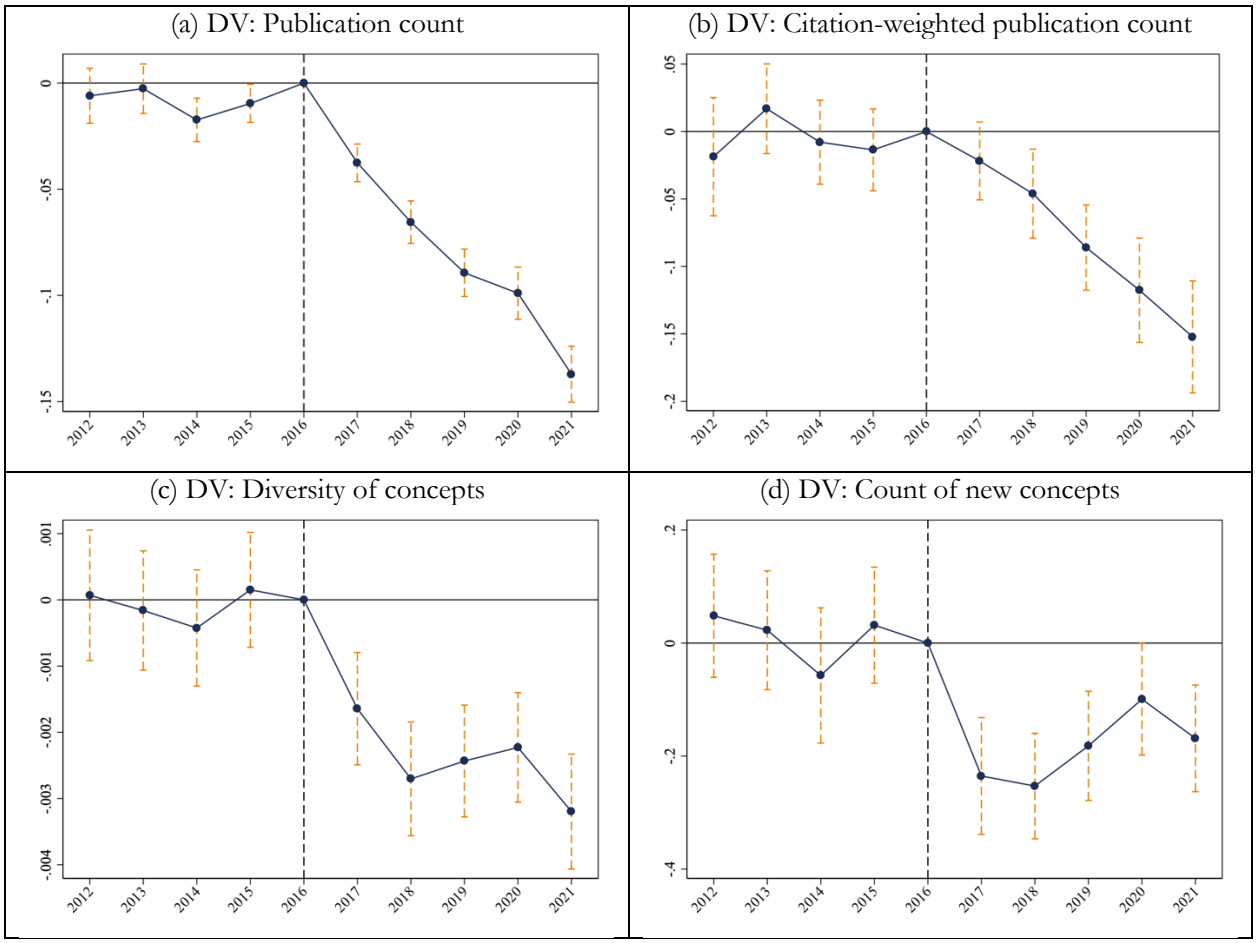
*Note*s. These figures replicates Figure 2 but only considers as collaboration in papers with (a) two coauthors, (b) less than or equal to four coauthors, (c) less than or equal to six coauthors, and (d) less than or equal to nine coauthors, (e) less than or equal to thirteen coauthors. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist pair*×*Year* and thus describes the relative difference in likelihood of collaboration between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A3b. Only considers as collaboration in papers with (a) only two coauthors, (b) only four coauthors, (c) only six coauthors, and (d) only nine coauthors



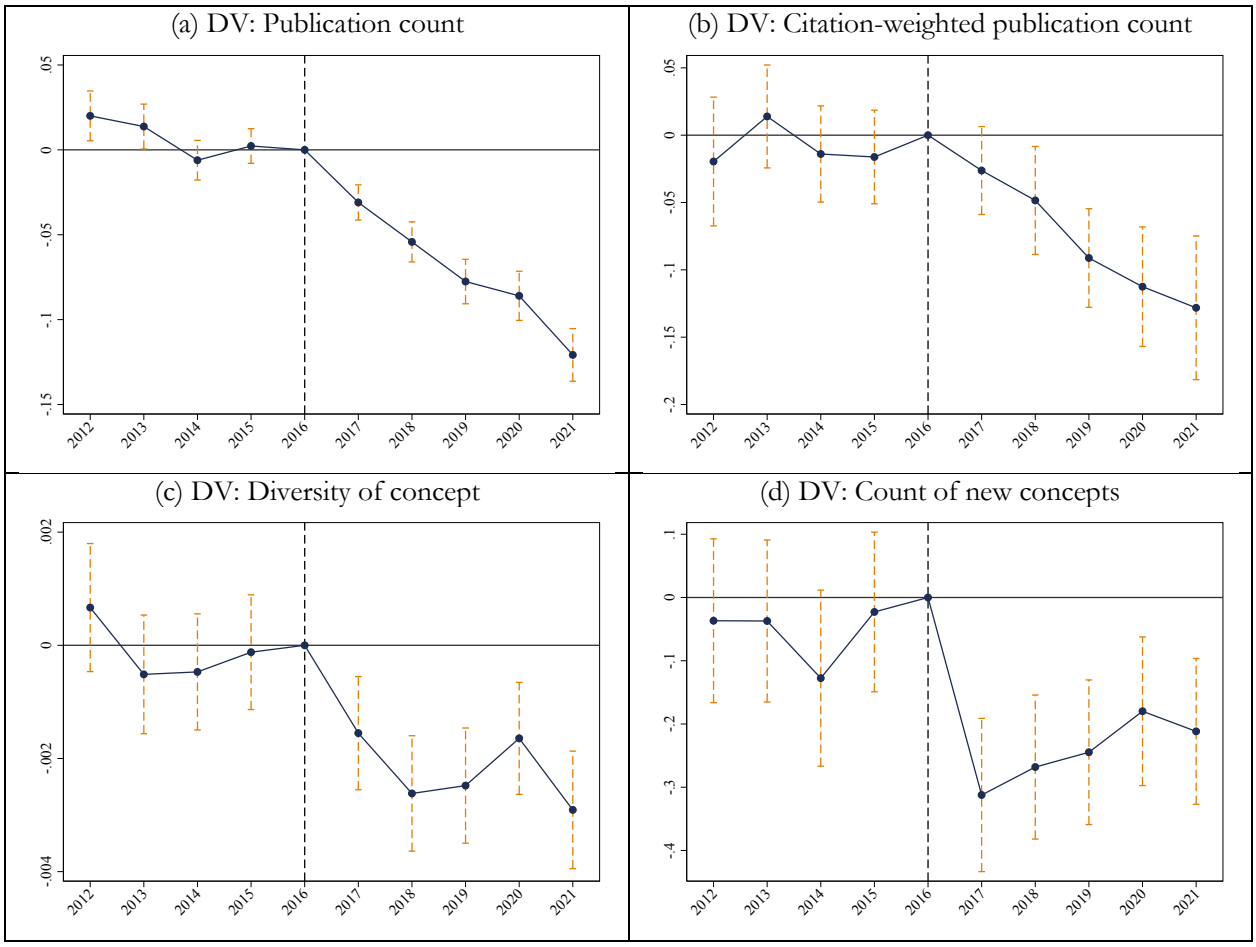
*Note*s. These figures replicates Figure 2 but only considers as collaboration in papers with (a) two coauthors, (b) less than or equal to four coauthors, (c) less than or equal to six coauthors, and (d) less than or equal to nine coauthors. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist pair*×*Year* and thus describes the relative difference in likelihood of collaboration between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A4. Figure 2 robust when restricting to scientists with at least one paper after 2016.



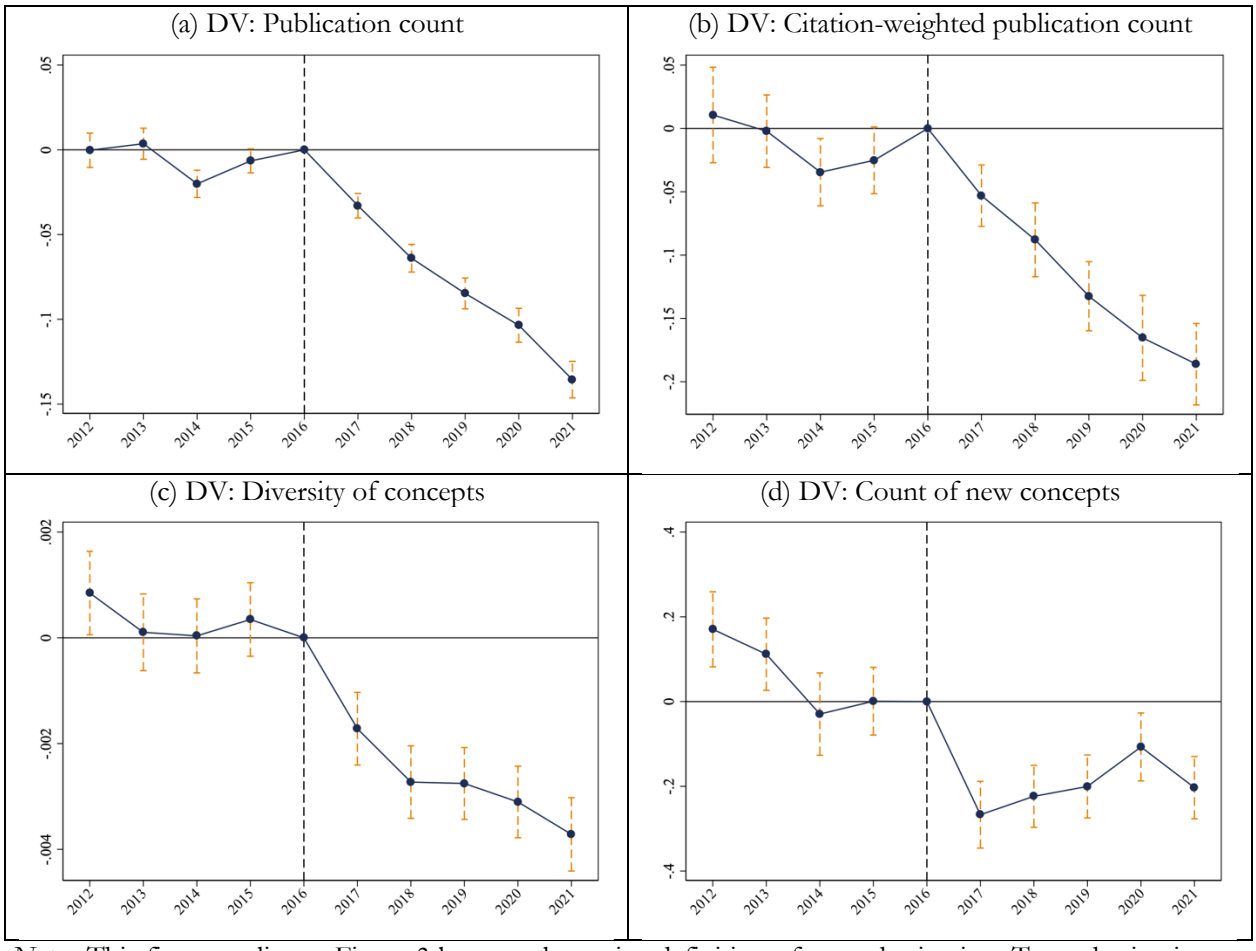
*Note*s. This figure replicates Figure 2 but restrict the scientists to those with at least one paper after 2016. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a) publication count (b) citation-weighted publication count, (c) diversity of concepts, and (d) count of new concepts between treated and control scientists in that year. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A5. Figure 2 robust when restricting to scientists who have never changed party affiliation during our sampling period.



