Inside Job or Deep Impact?
Using Extramural Citations to Assess Economic Scholarship*

Josh Angrist, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu†

August 2017

Abstract

Does academic economic research produce material of scientific value, or are academic economists writing only for clients and peers? Is economics scholarship uniquely insular? We address these questions by quantifying interactions between economics and other disciplines. Changes in the impact of economic scholarship are measured here by the way other disciplines cite us. We document a clear rise in the extramural influence of economic research, while also showing that economics is increasingly likely to reference other social sciences. A breakdown of extramural citations by economics fields shows broad field impact. Differentiating between theoretical and empirical papers classified using machine learning, we see that much of the rise in economics' extramural influence reflects growth in citations to empirical work. This parallels a growing share of empirical cites within economics. At the same time, the disciplines of computer science and operations research are mostly influenced by economic theory.

*Thanks go to our outstanding research assistants, Alonso Bucarey, Gina Li, Jose I. Morales, and Suhas Vijaykumar, and to Ben Jones and Heidi Williams for help and advice regarding Web of Science data. We're also grateful to Satyam Mukherjee and Brian Uzzi for facilitating access to the Northwestern University Web of Science database for exploratory analysis, and to Liz Braunstein for help with *Econlit*. Thanks also to Stefano DellaVigna and seminar participants at the University of Rome Tor Vergata and Microsoft Research for helpful comments.

†MIT, IZA, and NBER; MIT and NBER; MIT and NBER; MIT; and Purdue and Northwestern Universities.

1 Introduction

Academic scholarship is often evaluated by the nature and extent of academic citations. This essay looks at the impact of economic scholarship through the lens of *extramural citations*, quantifying the extent to which economic research in peer-reviewed journals is cited by scientists and scholars working in other disciplines. Our first goal is to look at changes in extramural citation patterns since 1970. We compare the frequency with which papers in a variety of other disciplines cite economics to the frequency with which they cite other social sciences. After showing that economics has a large and in some cases increasing impact on a number of our sister social sciences as well as on disciplines further afield, we describe the features of economic research that are responsible for economics' extramural influence.

Our inquiry comes in the wake of intellectual fallout from the Great Recession, which led media and professional observers to cast a harsh spotlight on the typically grey world of academic economic scholarship. The popular indignation captured in the film *Inside Job* is echoed by concerns within our discipline that economists might indeed be captured by special interests (see, *inter alia*, Zingales (2013)).[1]

Fourcade, Ollion and Algan (2015) offer an especially jaundiced view of economic scholarship. They argue that the influence of academic economics on public discourse and policy is the fruit of a disciplined but insular professional culture that discounts contributions by other social scientists, imposes dogmatic standards in teaching and training, and uses tightly controlled hiring practices to enforce conformity with the norms and goals of a guild-like professional elite. By these lights, economic scholarship benefits the well-paid and clever economists who practice it, and the finance industry for which economists provide sophisticated rationalization, but generates little value for the rest of society.[2]

We investigate economics' extramural scientific influence in a limited but empirically grounded manner. Although our conclusions regarding the value of academic economic scholarship differ from those of Fourcade, Ollion and Algan (2015), the picture painted here has features in common with theirs. Historically, economics has indeed been the least outward-looking social science. Our discipline is likely to benefit from explorations further afield; the increasingly fruitful interactions between economics and psychology are a leading example. The value of bibliometric research on interdisciplinary interactions, illustrated by Fourcade, Ollion and Algan (2015), emerges from our analysis as well. Extramural citation flows provide evidence of influence and scientific value, while mitigating concerns about within-discipline strategic considerations, career concerns, and faddishness.

Our analysis uses the *Web of Science* (hereafter, *WoS*) citation database for the period 1970-2015 to quantify citation patterns between economics and sixteen other disciplines. We do this by associating "disciplines" with sets of journals. This analysis necessarily looks only at disciplines with a largely journal-based academic literature. Our analysis starts by examining citation flows among social science disciplines that

[1]Partly in response to this concern, major scholarly institutions, most notably the American Economic Association and the National Bureau of Economic Research, have raised requirements for disclosure of potential conflicts of interest.