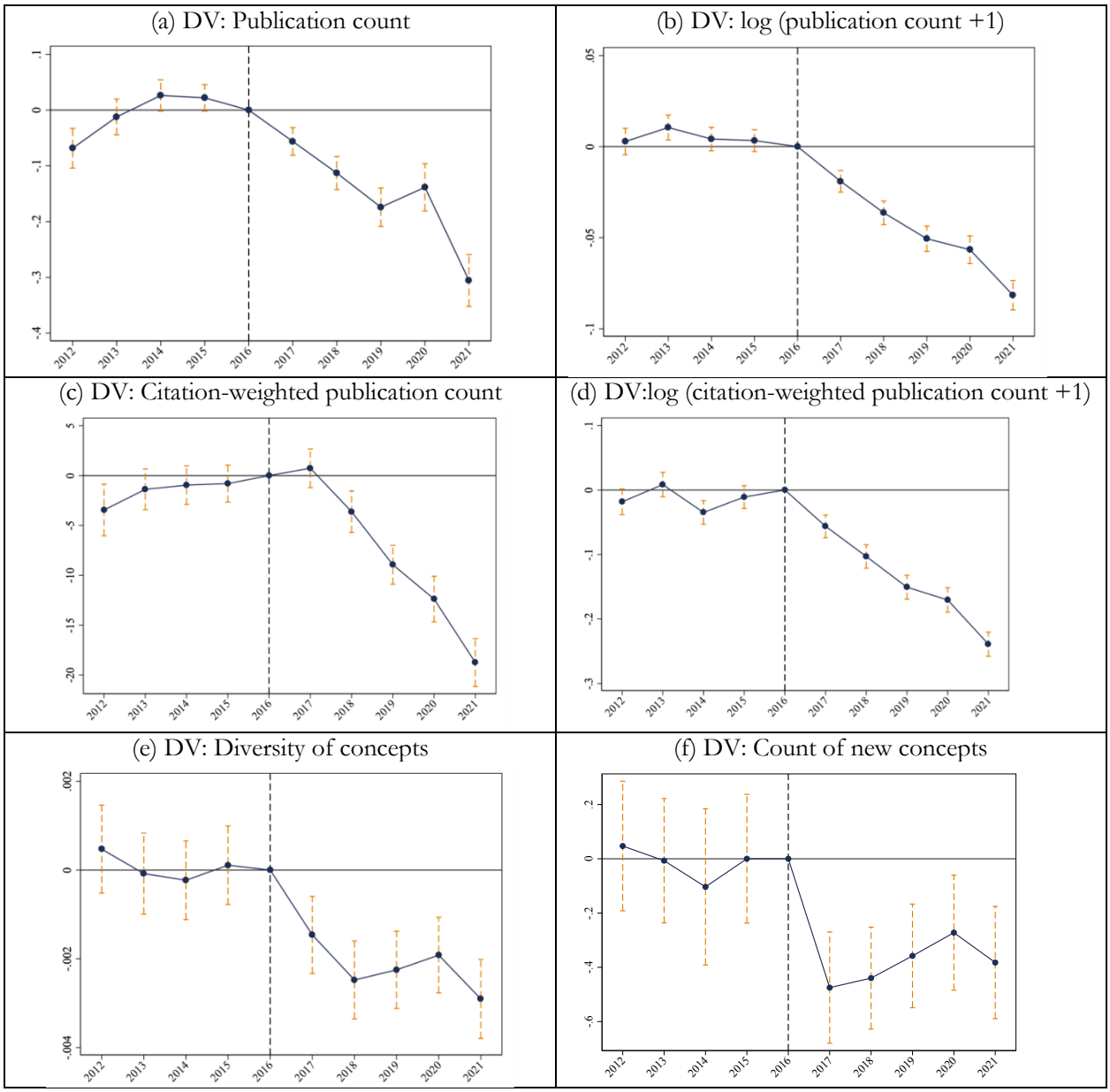
*Note*s. This figure replicates Figure 2 but restrict the scientists to those who have never changed party affiliation during our sampling period. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a) publication count (b) citation-weighted publication count, (c) diversity of concepts, and (d) count of new concepts between treated and control scientists in that year. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A6. Figure 3 robust when using Poisson model with treated scientists defined as those with **any coauthors** (rather than top quartile in share) from opposing party before 2016.



*Note*s. This figure replicates Figure 3 but use alternative definition of treated scientists. Treated scientists are defined as scientists with **any** coauthors from opposite party during 2012-2016 (rather than share of coauthors from opposite party in the top quartile, as in Figure 3), and control scientists are defined as scientists who did not have any coauthors from opposite party during 2012-2016. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a) publication count (b) citation-weighted publication count, (c) diversity of concepts, and (d) count of new concepts between treated and control scientists in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure A7. Figure 3 robust when using OLS model with treated scientists defined as a continuous treatment measure: share of coauthors from opposite party (during 2012-2016)



*Note*s. This figure replicates Figure 3 using an OLS model. Each point on the graph represents the coefficient value on the covariate *Treated scientist*×*Year* and thus describes the relative difference in (a) publication count, (b) publication count (logged), (c) citation-weighted publication count, (d) citation-weighted publication count (logged), (e) diversity of concepts, and (f) count of new concepts between treated and control scientists in that year. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Table A1. Table 2 robust to alternative models and count of co-publication as DV.

	Poisson model	OLS model	
	(1)	(2)	(3)
	DV: Number of copublications	DV: Any collaboration	DV: Number of copublications
Treated scientist pair × After 2016 US Presidential Election	**-0.037*** (0.003)**	**-0.002*** (0.000)**	**-0.003*** (0.000)**
LL/R^2	-14,382,476	0.001	0.001
Number of pairs	3,846,770	12,435,807	12,435,807
Number of observations	33,391,362	105,990,037	105,990,037
Age control	No	Yes	Yes
Scientist pair FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Notes. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Column (1) presents result from Poisson model, and Columns (2) and (3) present results from OLS model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE).*p < 0.10; **p < 0.05; ***p < 0.01.

Table A2. Table 2 robust when restricting to alternative sample construction approaches.

DV: Any collaboration	(1) Subsample: scientists with at least one paper after 2016	(2) Subsample: scientist pairs with at least one Democrat or one Republican scientists	(3) Subsample: scientists who have never changed party affiliation during our sampling period
Treated scientist pair × After 2016 US Presidential Election	**-0.039*** (0.002)**	**-0.031*** (0.002)**	**-0.043*** (0.003)**
LL	-9,915,485	-9,238,704	-5,771,563
Number of pairs	3,529,456	3,402,610	2,139,939
Number of observations	32,079,568	29,768,199	18,603,591
Scientist pair FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Notes. This table replicates Table 2 but employs various alternative sample construction approaches. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Logit model with estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table A3a. Restrict to papers with less than 2, 4, 6, 9 coauthors.

DV: Any collaboration	(1)	(2)	(3)	(4)	(5)
	two authors	Less than or equal to four authors	Less than or equal to six authors	Less than or equal to nine authors	Less than or equal to thirteen authors
Treated scientist pair × After 2016 US Presidential Election	**-0.020*** (0.008)**	**-0.033*** (0.004)**	**-0.035*** (0.003)**	**-0.037*** (0.003)**	**-0.036*** (0.003)**
LL	-1,166,370	-4,090,893	-6,260,905	-8,048,648	-9,026,144
Number of pairs	450,771	1,546,654	2,337,194	2,972,831	3,313,992
Number of observations	3,905,492	13,291,678	20,129,157	25,729,545	28,804,858
Scientist pair FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes. This table replicates Table 2 but only considers as collaboration in papers with (1) 2 coauthors, (2) less than or equal to 4 coauthors, (3) less than or equal to six coauthors, and (4) less than or equal to nine coauthors. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Synthetic difference-in-differences model. Estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table A3b. Restrict to papers with 2, 4, 6, 9 coauthors.

DV: Any collaboration	(1)	(2)	(3)	(4)
	Two authors	Four authors	Six authors	Nine authors
Treated scientist pair × After 2016 US Presidential Election	**-0.020*** (0.008)**	**-0.033*** (0.006)**	**-0.045*** (0.006)**	**-0.040*** (0.008)**
LL	-1,166,370	-2,021,373	-1,599,050	-850,279
Number of pairs	450,771	807,504	646,364	347,235
Number of observations	3,905,492	6,980,040	5,678,236	3,138,006
Scientist pair FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. This table replicates Table 2 but only considers as collaboration in papers with (1) 2 coauthors, (2) less than or equal to 4 coauthors, (3) less than or equal to six coauthors, and (4) less than or equal to nine coauthors. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Synthetic difference-in-differences model. Estimations presented as coefficient (SE).*p < 0.10; **p < 0.05; ***p < 0.01.

Table A4. Table 3 robust when restricting to scientists with at least one paper after 2016.

	(1) DV: Publication count	(2) DV: Citation-weighted publication count	(3) DV: Diversity of concepts	(4) DV: Count of new concepts
Treated Scientist × After 2016 US Presidential Election	**-0.077*** (0.004)**	**-0.061*** (0.010)**	**-0.002*** (0.000)**	**-0.194*** (0.025)**
LL/R^2	-2,658,517	-46,526,586	0.002	-3964978
Observations	1,820,272	1,820,272	1,380,335	1,327,833
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. This table replicates Table 3 but restricts the scientists to those with at least one paper after 2016. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). Column (1), (2) and (4) present results from Poisson model. Column (3) presents result from OLS model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table A5. Table 3 robust when restricting to scientists who have never changed party affiliation during our sampling period.