[2]Piketty (2014) adopts this posture as a promotional strategy:"There is one great advantage to being an academic economist in France: here, economists are not highly respected in the academic and intellectual world or by political and financial elites. Hence they must set aside their contempt for other disciplines and their absurd claim to greater scientific legitimacy, despite the fact that they know almost nothing about anything."

can be seen as offering complementary or competing paradigms for the study of human behavior: anthropology, economics, political science, psychology, and sociology. We also look at social science citation rates to non-social-science disciplines. Comparisons of citation flows between social sciences reveal large differences and important changes over time. Economics is among the more insular social sciences. But economics is increasingly outward-looking, with a citation rate to extramural social sciences well ahead of that from psychology since the turn of the millennium. Economics is also roughly tied with psychology and anthropology in second place for citations to the group of non-social-science and non-business disciplines.

Our main focus is the extramural influence of economics, that is, the extent to which others cite us. We begin by documenting levels and trends, comparing extramural citations to economics with the extramural citation rate to other social sciences. We gauge the relative influence of economics and sociology, for example, by comparing the rates at which economics and sociology are cited by political science and computer science since 1970. This analysis answers questions distinct from the question of whether individual social sciences are insular or outward-looking. The high level of extramural citations to economics suggests that many non-economists find our work interesting, though other social sciences also have considerable extramural impact. We're especially interested in changes over time: since the 1990s, economics' extramural impact on many disciplines, including psychology, computer science, public health, operations research, and medicine, has grown steadily.

We also distinguish the features of economics that appear to be responsible for a high or increasing level of extramural influence. Our categorization of features begins with fields. An interesting finding here is that the diversity of economics is an important source of its strength. Different disciplines cite papers from different fields and many fields are cited by at least one discipline. For example, sociologists cite labor economics; public health and medicine cite public finance; marketing cites industrial organization; computer science, psychology, and operations research cite microeconomics; and statistics cites econometrics.

Our examination of extramural citations is also motivated by a marked shift towards empirical work in economic research. We use machine learning techniques to document this shift for a longer period and wider sample of journals than earlier analyses of changes in economics scholarship.[3] Angrist and Pischke (2010) argued that the growing importance of empirical work has been concomitant with increasing quality, a phenomenon the Angrist-Pischke essay calls a "credibility revolution." By this account, empirical work in economics has benefited from the increased use of randomized trials and the quasi-experimental research designs. Here we investigate the extent to which the increasingly empirical orientation of economics has stimulated interaction with other disciplines.

In 1990, only about one-third of citations from top economics journals were to empirical papers. The empirical citation deficit was reflected in both worse journal placement of empirical papers and the fact that, conditional on the publishing journal, empirical publications received fewer cites. For example, empirical papers were less likely to appear in the *American Economic Review* (AER), and empirical papers in the

[3] See Hamermesh (2013, 2018), among others. Backhouse and Cherrier (2014) uses the fields of Clark medalists as an indicator of economics' growing empiricism. Biddle and Hamermesh (2017) also documents trends within empirical microeconomics, focusing on the movement away from Cowles-Commission-inspired structural empirical work.

AER were cited less often than theoretical papers in the AER. This has since changed: the empirical share of citations from top economics journals has increased by about 20 percentage points. And empirical papers now receive more citations than theoretical papers published in the same journal. The pattern of many disciplines' extramural citations to economics research styles exhibits a parallel shift. In citation flows from most of the disciplines where economics has long been influential, and in some where economics is growing, we see a shift toward empirical work. But there are some noteworthy exceptions to this pattern, where recent surges in extramural influence reflect growing interest in theoretical work.

2 Measuring Influence

2.1 Defining Disciplines

For our purposes, disciplines are defined by their journals. This is accomplished by first identifying "trunk journals" for each discipline. Trunk journals are mostly flagship journals published by a leading American professional association. For example, the AER, published by the American Economic Association, and the *American Sociological Review* (ASR), published by the American Sociological Association, provide the economics and sociology trunks. Each discipline's journal list is built from the journals most highly cited by its trunks. The economics and sociology disciplines therefore consist of journals highly cited by the AER and ASR. Appendix Table A1 lists professional associations and trunk journals for each discipline. For disciplines with no single obvious trunk, we chose one or two leading journals.[4] We also consider a distinct "multidisciplinary science" discipline, defined as the set of publications in three highly-regarded multidisciplinary journals, *Science*, *Nature*, and *Proceedings of the National Academy of Science Citations* (PNAS).