	(1) DV: Publication count	(2) DV: Citation-weighted publication count	(3) DV: Diversity of concepts	(4) DV: Count of new concepts
Treated Scientist × After 2016 US Presidential Election	**-0.078*** (0.005)**	**-0.059*** (0.011)**	**-0.002*** (0.000)**	**-0.199*** (0.030)**
LL/R^2	-2,033,805	-35,306,011	0.001	-3,228,190
Observations	1,492,578	1,492,578	1,110,361	1,040,013
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. This table replicates Table 3 but restricts the scientists to those who have never changed party affiliation during our sampling period. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the top quartile (i.e., share coauthors from opposite party greater than 23%), and control scientists are defined as scientists with share of coauthors from opposite party (during 2012-2016) in the bottom quartile (i.e., equals to 0). Column (1), (2) and (4) present results from Poisson model. Column (3) presents result from OLS model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table A6. Table 3 robust when using Poisson model with treated scientists defined as those with **any coauthors** (rather than top quartile in share) from opposing party before 2016.

	(1) DV: Publication count	(2) DV: Citation-weighted publication count	(3) DV: Diversity of concepts	(4) DV: Count of new concepts
Treated Scientist × After 2016 US Presidential Election	**-0.078*** (0.004)**	**-0.095*** (0.008)**	**-0.003*** (0.000)**	**-0.242*** (0.021)**
LL/R^2	-3,764,563	-70,874,794	0.001	-5,408,180
Observations	2,564,331	2,564,331	1,982,172	1,857,848
Scientist FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes. This table replicates Table 3 but uses alternative definition of treated scientists. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with any coauthors from opposite party during 2012-2016 (rather than share of coauthors from opposite party in the top quartile, as in Figure 3), and control scientists are defined as scientists who did not have any coauthors from opposite party during 2012-2016. Column (1), (2) and (4) present results from Poisson model. Column (3) presents result from OLS model. Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table A7. Table 3 robust when using OLS model with treated scientists defined as a continuous treatment measure: share of coauthors from opposite party (during 2012-2016).

	(1)	(2)	(2)	(4)	(3)	(4)
	DV: Publication count	DV: Log publication count	DV: Citation-weighted publication count	DV: Log citation-weighted publication count	DV: Diversity of concepts	DV: Count of new concepts
Treated Scientist × After 2016 US Presidential Election	**-0.249*** (0.025)**	**-0.059*** (0.004)**	**-3.517*** (0.989)**	**-0.104*** (0.010)**	**-0.002*** (0.000)**	**-0.356*** (0.108)**
R^2	0.009	0.010	0.011	0.025	0.001	0.000
Observations	2,574,330	2,574,330	2,574,330	2,574,330	1,982,172	1,960,283
Age control	Yes	Yes	Yes	Yes	Yes	Yes
Scientist FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table replicates Table 3 but uses alternative definition of treated scientists. The data are a panel at the scientist-year level based on publication data for 2012–2021. Treated scientists are defined as scientists with any coauthors from opposite party during 2012-2016 (rather than share of coauthors from opposite party in the top quartile, as in Figure 3), and control scientists are defined as scientists who did not have any coauthors from opposite party during 2012-2016. All models are OLS with Robust standard errors clustered at the scientist level. Estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Appendix B

Additional Analyses

Figure B1. Raw graph versions of (a) Figure 2 where I plot the (a) average likelihood of any collaboration, and (b) average number of co-publication by types of different scientist pairs by year

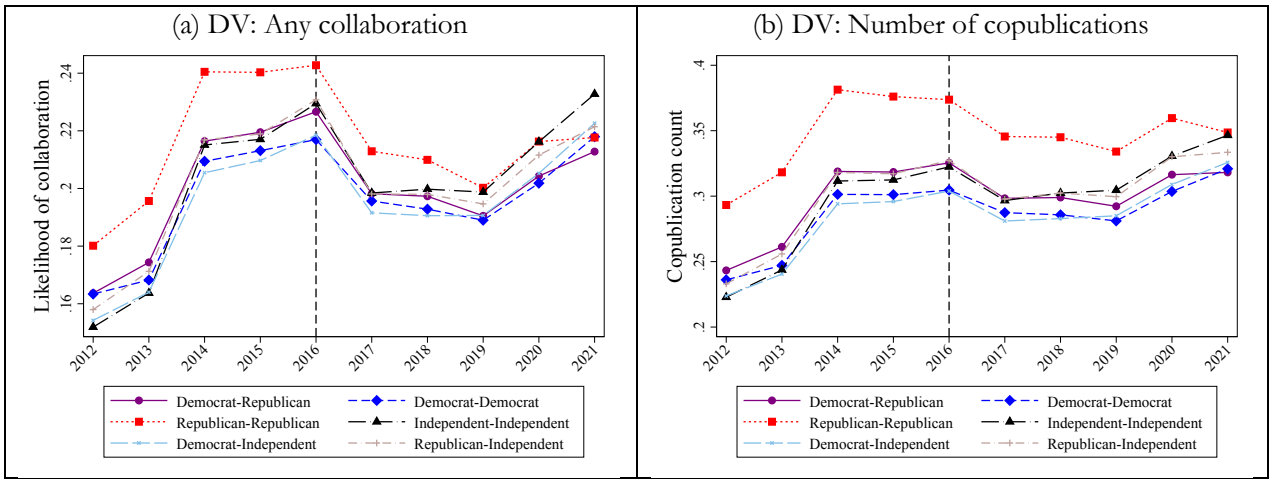
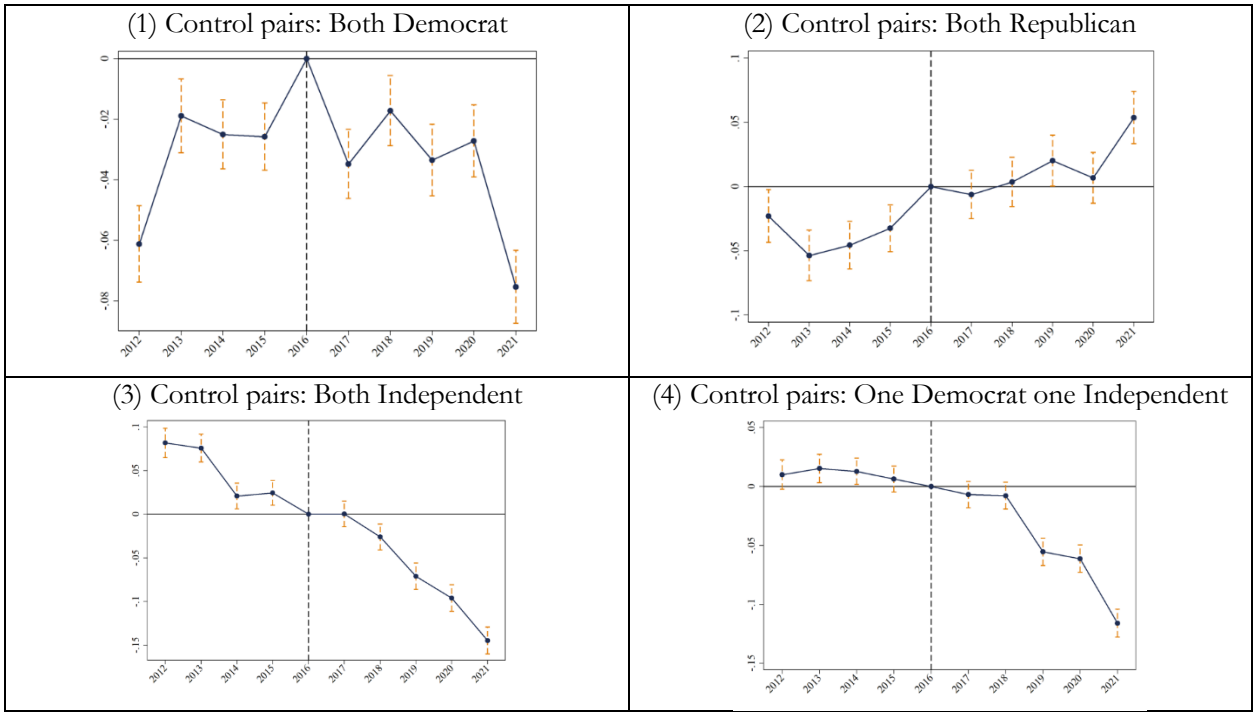
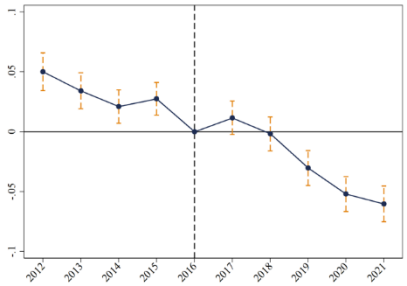


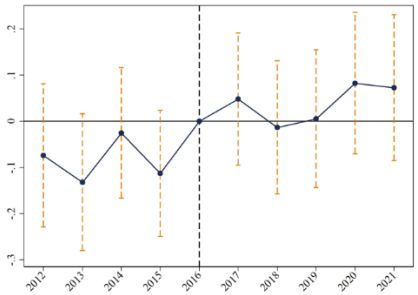
Figure B2. Main results using each party combination of control scientist pairs as control, once at a time.



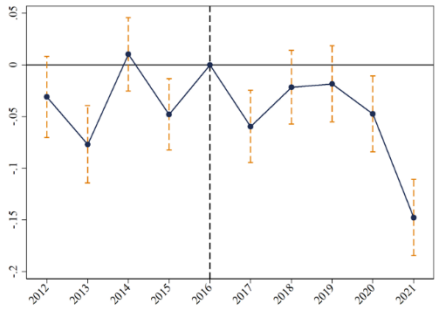
(5) Control pairs: One Republican one Independent



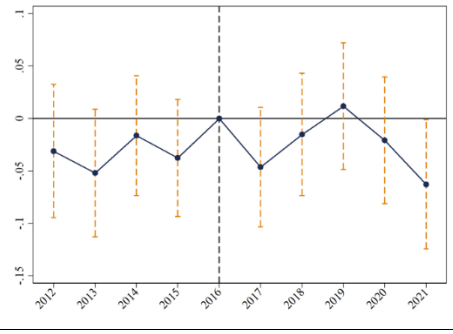
(6) Both other parties



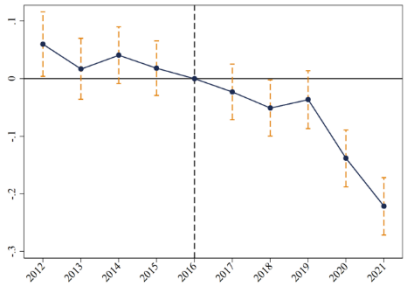
(7) Control pairs: One Democrat one other parties



(8) Control pairs: One Republican one other parties



(9) Control pairs: One Independent one other parties



Notes. These figures replicates Figure 2 use each party combination of control scientist pairs as control, once at a time. Other parties refer to parties that are non-Democrat, Non-Republican, and non-Independent. The figures plot the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist pair*×*Year* and thus describes the relative difference in likelihood of collaboration between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Figure B3. Number of scientists, new scientists, and papers, and average number of papers per scientist by party. (Note: this is based on full matched data, not our sample where we restrict to scientists with at least one paper before 2016).