The journal list for each discipline begins as the 50 journals most cited by the relevant trunk in any decade starting in 1970 (specifically, decades are defined as 1970-1979, 1980-1989, 1990-1999, 2000-2009, and 2010-2015). We count references to papers published at any time. These initial journal lists have some overlap and some assignments that seem incorrect. *Econometrica*, for example, is cited heavily by both the economics and statistics trunks, and the AER is among the 50 journals most cited by the *American Political Science Review*. The appendix describes how journals appearing on any preliminary list were assigned to at most one discipline, with the goal of eliminating overlap and correcting seeming mistakes. The assignment procedure considers journal rank in an ordered list of trunk journal citations and the frequency that journals cite each trunk in each decade.

Although our journal selection process involved editorial judgements, it is important to note that, by virtue of the weighting scheme discussed below, down-list journal choice has little effect on the overall picture of citation patterns. It's also worth noting that our journal selection procedure moves journals into economics if they cite core economics journals substantially more than they cite core journals in their initially assigned

[4]Medicine trunks are *Journal of the American Medical Association* and *New England Journal of Medicine*. The American Mathematical Society's leading research journal is relatively new, so the Math trunk is *Annals of Mathematics*, a historically important and leading journal published by Princeton University.

discipline. Disciplines that might be said to have emerged from economics, such as finance and accounting, lose journals to economics this way.[5]

We see our set of 16 non-economics disciplines as interesting, relevant, and suitable for bibliometric analysis. The social science discipline group consists of anthropology, political science, psychology, and sociology, in addition to economics. The discipline of history, also a social science, is omitted because it's less journal-based, relying more on books. The same consideration rules out humanities disciplines like literature and gender studies. It seems likely that humanities disciplines interact more with sociology than with economics, an important qualification to the results presented here.

In the universe of social science scholarship, books are more important to political science and sociology than to academic economics, so the results reported here should be understood as representative of journal-based social science scholarship alone. A focus on journals also allows us to rely on the *WoS*, which has good coverage of scholarly journal output published since the mid-20th Century. Journal-based classification also lends itself to impact-factor-type quality weighting of citation flows.

We are also interested in interactions with non-social science disciplines. Some of these disciplines, like statistics and marketing, have a long history of interaction with economics. Others, like math, medicine, and physics, cite social science rarely. These disciplines are included because we see evidence of increasing citation flows between them and the social sciences, while others that we've omitted, like chemistry, remain isolated from social science. We omit most engineering fields, though some of these, like civil engineering, interact with economics. Engineering disciplines rely heavily on conference proceedings and are therefore ill-suited to our journal-based classification scheme. On the other hand, our list includes two fields that might be considered engineering disciplines, operations research (OR) and computer science (CS), because these disciplines publish heavily in traditional journals as well as in conference proceedings.[6]

2.2 Data

The citation data analyzed here come from the *WoS*, with citing articles published between 1970-2015 and cited articles published since 1955. These restrictions are motivated by the fact that the *WoS* appears to be less complete and less accurate in earlier years. We match economics articles in the *WoS* data to more detailed bibliographic content found in the *EconLit* database, including abstracts and keywords. Importantly, *EconLit* includes *Journal of Economic Literature* (JEL) codes, which are used to classify papers into economics fields. The *WoS* indexes a much broader set of publications than does *EconLit*, including book reviews, conference notes, and editors' introductions. The cited economics sample used in the fields and styles sections is therefore smaller than the sample used for discipline level analyses. Since most of these additional publications neither cite nor are cited, their omission has little impact on our statistics. We

[5]Core journals are defined as the minimal set of journals that comprise at least 30% of citations from trunks. Online Appendix Table W1 lists the journals ultimately assigned to each discipline. A few journals remain unassigned and a handful of assignments are discretionary. See the appendix for details.

[6]The *WoS Conference Proceedings Citation Index* coverage of cited references begins only in 1999, a further consideration weighing against inclusion of most engineering disciplines. Along the same lines, it's also worth noting that the *WoS Book Citation Index* starts only in 2005.

matched 94% of *EconLit* articles to the *WoS* and use this matched sample to examine citation to economics fields and styles.[7]

Our sample covers a period of changing journal influence and prestige, and, especially, changing composition of journal output. We therefore allow for changes in a journal's intellectual importance using a weighting scheme that implicitly values citations by the importance of the *citing* journal, a kind of customized "impact factor." This produces a weighted citation rate along the lines of the impact-factor-weighted indices seen in popular scholarship rankings. As a robustness check, we consider two weighting schemes when looking at citations from economics.

2.3 Quantifying Extramural Influence: Conceptual Framework

Extramural influence is defined by citations from the journals of one discipline to journals from another. Citations from discipline d to discipline d' are measured using a weighted average that can be written:

$$s^t_{dd'} \equiv \sum_{\{j|D(j)=d\}} w^t_j s^t_{jd'}, \tag{1}$$

where $s^t_{jd'}$ is the fraction of year t citations in journal j (among citations for which we can identify the discipline of the cited reference) made to articles in the journals of discipline d'. The sum runs over all journals, indexed by j, classified as belonging to discipline d (a set denoted $\{j|D(j) = d\}$). The weights, w^t_j, emphasize journals that are more important to discipline d at time t. Specifically, the w^t_j are proportional to the number of citations from discipline d's trunk journal(s) in year t to journal j, rescaled so that in each year they sum to one across the journals in each discipline. The measure $s^t_{dd'}$ can be thought of as a citation *share*, showing, for example, the (weighted) fraction of citations in economics papers published in 1997 to articles in sociology journals.[8]

Some of the statistics discussed below characterize citations to groups of disciplines. The share of citations from discipline d to a group of disciplines, denoted by G, is described using the sum,

$$s^t_{dG} \equiv \sum_{d' \in G} s^t_{dd'}.$$

For example, the citation share from economics to the group of business disciplines is the sum of shares of economics cites to finance, marketing, management, and accounting. Most citations are within-discipline, so counts of extramural citations can be small, even when grouped. Our plots therefore show 5-year moving averages to smooth some of the resulting variation.

[7]The match rate is 71% with *WoS* publications as the denominator. Unmatched *WoS* items are mostly documents like book reviews, announcements, and problems (for teachers) that make and receive very few citations, though there are some differences in coverage, especially in the earlier years. See the appendix for details.

[8]Note that the journal-to-discipline shares, $s_{jd'}$, are defined so as to sum to one across all disciplines, indexed by d', for each journal j in each year t. Because the w^t_j sum to one across journals, the discipline-to-discipline shares, $s^t_{dd'}$, also sum to one across disciplines d' for each d and t. Note also that the weights w^t_j are a rescaling of journal j's citations from the trunk journal *in the discipline to which journal j belongs*. Subscript j therefore identifies the citing discipline as well as the citing journal.

Citations are interpreted here as a measure of influence. Some citations, of course, are critical, and so the citing article might be fairly described as rejecting or criticizing the content of the cited paper. But even negative citations reflect a measure of influence in the sense that the critically citing author finds the content or arguments of the paper being cited to be worth responding to by argument or with evidence. In any case, the bulk of citations from social science disciplines that might have a critical view of economics scholarship are to empirical papers. It seems likely that even critical references to empirical papers are reacting to rather than simply dismissing the cited work.[9]

3 Economics Insularity

In an earlier examination of citation data, Pieters and Baumgartner (2002) conclude that "no area of economics appears to build substantially on insights from its sister disciplines." Examining an interdisciplinary network derived from cross-journal citation data originally compiled by Leydesdorff (2004), Moody and Light (2006) find that several sociology journals are among the most central in the network. By contrast, political science, psychology, and especially economics journals occupy distinct, well-differentiated clusters, a marker for being more self-referential. More recently, Fourcade, Ollion and Algan (2015) compare citation rates between economics, sociology, and political science trunk journals (as defined here), arguing that economics is uniquely insular among social sciences. As a prelude to our evaluation of economics' extramural influence, we ask whether our data support this view. In addition to being of intrinsic interest, evidence of insularity helps calibrate differences in extramural citation flows. Sociology might cite economics more than political science does simply because sociology cites all other social sciences more than political science does.

The left panel of Figure 1, which compares extramural citation rates from each social science discipline to the group consisting of the other four, shows large differences in insularity across disciplines, as well as important changes in extramural citation rates. Political science is the most outward-looking, though political science's extramural citation rates were trending down through about 1990. Sociology is the second most outward-looking social science, with a mostly increasing extramural citation rate.

Economics is less outward-looking than sociology and political science, but not uniquely or irredeemably insular. Since around 1990, economics has paid more attention to other social sciences than has psychology. Moreover, economics' citation rates to other social sciences have been increasing for most of our sample period (though have leveled off recently).