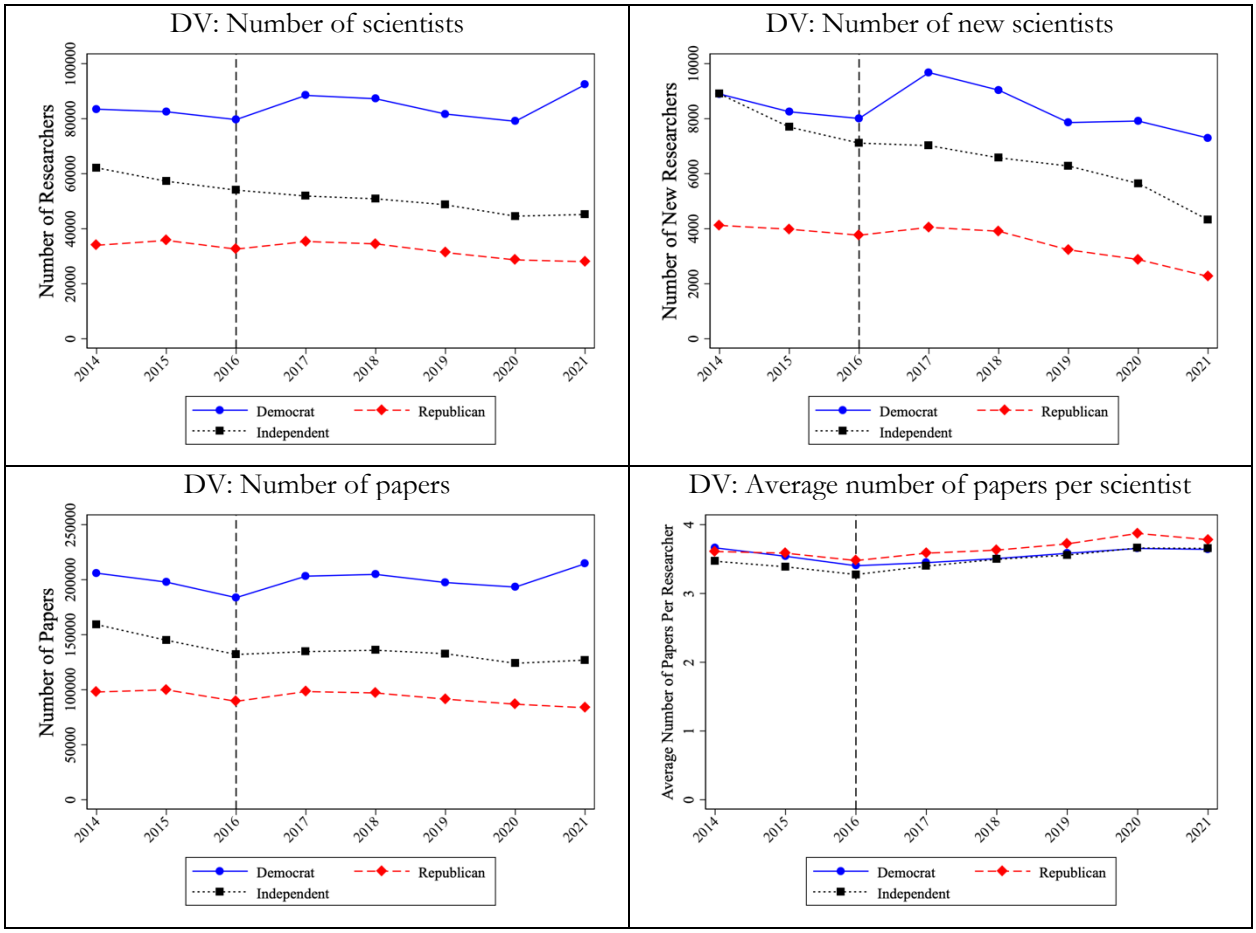
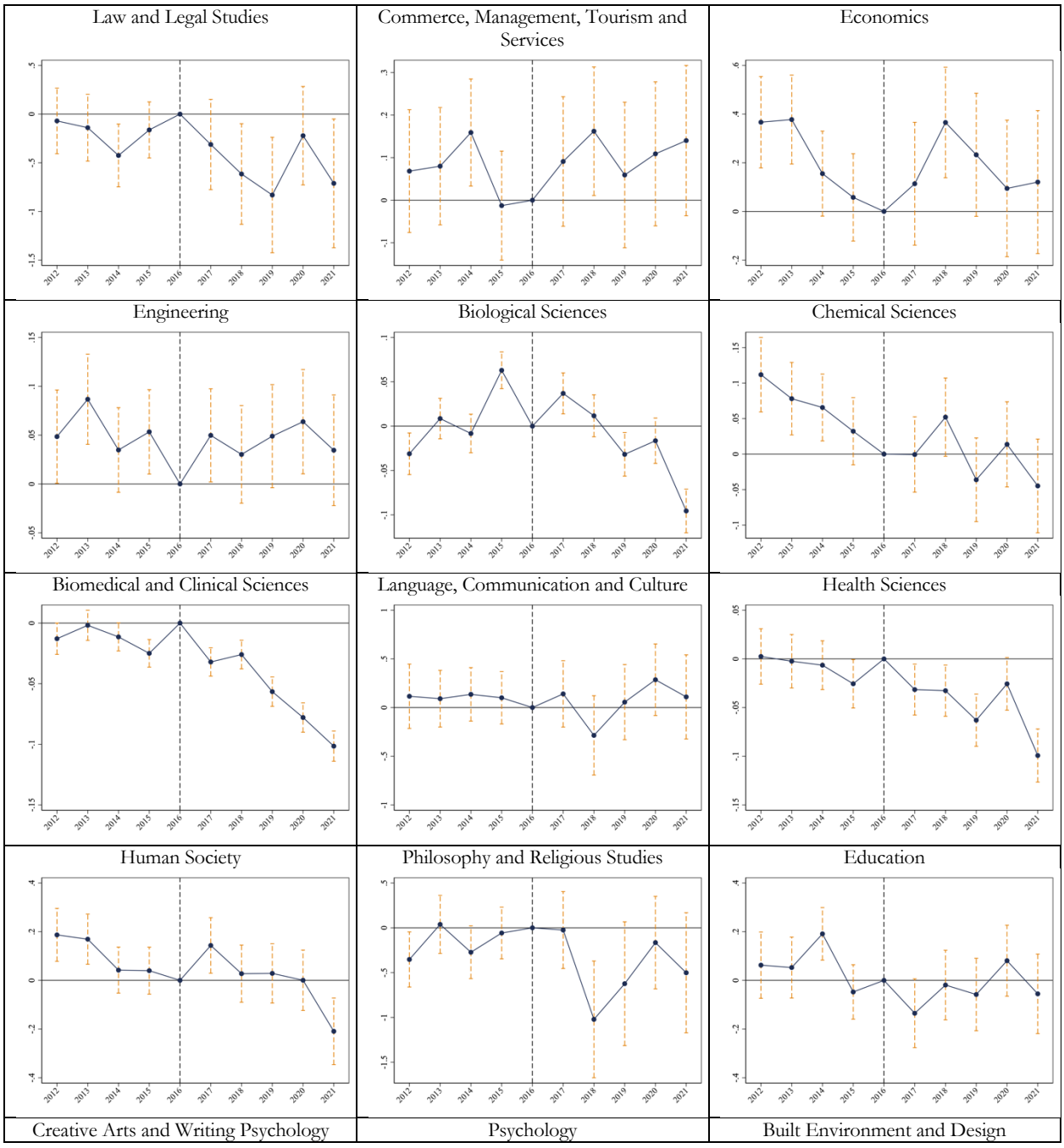


Figure B4a. Heterogeneities of Figure 2 by field of research (DV: any collaboration)