Fourcade, Ollion and Algan (2015) note that economics' citations to finance rose in the 1980s and 1990s, a trend they see as sinister. The middle panel of Figure 1 compares citation shares going from each social science discipline to the four business disciplines in our sample (finance, accounting, marketing, and management). Not surprisingly, economics is indeed the most business-focused of the social sciences, with modest growth in the economics citation share to business since 1980. This pattern is consistent with the Fourcade analysis.[10]

[9] Lynn (2014) similarly gauges interdisciplinary impact using extramural and intradisciplinary citation flows, without regard to the tone of the underlying references.

[10] The largest recipient of economics citations among business disciplines is finance. In sociology the largest share goes to

But sociology is not far behind, and citations from sociology to business disciplines also mostly show modest growth over the same period. Since 2000, the strongest growth in extramural social science citations to business disciplines has been in citations from psychology. It also seems noteworthy that the economics citation share to business disciplines is not much larger than than the share of economics citations going to the social science group.

Social science citation shares to seven other disciplines are about on par with those to business. This can be seen in the third panel of Figure 1, which plots extramural citations shares from social science disciplines to operations research, statistics, CS, mathematics, physics, medicine, and public health; this group includes all of our remaining disciplines except multidisciplinary science.[11] Sociology has a modest lead here, but economics, which is roughly tied with psychology and anthropology, does not appear unusual among social sciences in its interactions with these disciplines.

While contradicting polemical claims of economics' unique insularity, Figure 1 also highlights the importance of context when comparing extramural citation rates. Our assessment of economics extramural influence relies primarily on relative citation rates; we compare for example, the extent to which sociology cites economics and political science, thereby holding fixed sociology's propensity to make extramural citations. Such comparisons are also made for changes and trends in extramural citations. Our analysis of the influence of economics fields and styles likewise relies on relative measures, comparing, say, the *share* of extramural citations received by macro and micro economics, and by theoretical and empirical economics.

4 Extramural Influence

Among the four non-economics social sciences, economics has the most influence on sociology and political science. These two historically outward-looking disciplines are also increasingly likely to cite us. This trend is documented in Figure 2, which plots extramural citation rates for five social sciences. Political science saw rapid growth in citations to economics in the 1970s and 1980s, with ups and downs around a modest upward trend thereafter. At the same time, citations from political science to sociology declined. Extramural citations from sociology to economics also rose steeply in the 1970s and early 1980s, overtaking the citation rate from sociology to psychology and political science. Sociology's attention to economics flattened in the 1990s, but has been trending up since the early 2000s.

Psychologists and anthropologists appear to read less economics than do sociologists and political scientists. Sociology has historically had more influence on these fields than we do, and psychology also has more influence than we do on anthropology. Yet economics' influence on psychology has recently accelerated, more than doubling since the early 2000s. This presumably reflects the influence of behavioral economics on both disciplines. The extramural citation rate from psychology to economics now roughly matches the

management. In psychology, the largest share goes to marketing.

[11]Multidisciplinary science is an exception to our rule that disciplines are defined by their journals. It seems interesting to ask what top multidisciplinary science journals are *citing*, as we do below. It's harder to interpret inbound citations to these journals as indicative of other disciplines' extramural interests, since the content of these journals comes almost exclusively from traditional disciplines.

corresponding rate to sociology (at a little over 1%).

The bottom panel of Figure 2 shows citations from economics to other social sciences. Political science emerged in the 1990s as the most influential social science for economics, now capturing about 2.5% of (weighted) economics citations. This puts economics in second place in citations to political science, behind sociology (with 5%). Economics citations to psychology and sociology have also grown since 1990, with both now running a little over one percent.

Not surprisingly, economics is widely read by scholars working in business-related disciplines, especially those in finance and accounting. This can be seen in Figure 3, which shows extramural citation rates from finance, accounting, marketing, and management.[12] Economics has long been the dominant social science influence on finance, garnering over 40% of finance citations in the 1970s. But the attention paid to economics in finance declined markedly in the 1980s, and has remaining at a lower level (just under 30%s) since.