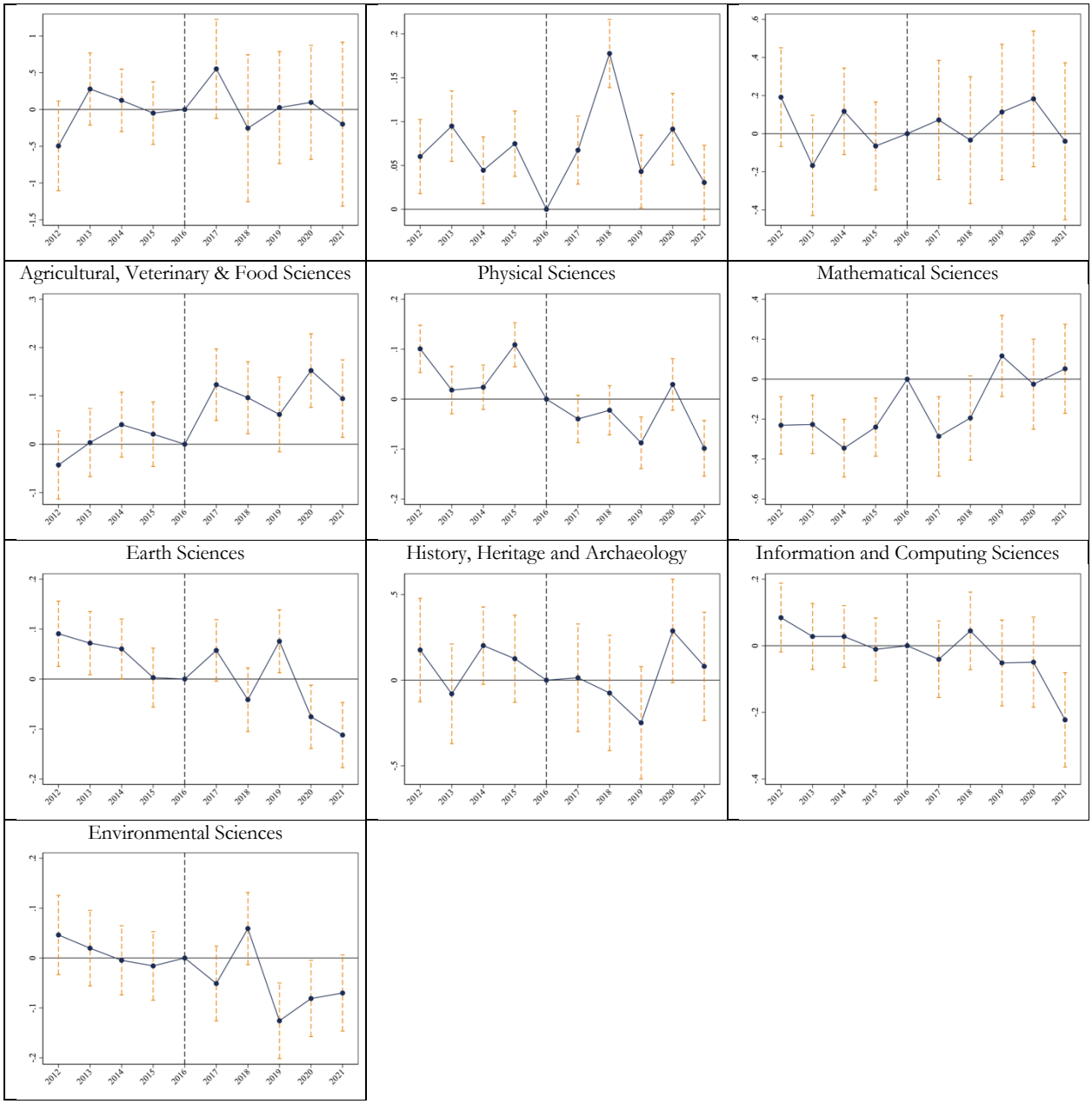
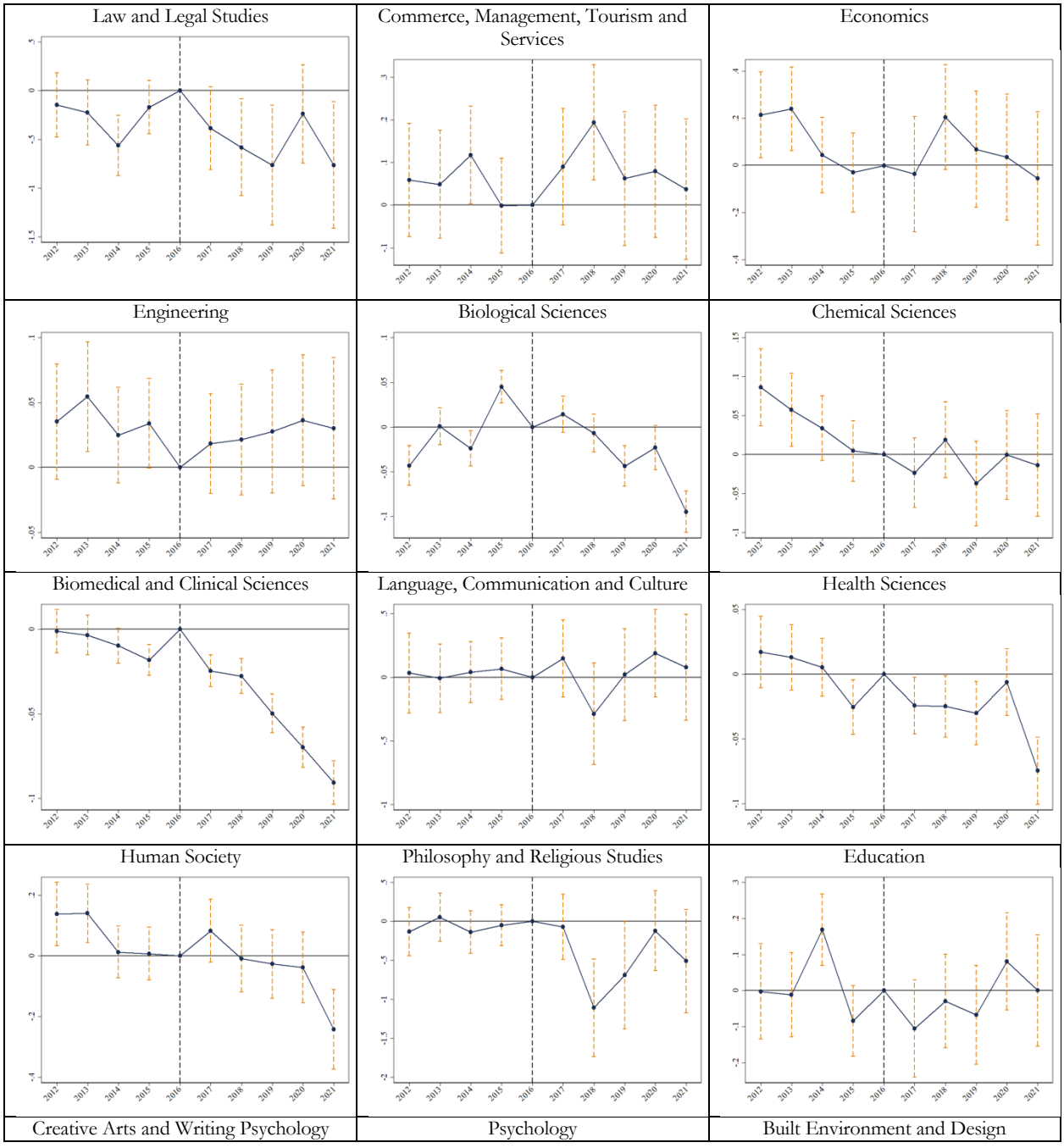

Agricultural, Veterinary & Food Sciences
Physical Sciences
Mathematical Sciences
Earth Sciences
History, Heritage and Archaeology
Information and Computing Sciences
Environmental Sciences

Figure B4b. Heterogeneities of Figure 2 by field of research (DV: number of co-publication)



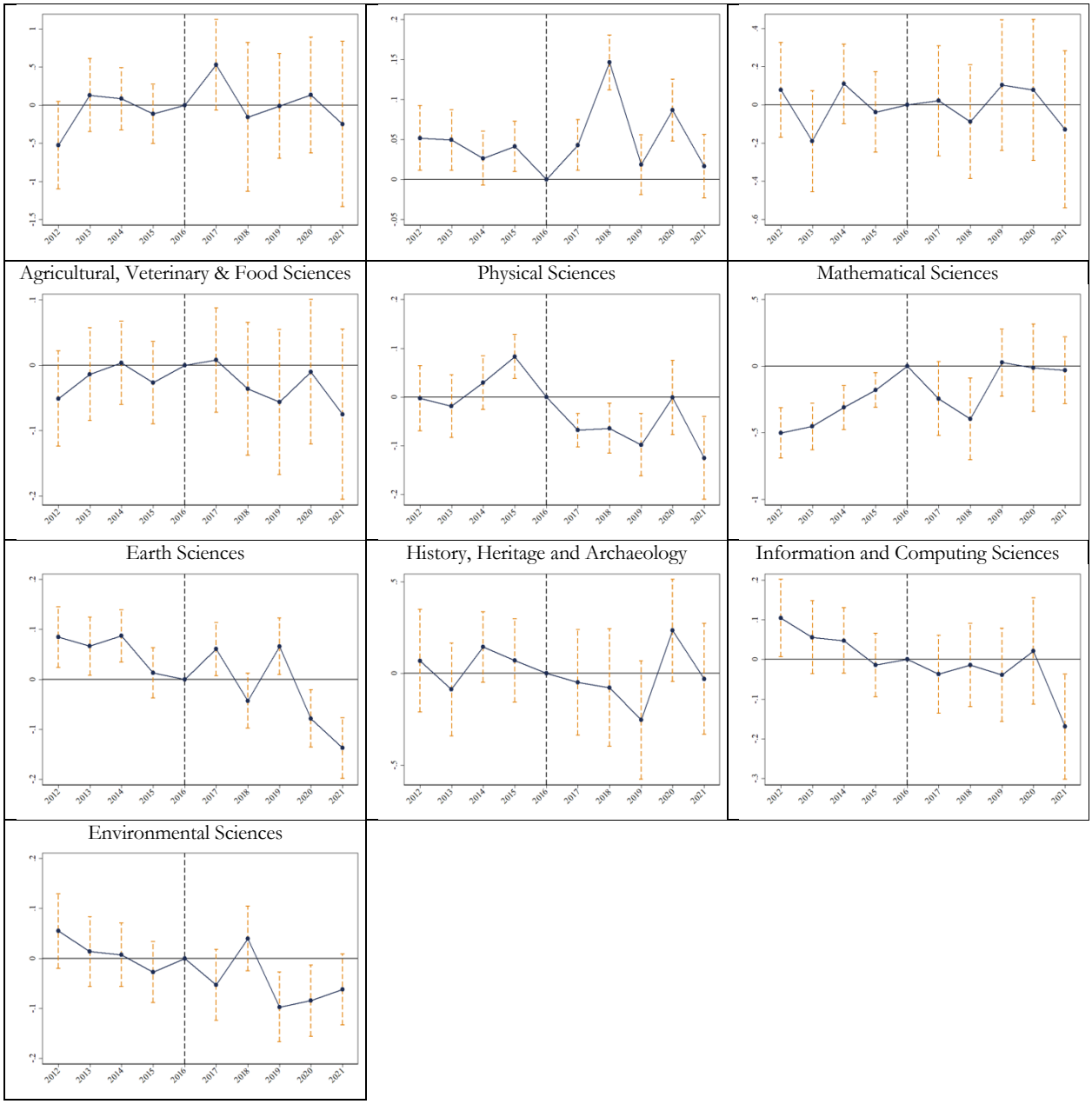
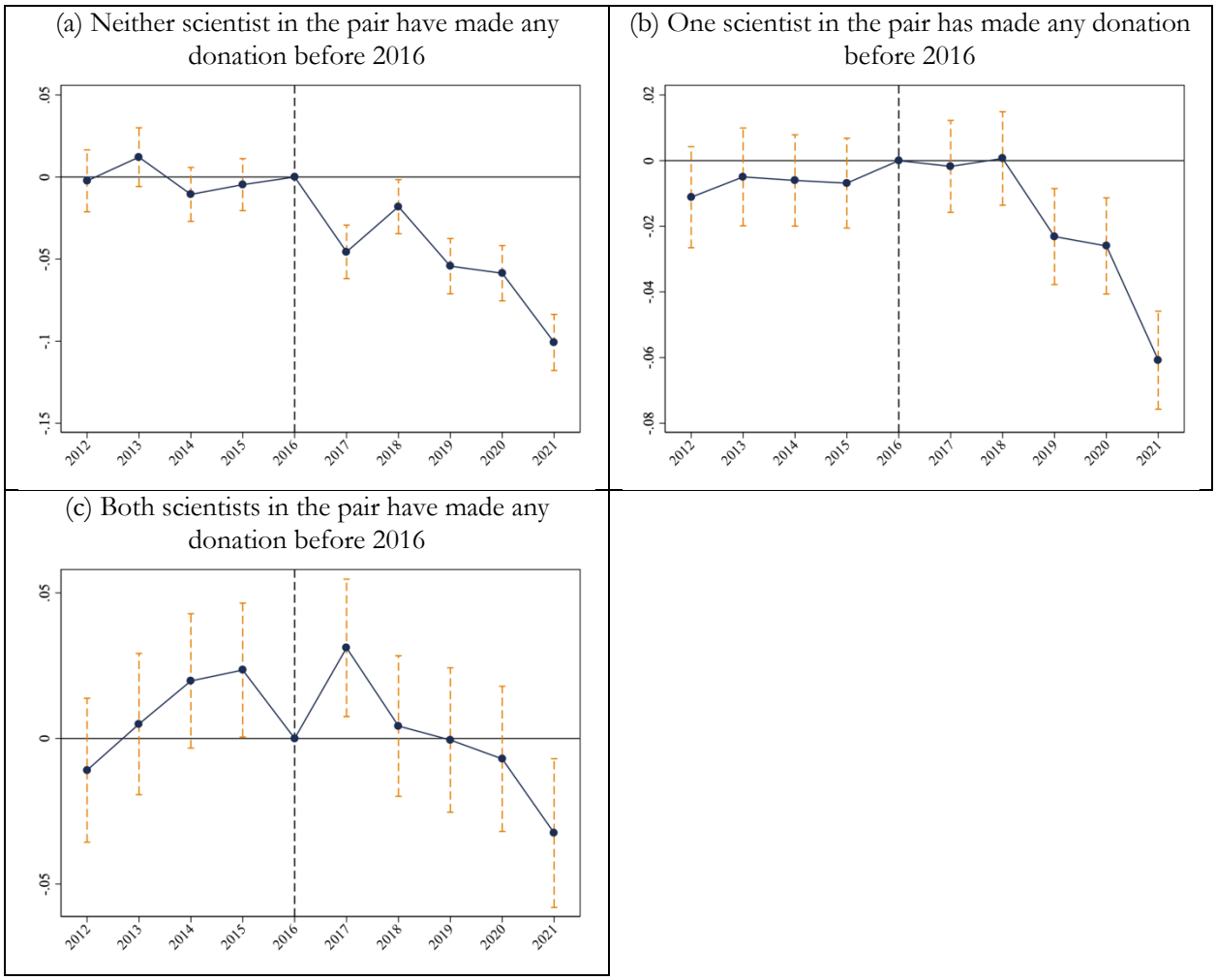

Agricultural, Veterinary & Food Sciences
Physical Sciences
Mathematical Sciences
Earth Sciences
History, Heritage and Archaeology
Information and Computing Sciences
Environmental Sciences

Figure B5. Repeat Figure 2 splits scientist pairs into different subsample based on whether they have made any donations before 2016.