Extramural citation rates from accounting show an up and down pattern. Economics and psychology both had dramatically increasing influence on accounting in the 1970s and 1980s. Economics' citation share from accounting peaked in the mid to late 1980s and has since fallen to about half of its peak. Psychology's influence fell off more steeply starting in the early 1980s, and psychology is now close to invisible in accounting journals, though it was once remarkably influential.

Management and marketing are more attentive to psychology than to economics, a gap that appears to have widened in the past 10 years, especially for marketing. But the gap in extramural citation rates from management to economics and sociology narrowed in the 1980s. Social sciences other than psychology and economics receive little attention from scholars publishing in business disciplines, with the exception of management, which also cites sociology.

The four mathematically oriented disciplines covered by our analysis are operations research (OR), statistics, computer science (CS), and mathematics. We expect OR, which emphasizes optimization, and statistics, which overlaps with econometrics, to pay much more attention to economics than to other social sciences. This is borne out by Figure 4, which plots extramural citation rates for math disciplines. Moreover, the share of OR's cites going to economics has roughly doubling since the late 1990s, cresting recently at around 13%. After declining in the 1970s and 1980s, the importance of economics (mostly econometrics, as we show below) to statistics has also increased since about 1990.

Economics has had a much lower level of influence on computer science and mathematics than on OR and statistics. Interestingly, however, citations from computer science to economics have grown from a vanishingly small share before 1990 to claim about 1% in the 2000s. Also noteworthy is the fact that citations from CS to economics caught up with those from CS to psychology in the mid-2000s. The citation rate from math to economics is very low, and annual citation shares from math are noisy and heavily influenced by a few citing articles. Here too, however, there are signs of growing (though still small) influence since 1990.[13]

[12] We elected to start the management series in 1980 because one of the management trunks (Academy of Management Review) started in 1976 and is not indexed by the Web of Science before 1983, while indexing of the other (Academy of Management Journal) appears to be substantially incomplete until the mid-seventies.

[13] The table-top-shaped spike around 2003 is an artifact of our use of a five-year moving average and the 2003 publication by the highly ranked *Bulletin of the American Mathematical Society* of an article, Hofbauer and Sigmund (2003), which cites

Extramural citation rates from other disciplines are plotted in Figure 5. This group includes multidisciplinary science, public health, medicine, and physics. Psychology takes pride of place for extramural citation rates from three of four of these disciplines. Economics emerges with a small but growing share of extramural cites in this group starting around 2000, taking second place in three cases, and bypassing sociology for extramural citation rates from medicine after 2000. Although citation rates to economics remain low in these four disciplines, the attention they pay to our scholarship is significant by historical standards and in comparison with anthropology and political science. Rising from virtual invisibility, economics now garners 2% of citations from public health and almost 1% from multidisciplinary science.

These comparisons show economics to be the most widely cited social science in 7 of the 16 disciplines we examined and is now essentially tied for first in two more. Sociology is roughly comparably influential in the social sciences. Outside the social sciences, psychology is our main competitor for extramural influence. In particular, psychology is the most influential social science in marketing and management and also ahead of economics in our "other sciences" discipline group.

Interest in economics also appears to be growing in many disciplines and has surged recently in some. Even among scholars who have historically read no economics, interest has ticked up. This weighs against narratives describing economic scholarship as narrow or captured by narrow special interests. Of course, as noted above, the analysis here covers journal scholarship only, and does not isolate critical or negative citations from the rest. Likewise, our findings do not shed light on interactions with the humanities.

We turn next to a finer-grained analysis of the sources of economics' extramural impact. This deeper investigation classifies the economics papers that have been influential in each discipline into economics fields (like macroeconomics and labor economics) and research styles, distinguishing empirical and theoretical work.

5 Sources of Impact: Economics Fields

5.1 Defining Fields

As with our analysis of extramural citation rates by discipline, the first step in an investigation of citations to fields is classification. We classified cited economics articles into fields using information in article titles, keywords, and JEL codes. Because the *Web of Science* omits article keywords and JEL codes, the field analysis looks only at articles matched to the AEA's *Econlit* database, which provides JEL codes and keywords for each paper. *EconLit* started in the 1960s but coverage seems patchy in the first few years, so fields are classified for articles published since 1970. Because citations are necessarily backward-looking and it takes time for citation patterns to emerge, the universe of citing articles in our study of fields includes papers published since 1980.