*Note*s. This figure replicates Figure 2 but splits scientist pairs into different subsample based on whether they have made any donations before 2016. This figure plots the difference-in-differences coefficient estimates. Each point on the graph represents the coefficient value on the covariate *Treated scientist pair*×*Year* and thus describes the relative difference in likelihood of collaboration between treated and control research pairs in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2016.

Table B3. Main results using each party combination of control scientist pairs as control, once at a time.

DV: Any collaboration	(1) Control pairs: Both Democrat	(2) Control pairs: Both Republican	(3) Control pairs: Both Independent	(4) Control pairs: One Democrat one Independent	(5) Control pairs: One Republican one Independent	(6) Both other parties	(7) Control pairs: One Democrat one other parties	(8) Control pairs: One Republican one other parties	(9) Control pairs: One Independent one other parties
Treated scientist pair × After 2016 US Presidential Election	**-0.014*** (0.003)**	**0.043*** (0.005)**	**-0.096*** (0.004)**	**-0.055*** (0.003)**	**-0.047*** (0.004)**	**0.100*** (0.036)**	**-0.031*** (0.009)**	**-0.002 (0.015)**	**-0.113*** (0.012)**
LL	-4,499,684	-2,330,013	-2,814,193	-4,736,761	-2,981,402	-1,831,719	-1,946,634	-1,868,206	-1,886,467
Number of pairs									
Number of observations	14,501,641	7,444,525	9,040,222	15,307,252	9,567,524	5,881,584	6,253,156	5,998,986	6,062,533
Scientist pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table replicates Table 2 within different fields of research. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Different columns correspond to different subsamples. Logit model with estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table B2. Average number of coauthors per paper by field of research

Field	Average number of coauthors
Law and Legal Studies	3.3
Commerce, Management, Tourism and Services	3.07
Economics	3.29
Engineering	4.84
Biological Sciences	6.51
Chemical Sciences	5.67
Biomedical and Clinical Sciences	6.72
Language, Communication and Culture	2.46
Health Sciences	5.47
Human Society	3.21
Philosophy and Religious Studies	2.05
Education	3.18
Creative Arts and Writing	2.51
Psychology	5.01
Built Environment and Design	3.67
Agricultural, Veterinary and Food Sciences	5.33
Physical Sciences	18.25
Mathematical Sciences	14.7
Earth Sciences	5.53
History, Heritage and Archaeology	2.95
Information and Computing Sciences	4.07
Environmental Sciences	5.51

Table B3a. Heterogeneities of Table 2 by field of research

	DV: Any collaboration	DV: Number of copublications	Number of pairs	Number of treated pairs
Law and Legal Studies	**-0.357*** (0.127)**	**-0.315** (0.135)**	3,500	521
Commerce, Management, Tourism & Services	**0.057 (0.039)**	**0.057 (0.042)**	18,506	3,570
Economics	**0.028 (0.060)**	**-0.021 (0.064)**	10,363	1,751
Engineering	**0.004 (0.012)**	**-0.002 (0.015)**	174,705	30,119
Biological Sciences	**-0.026*** (0.006)**	**-0.028*** (0.006)**	790,421	131,598
Chemical Sciences	**-0.050*** (0.014)**	**-0.041** (0.016)**	155,335	25,146
Biomedical and Clinical Sciences	**-0.046*** (0.003)**	**-0.044*** (0.004)**	2,503,589	470,044
Language, Communication and Culture	**-0.012 (0.091)**	**0.016 (0.095)**	4,520	672
Health Sciences	**-0.042*** (0.007)**	**-0.031*** (0.008)**	549,129	99,671
Human Society	**-0.057* (0.03)**	**-0.078** (0.031)**	39,674	5,736
Philosophy and Religious Studies	**-0.251* (0.13)**	**-0.339** (0.158)**	3,660	572
Education	**-0.090** (0.036)**	**-0.053 (0.048)**	24,281	4,317
Creative Arts and Writing	**0.120 (0.195)**	**0.174 (0.194)**	1,578	268
Psychology	**0.035*** (0.010)**	**0.035*** (0.013)**	253,767	42,597
Built Environment and Design	**0.045 (0.084)**	**0.001 (0.098)**	5,523	938
Agricultural, Veterinary and Food Sciences	**0.101*** (0.018)**	**-0.016 (0.050)**	72,350	13,870
Physical Sciences	**-0.090*** (0.012)**	**-0.088*** (0.022)**	168,596	27,987
Mathematical Sciences	**0.106** (0.051)**	**0.100 (0.121)**	14,729	2,433
Earth Sciences	**-0.055*** (0.015)**	**-0.068*** (0.016)**	116,034	17,212
History, Heritage and Archaeology	**-0.072 (0.078)**	**-0.082 (0.079)**	6,349	827
Information and Computing Sciences	**-0.072** (0.030)**	**-0.074** (0.036)**	37,708	6,013
Environmental Sciences	**-0.055*** (0.019)**	**-0.053*** (0.019)**	83,753	12,524

Notes. This table replicates Table 2 within different fields of research. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Different rows correspond to different fields of research subsample. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Logit model with estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

Table B4. Repeat Table 2 splits scientist pairs into different subsample based on whether they have made any donations before 2016.

DV: Any collaboration	(1)	(2)	(3)
	Subsample: Neither scientist donate before 2016	Subsample: Either scientist donate before 2016	Subsample: Both scientists donate before 2016
Treated scientist pair × After 2016 US Presidential Election	**-0.052*** (0.004)**	**-0.015*** (0.004)**	**-0.007 (0.006)**
Log likelihood (LL)	-4,020,170	-4,740,910	-1,529,098
Number of pairs	1,527,133	1732739	540,471
Number of observations	12,990,298	15269,638	4,909,671
Scientist pair FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Notes. This table replicates Table 2 but split scientist pairs into different subsamples based on whether they have made any donation before 2016. The data are a panel at the scientist pair-year level based on publication data for 2012–2021. Treated scientist pairs are pairs with one Democrat and one Republican scientist, and control scientist pairs are other pairs, such as one Democrat and one Independent scientist, or one Democrat and one scientist from others. Different columns correspond to different subsamples. Logit model. Estimations presented as coefficient (SE). *p < 0.10; **p < 0.05; ***p < 0.01.

APPENDIX C

More information about the matching algorithm

The Algorithm

The central challenge associated with constructing our dataset is the fact that we lack a common high-quality discriminating variable which can facilitate the linking of inventors with voter registration records using conventional string matching. To overcome this challenge, our matching strategy builds on the computer science and statistical record-linking literature to implement a matching algorithm which allows us to generate reliable matches without a common high-quality discriminating variable (e.g., Enamorado 2021). Specifically, our record-linking strategy is predicated on our ability to uncover and utilize implicit heterogeneity in matching variables allowing us to infer how rare or infrequent specific combinations of matching variables are in the overall population. The general intuition is that even if datasets lack a common high-quality discriminating variable, we can leverage the commonness or infrequency of combinations of a set of lower-quality matching variables (e.g., first name, last name, geographical location) to probabilistically identify correct matches. The matching procedure works by identifying potential matches and then weighting these matches by the probability such a match would be observed at random (Christen 2012). Our matches are built using the fastLink probabilistic record-linking model (Enamorado et al. 2017), a model which uses the Expectation Maximization (EM) algorithm to fully estimate a detailed probabilistic matching that draws heavily from the canonical Fellegi and Sunter (1969) statistical record linkage paradigm (Enamorado et al. 2019). Conceptually, this paradigm assesses the likelihood ratio of correct versus incorrect matches and picks correct matches as those with the highest ratio. Calculation of the likelihood of correct and incorrect matches is assessed by the relative rarity (or commonness) of different combinations of matching features (e.g., first name, last name, geographical location). The fastLink algorithm uses this information to estimate the posterior probability that any two records between the datasets of interest are a correct match.

For example, in attempting to find a match for an individual named Daniel Wilson, the algorithm will separately consider the frequency of the name Daniel and that of Wilson in the two datasets to be matched. If both names are rather frequent, the algorithm will assign a low posterior probability that characterizes the match quality. The posterior probability will have a higher value for an individual named Daniel Dufresne since Dufresne is a less common name than Wilson. Similarly, a Wylie Dufresne will have an even higher posterior probability because Wylie is a less common name than Daniel. In addition to considering the frequency of each name separately, the algorithm also considers the frequency of combinations of first and last names. In our example, the level differences in the posterior probabilities of the three examples will depend not only on the frequency of the first names Daniel and Wylie and that of the last names Wilson and Dufresne, but also on the frequency of their combinations. The algorithm extends the same process to any set of matching variables identified by the analysis.

In our PatentsView to L2 matching process, we consider six matching variables: first name, last name, middle initial, gender, year, and commuting region. We detail the definition of commuting region below.

Given the approach of evaluating pairwise combinations of matching variables, the fastLink algorithm is computationally intensive. Matching across the PatentsView and L2 data sets would entail $1.3 * 10^{14}$ (i.e., 722,935* 182,464,067) comparisons. This amounts to approximately 78,225 hours or 4.18 years of run time when considering linear processing at a speed of one microsecond per row – an exceedingly fast computation time for standard string comparison libraries.

We address this computational complexity through two steps: reducing the total number of comparisons and parallelizing data processing. To reduce the total number of comparisons, we employ a common practice in statistical record linking called blocking, where we limit comparisons of matching variables across the two datasets to blocks of data that share a common value for one or more of our matching variables. The intuition behind this practice is to exclude pairwise comparison of records that

have a low *ex ante* likelihood of being a match. We block on commuting regions such that PatentsView inventors and L2 registered voters are compared only within each commuting region. To further reduce computational complexity, we process our matching on a high-performance computing cluster where we parallel process each commuting region across clusters.

We make our resulting code applying the fastLink algorithm publicly available via a GitHub repository.[15] The code repository is meant to enable researchers to execute our code directly in their chosen environment. While the code for the fastLink algorithm described in Enamorado (2019) is already publicly available, in what follows we describe the additional steps we take to apply the code to our specific empirical setting. Specifically, in this supplemental information, we provide a description of the different files in our repository and steps our code takes to achieve the matching.

Commuting Region Definition

L2 provides the registered home addresses of all voters in the national voter file as part of the required proof of eligibility to vote in a particular district. We use this data as a matching variable and to reduce the number of comparisons between L2 and PatentsView by making use of the concept of a commuting region. Specifically, for each address in our data, we assign a commuting region code derived from the U.S. Census definitions of Core-Based Statistical Areas (CBSA). The U.S. Census Bureau routinely measures the daily commutes of a large representative sample of U.S. workers and uses this commute data to define geographical areas. Based on the results of the survey, each county in the United States is classified by the Bureau as either a Metropolitan Statistical Area (population of 50,000 or more) (MSA), a Micropolitan Statistical Area (a population of at least 10,000 but less than 50,000) or Rural. MSAs can be comprised of multiple counties. The Bureau groups counties that have substantial economic relationships, defined by having at least 25% of workers commuting between the counties, into MSAs. Thus, the MSAs can be comprised of one or multiple counties. In addition, the Census

[15] To preserve anonymity in the review process, we do not disclose the GitHub link in our submission but will put the URL of the GitHub repository here. If reviewers wish to view the code, we can provide an anonymized version of the code.

Bureau defines Combined Statistical Areas (CSAs) as groupings of MSAs based on wider commuting criteria across MSAs. Not all MSAs are grouped into CSAs.

We leverage this nested U.S. Census Bureau's coding scheme to define our commuting region codes. For each county in the United States, we assign its commuting region code as its CSA code if the county is a member of a CSA. If the county is not a CSA member, we assign the commuting region as its MSA code, and if the county is not part of an MSA, we assign the commuting region as its Micropolitan Area code. Lastly, if the county is not a member of a CSA, MSA, or Micropolitan Area, we assign the commuting region as its county code. Using this coding scheme, we are able to assign all addresses in the L2 and PatentsView data sets to a commuting region code.

Dimensions Data Pre-Processing

We downloaded the full Dimensions data set in February of 2022 of academic journal articles. Before the data can be put through the fastLink algorithm, it needs to be processed from its original raw form to fit the algorithm's required format. We achieved these processing steps through a set of computer programs contained in the `preprocess_academic/` subfolder of our GitHub repository. The full sequence of programs for processing the inventor data is initiated with the file `0_proc_regional_academic.job`.[16] We provide a visual overview of the processing steps in Figure B1 to accompany our description of these processing steps.

The first step in our process is to take the raw data from Dimensions dataset that contain and restrict our analysis to academic papers with at least one co-author in the United States published in the years 2012-2021. The Dimensions data provides information on each author's academic affiliations which we use to create our commuting zone blocks. From this data set, we then create a researcher-year-city level data where each row corresponds to an inventor that published a paper in a given year in a

[16] We completed all data processing and the execution of our matching in USC's Advanced Research Computing high performance cluster computing environment which uses the SLURM workload management system. Within this system, .job files are used to schedule and execute jobs.

given city. To this data, we then append our commuting zone code described above using a city to commuting zone crosswalk. At this stage, the data is now at the researcher-year-commuting-zone level.

Because not all researchers publish every year, there might be researchers who are not included in the inventor-side matching consideration set in a given year even though they continue their inventive activity as evidenced by later patents. For these inventors, we apply an imputation algorithm so that we can track the evolution of their partisan affiliation over time including years in between patenting events. Our imputation algorithm identifies two cases: (1) the researcher has multiple papers showing the same location in different years with non-publishing years in between and (2) the researcher has multiple papers showing different locations in different years with non-publishing years in between.

To address case (1), we impute the rows between publishing years with the same location data as the publishing years. For example, if a researcher had papers published in 2014 and in 2017 which both have the same city and state, we include an entry for that researcher for that city and state in the years 2015 and 2016. For case (2), because we cannot identify when the researcher moved, we impute the rows between patenting years with both the location data of the first and the second publishing event. For example, if a researcher published in Los Angeles, CA in 2014 and then in New York, NY in 2017, we would include rows for that researcher in both the Los Angeles, CA and New York, NY for 2015 and 2016, for a total of four additional rows. As it will become clear in our description of the matching process below, the final match for the researcher in each year will then be the voter who best matches the researcher in each year from all inputted potential matches.

The last step in our data preparation is to create separate data files for each year and commuting zone in which an inventor is observed either directly or through imputation. Each of these data files is used iteratively in the fastLink matching process described below.

The final output of the researcher preprocessing process is a set of files with data at the year-commuting zone level where each file contains all researchers that either published a paper in that year and identified that commuting zone as the area of affiliation or published a paper in that year and

showed evidence of being in that commuting zone in that year, as per our data imputation process described above.

L2 Voter File Data Pre-Processing

The code to process the L2 data files is contained in the `preprocess_l2_by_year/` folder of our GitHub repository. Each of the steps described in the process is initiated by launching the `l2-master.job` batch file in our cluster environment. In Figure B2, we provide a visual summary of the logic contained in the process.

The raw L2 Voter Registration data contains separate files for each state and for each of the years in our study (2012-2021). Because voters are required to provide their home address as part of the voter registration process, L2 provides information on the county of residence for each voter in the data. We match the county to commuting zones using the crosswalk to obtain a commuting zone for each row of the L2 data. Because commuting zones can cross state lines, our data processing algorithm then takes each state-year file and iteratively appends each row of data to the proper commuting zone-year file. At the end of this process, we obtain separate files for each commuting zone and year in which a voter is observed. Each of these data files is used iteratively in the fastLink matching process described below.

Executing the fastLink Algorithm

The code associated with the execution of the matching algorithm is contained in the `fastlink/`folder of our GitHub repository and is initiated by the `run_fastlink.job` batch file in our cluster environment. We provide a visual description of the full matching process including the step below in Figure B3. For each year and commuting zone in our data, the code loads each of the voter and inventor files into the fastLink algorithm and then processes and outputs the results. The core fastLink algorithm is part of the inherited library downloaded directly from the CRAN Network.

Final Matched Data

The code associated with the production of the final matched data is contained in the `crosswalk/` folder of our GitHub repository. The code is initiated and executed by the `agg_match.job` batch file in our cluster environment. The output files from the fastLink matching

process have multiple candidate matches with a posterior probability above our threshold of choice (0.75). The first step in this process is to aggregate all the commuting zone files for each year into one file. If we find multiple matches with different posterior probabilities for one inventor, we choose the match with the highest posterior probability as it signifies the most likely match. Next, we eliminate all matches that were indeterminate because there were multiple inventor-voter matches that received the same posterior probability (i.e., we exclude matches for inventors where we did not observe a 1:1 match to the voter file that was clearly ranked above others in terms of posterior probability). Lastly, for our main analysis, we exclude any matches that did not receive a posterior probability score at or above 0.95. However, our matched dataset includes the posterior probabilities for each match, thereby allowing researchers to specify the degree of matching accuracy they require for their specific application. Table B1 describes the number of matches at different levels of posterior probabilities.

Benchmarking the Matching Results

To validate our matching approach and to demonstrate that our matching strategy generates quality matches, we benchmarked our matching results against prevailing record-linking approaches which use a high-quality discriminating variable in the matching process. Because no such variable exists between PatentsView and L2, we integrated a third database to assist in this benchmarking exercise, the Database on Ideology, Money in Politics, and Elections (DIME), a comprehensive database which aggregates political campaign contributions across a variety of sources (Bonica 2014). The intuition underlying our approach is that the DIME database has information which overlaps with both PatentsView and with the L2 data, and hence this overlapping information can be used to assess the quality of matches. Specifically, the DIME data includes information on the donor's employer (contained in PatentsView) and it contains the donor's home address (contained in L2). We focus on the DIME

data that covers election cycles occurring in 2014, 2016 and 2018. There are 5,734,103 donors in the 2014, 2016 and 2018 election cycles that have their last name, first name and employer listed in DIME.

We first match inventors in PatentsView with donors in the DIME database using prevailing string-matching methods which link records using first and last name along with employer as the discriminating variable, creating a crosswalk of matched records between PatentsView and DIME for each year of our panel. Given the benchmarking purpose of this exercise, we match the PatentsView data to the DIME data using the most stringent implementation of string matching – an exact match criterion on each matching variable. While this limits the number of total matches, the benefit is it minimizes the rate of false positives, thereby ensuring that our PatentsView to DIME matches can be used as "ground truth" data to assess the quality of our PatentsView to L2 matches.

Next, we use this crosswalk to assess the quality of our matches between PatentsView and L2. To do so, we leverage the fact that we can isolate records in PatentsView which we successfully linked to DIME (using name string-matching with employer) and the L2 database (using the fastLink EM algorithm), effectively creating a PatentsView-DIME-L2 crosswalk. The resulting three-way match data sets is comprised of 10,085 researcher-year records.

If our EM algorithm performed well, then we would expect that, for these records, the home address listed in DIME and the home address listed in L2 would have a high degree of overlap. That is, a high degree of overlap would suggest that our matches using the EM algorithm are identifying the same matches which would be generated using string-matching using employer as a discriminating variable. After tabulating the distances between the address coordinates reported in the two databases for a given inventor, we find that 56.23% of our benchmark matches using the fastLink algorithm are less than 2 miles apart, 62.34% of our matches are less than 10 miles apart, and 68.53% are less than 100 miles apart. These results are displayed graphically in Figure B2. Panel A shows the distribution of distances between the DIME and L2 addresses for the full data, revealing that a preponderance of our data clusters in the smallest distance bin. Since the smallest distance bin still covers distances up to 150 miles,

we zoom in by excluding the top 10 percentile of inventors by distance. We plot the remaining data in Panel B and show that the majority of our matched records cluster in the smallest bin of distance i.e., below 25 miles. We continue to zoom in by eliminating the top 25 percentile of inventors by distance (Panel C) and the top 50 percentile of inventors by distance (Panel D), revealing that the majority of our matched records have a distance less than 2 miles apart. More specifically, we find that 51.48% of our inventors have addresses reported in L2 and DIME which are less than one mile apart.

It is important to reiterate that the geographic data in the DIME is self-reported and thus may not reflect the donor's true home address. It is with this in mind that we interpret the results of the benchmarking exercise (i.e., 51.48% of records within 1 mile) as strongly suggesting our algorithm performs well when compared to conventional name and employer string-matching.