Our field classification scheme exploits three types of information: the JEL codes chosen by authors; words appearing in titles and keywords; and the JEL codes of the articles in a paper's reference list. We process this information in two steps. The first uses articles' JEL codes, titles, and keywords as inputs to

many game theory papers published in economics journals.

a random forest algorithm that assigns papers to one of the seventeen economics fields defined in Ellison (2002). The second step applies a clustering algorithm to boil these 17 "initial fields" down to a group of 9.

Our 9 final fields are development, econometrics, industrial organization (IO), international, labor, macroeconomics, microeconomics, public finance, and a group of miscellaneous smaller fields. Note that—as with John Pencavel's 1988 revision of JEL codes—our field taxonomy is meant to distinguish between substantive areas of economic research, such as between inquiries related to product market structure, labor market behavior, taxation, and business cycles.[14] The distinction between research styles, that is differences between purely theoretical and empirical work, is tackled separately. The "miscellaneous" field includes most of the papers classified initially as economic history, environment, (lab) experiments, finance, law and economics, political economy, productivity, and urban economics, as well as papers that were simply hard to classify.[15] The papers in the miscellaneous category are about two-thirds empirical (interactions between fields and research styles are discussed below).

The challenge of classifying *Econlit*'s 140,000 or so papers into fields is aggravated by the fact that *Econlit* lists several JEL codes for most papers. The codes for any one paper are often diverse, pointing to a set of very different fields. Although some articles in *Econlit* are indexed with JEL codes in an informative order reflecting the authors' judgment of their relevance, papers published since 2004 are mostly indexed with codes in alphabetical order. We therefore constructed a large training data set containing papers whose JEL code order appears to be informative, supplemented with papers classified by hand.[16] The training data were used to train a random forest algorithm to classify papers as a function of fields associated with unordered JEL codes, (words in) titles, and keywords. This step classifies papers into the 17 fields defined by Ellison (2002).

The second field classification step uses k-means clustering to produce a set of 9 "final fields". The clustering algorithm looks at each article's initial field and the initial fields of papers on its reference list. In a random sample of 100 articles, the results of our machine learning classification scheme match those from one of two human raters about 74% of the time. It should be noted, however, that the human raters themselves agree on article fields only about 76% of the time. Other details related to field classification appear in Appendix B.

5.2 Economics Intramurals

Output by Field

To put extramural citation patterns to fields in context, we look first at economics field output and the economics discipline's own citation distribution over fields. Figure 6 traces the evolution of economics

[14] Cherrier (2017) outlines the history of JEL codes.

[15] Note that "finance" appears both as a miscellaneous field within economics and as a non-economics discipline. The distinction here is based on journals. For example, articles published in the *Journal of Finance*, whatever the topic, belong to the finance discipline, while papers on corporate finance in the *Quarterly Journal of Economics* mostly end up in the miscellaneous field in the economics discipline.

[16] The latter includes papers in field journals clearly associated with a single field, e.g. *The Journal of Labor Economics*. For purposes of training, JEL codes ordered informatively were given field labels following the algorithm used in Ellison (2002).

journal *output* by field for the period 1970-2015 using three weighting schemes. The unweighted share of articles published in field f in year t, reported in the left panel of Figure 6, is defined as $n_f^t / \sum_{f'} n_{f'}^t$ where n_f^t is the number of *WoS* papers matched to *Econlit* that are classified as belonging to field f, and published in journals on our economics journal list in year t. These shares are easily interpreted, but may be sensitive to the selection of journals on the journal list.

The middle panel, labeled "AER weighted," reports field shares using trunk journal importance weights, as defined in Section 2.3. The AER is the economics trunk journal. Specifically, weighted economics journal output share is computed as

$$\tilde{m}_f^t \equiv \frac{m_f^t}{\sum_{f'} m_{f'}^t},$$

with weighted publication shares by field, denoted m_f^t, defined by $m_f^t \equiv \sum_{\{j|D(j)=\text{econ}\}} w_j^t m_{jf}^t$, where w_j^t is the share of the AER's year-t references to economics journal j and m_{jf}^t is the fraction of papers published in journal j in year t classified in field f. This measure captures the relative prevalence of fields among papers published in year t in journals that are cited heavily by the AER.