Specifically, approximately 2.7% of DIME records report only their city and state or zip code, rather than their full home address. For such records, the longitude and latitude provided in DIME reflect the center of the city or zip code. Second, different geocoding algorithms can produce different coordinates if addresses do not provide fully deterministic information in terms of geolocation (i.e., when addresses are somewhat incomplete). To demonstrate, we ran the DIME addresses for all matched records through Google's Geocode API. We found that roughly 27% of the coordinates produced by the Google API were greater than 1 mile apart from the coordinates for the same address provided in DIME. Third, the self-report nature of the DIME data means that some contributors in DIME provide a business address rather than their home address. To quantify this, we once again used Google's geocoding API, the results of which suggest approximately 22.19% of DIME records are likely commercial addresses. [17] Lastly, DIME contributors may change home addresses for various reasons (e.g., approximately 6.4% of individuals moved in calendar years 2018-2019 according to the IRS [18]). It is

[17] Determination of business and residential addresses was made using US Postal Service data made available through Google's Address Validation API (https://developers.google.com/maps/documentation/address-validation). We coded addresses as business addresses if the "business" flag in the API response was set to TRUE.

[18] https://www.irs.gov/pub/irs-pdf/p5508.pdf

with the above information in mind that we view our benchmarking results to be indicative of a strong performing matching algorithm and one that compares well against conventional matching.

Ruling out Bias in the Matching Process

One of the principal concerns to using our data in research projects is that our matching procedure systematically over or under samples inventors with certain political orientations relative to the full inventor population. For example, it is theoretically possible that individuals with common last names, such as Smith or Li, would, on average, skew more right or left on the political spectrum. The more common the name, the lower the probability that our algorithm would return a high matching posterior probability. More broadly, any reasonably strong correlation between the commonness of inventor matching attributes and their behavior or preferences would result in a matched sample that is not representative of the inventor population. Ideally, our matched sample should be as close as possible to a random sample of the inventor population. Statistically, this ensures generalizability of any research results to the population of inventors.

To this end, we check that last names in our L2 voter file are not systematically related to political partisanship in a way that might bias the results of our matching algorithm. Since our matching algorithm gives extra weight to infrequently observed names, a key concern would be that there is correlation between partisan affiliations and last name (un)commonness. To check this possibility, we tabulated the frequency of last names in the entire L2 database as well as the counts of registered Republicans and Democrats for each last name. Of the 6,657,215 distinct last names in the L2 data, we report the fifteen most common names in our data and their share of registered Democrats and Republicans in Table B2. While there is substantial variation in the share of registered Democrats and Republicans by last name, there is no clear pattern in the data relating name frequency and political partisanship. We further test for a potential relationship by measuring the correlation coefficient between name frequency in the L2 data and share of Republican registration and find a correlation coefficient of 0.0069. We also measure the correlation coefficient between the name frequency in the L2 data and share

of Democrat registration and find a correlation coefficient of -0.0029. The lack of correlation between last name frequency and political party affiliation suggests last name frequency is unlikely to be a source of bias for research preoccupied with the role of inventors' partisanship that uses our matched data.

Figure B1: Processing steps to prepare the Matched Dimensions L2 data

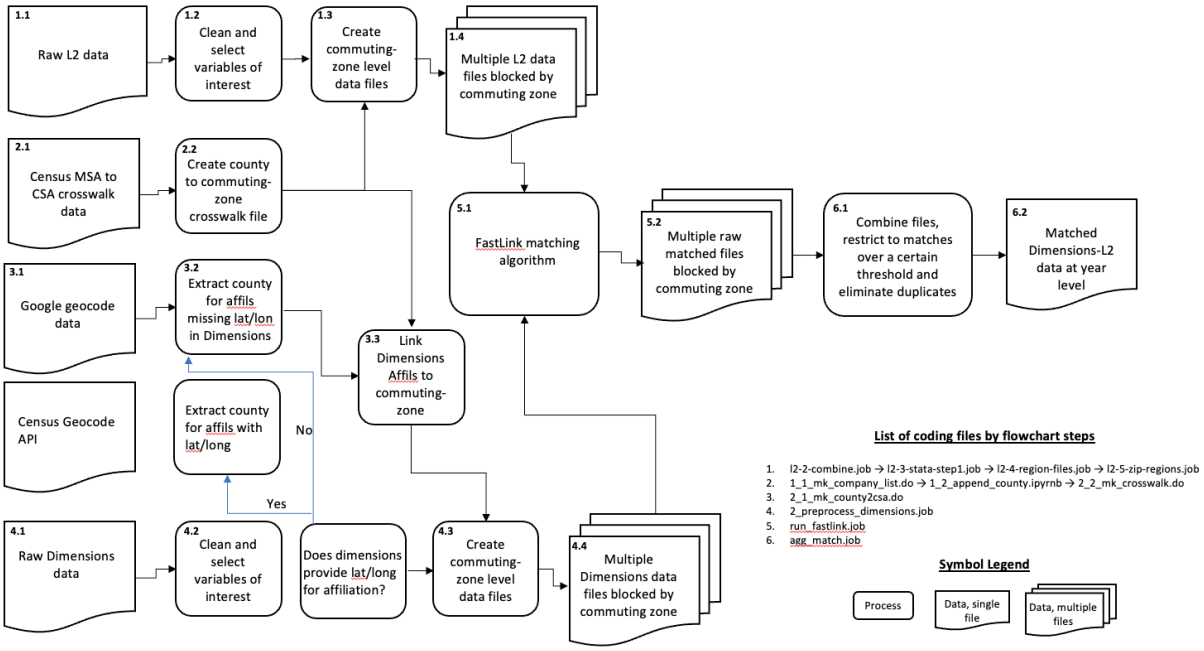
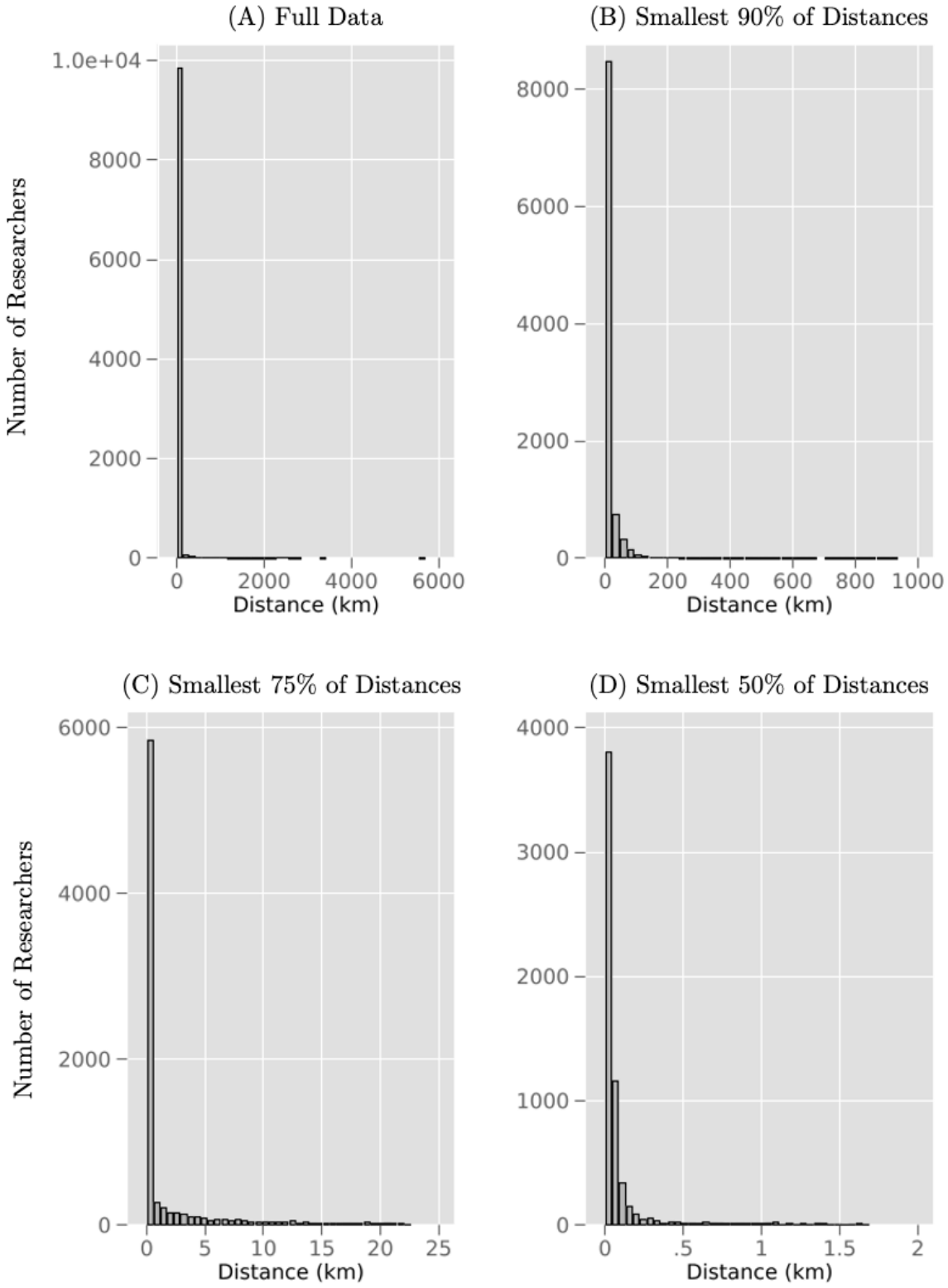


Figure B2: Frequency of geographic distances between home addresses reported in L2 and DIME for inventors matched to both data sources



Note: Using a ground truth data set from DIME where inventors self-report both their address and their employer, we measure the distance between those self-reported addresses and the addresses in our PatentsView – L2 match data. The plot in Panel A shows the distribution of the 6,162 inventors in our ground truth data. In Panel B, we restrict to the smallest 90% of distances. In panel C, we restrict to the smallest 75% of distances in our data. Finally in Panel D, we show the distribution of the smallest 50% of the data in terms of distance.

Table B1: Distribution of Posterior Probability of Correct Matches between L2 and Dimensions

Posterior Probability Range	**Number of Matches**	**% of Matches**	**% of Dimensions Sample**	**Cumulative % of Dimensions Sample**
95-100%	628,160	94.54%	65.00%	65.40%
90-95%	25,663	3.86%	2.7%	67.63%
85-90%	5,320	0.80%	0.60%	68.18%
80-85%	3,456	0.52%	0.4%	68.54%
75-80%	1,852	0.28%	0.2%	68.73%
Total	**286,994**	**100%**	**68.7%**	

Table B2: Frequency and Political Affiliation of the Fifteen Most Common Names in L2 Data in Commuting Zones with at Least One Inventor (2014-2020)

Last Name	**Frequency in L2 (2014-2020)**	**Republican Share**	**Democrat Share**
Smith	11,321,852	32.85%	39.02%
Johnson	9,057,045	27.62%	44.78%
Williams	7,592,997	22.33%	51.56%
Brown	6,724,800	27.16%	45.83%
Jones	6,645,011	26.66%	46.56%
Miller	5,394,510	37.10%	34.23%
Davis	5,231,720	29.27%	43.57%
Wilson	3,805,353	31.35%	41.05%
Anderson	3,751,136	34.03%	37.10%
Thomas	3,520,370	25.43%	47.71%
Taylor	3,515,561	30.49%	41.77%
Moore	3,417,790	30.89%	41.72%
Rodriguez	3,378,401	13.51%	59.37%
Jackson	3,295,622	19.91%	54.34%
Garcia	3,282,567	13.80%	60.02%