Papers published in *Econometrica* and other journals may, of course, distribute citations over journals differently than does the AER. To construct a picture of economics journal output that does not privilege the AER, the right panel of Figure 6, labeled "Top 6 weighted," uses a set of broader journal weights discussed in Angrist et al. (2017). These are derived from the citation behavior of a set of six top journals, which includes the usual top 5 plus *The Review of Economics and Statistics* (once cited as often as *The Quarterly Journal of Economics*). Top-6 weights are year-specific and come from applying the Google Page Rank algorithm to the matrix of cross-citations between these six journals.[17] Top-6 weighting emphasizes papers in journals cited heavily by *Econometrica* as well as the AER, thereby weighting more technical articles more heavily than the AER-only weighting scheme.

Unweighted shares show microeconomics to be the field that has grown the most over the past 30 or so years, with a publication share that roughly doubled since 1990 and is now around 17%. This growth partly reflects the proliferation of microeconomic theory journals and their expansion. For example, *Games and Economic Behavior* started in 1989 (indexed by WoS from 1991); *Economic Theory* started in 1991 (indexed by WoS from 1995), and the number of papers appearing in the *Journal of Economic Theory* more than doubled between 1980 and 2014.[18] The increase in microeconomics as a share of top journal

[17]Formally, let A^t be the 6×6 matrix with entries A_{kj}^t equal to the fraction of journal j's citations to all top six journals in year t made to journal k; and let μ^t be the solution to $\mu^t = dA^t\mu^t + \frac{1-d}{6}\mathbf{1}$, i.e. $\mu^t = (I - dA^t)^{-1}\frac{1-d}{6}\mathbf{1}$, where $d = 0.85$. We set $\tilde{w}_j^t \equiv \sum_k \mu_k^t s_{kj}^t$, where the sum is taken over the top six journals k, and s_{kj}^t is the number of citations from journal k to journal j in year t as a fraction of all year t citations from journal k to journals in our full economics list. The final top-6 weighting series, denoted w_j^{T6t}, is the five-year moving averages of the $\tilde{w}_j^t$. Figure 1 in Angrist et al. (2017) plots these weights. The right panel in Figure 6 plots weighted shares computed as

$$\tilde{m}_f^{T6t} \equiv \frac{m_f^{T6t}}{\sum_{f'} m_{f'}^{T6t}},$$

where $m_{f'}^{T6t} \equiv \sum w_j^{T6t} m_{jf}^t$.

[18]Kelly and Bruestle (2011) find that increasing the number of a field's specialty journals indeed increases that field's publication shares, though Card and DellaVigna (2013) note that citation rates to fields also change independently of the number of papers published.

publications depends on the weighting scheme used to measure this growth. The AER-weighted series portrays microeconomics as growing by about 50% over the past 35 years and only recently becoming the largest field. Under top-6 weighting, microeconomics has long been dominant, increasing primarily n the 1980s and accounting for for a little over 20% of weighted publication output ever since.

Another notable feature of the field distribution in journal output is the sharp increase in the weighted share of papers in development since about 2000. Weighted measures show development pulling ahead of IO around 2010. By contrast, labor economics and IO have suffered clear declines in weighted publication shares, falling markedly from late-80s peaks. It's also interesting that macroeconomics and the miscellaneous category have been in the top 3 since around 1990, though these fields' output shares seem to have peaked around 2000.

Field Citation Shares

We report citation rates to economics fields using importance-weighted measures analogous to those used to estimate extramural citation rates to the economics discipline as a whole. In this case, however, citation rates are constructed so as to normalize the overall size of the citing reference distribution. In particular, we graph the (weighted) shares of the citation distribution garnered by labor, international, and so on, using measures that sum to one across fields (likewise, for styles, as discussed below). These shares are constructed the same way for economics' own field citations and for those of other disciplines and are therefore explained here.

Let s^t_{jf} be the fraction of journal j's year t citations made to papers in economics field f. The extramural influence of field f on discipline d is built up from a weighted average of journal-specific citation rates across the set of journals in discipline d:

$$s^t_{df} \equiv \sum_{\{j|D(j)=d\}} w^t_j s^t_{jf}. \tag{2}$$

The sum over fields of the s^t_{df} equals the weighted share of citations in discipline d going to economics papers matched to *Econlit* and classified into fields.[19] The first set of weights (w^t_j) used in this formula is the same as that used for extramural citation rates by discipline; these weights are proportional to journal j's share of all year t citations from discipline d's trunk journal to journals in that discipline (these sum to one over journals in the discipline). We also report citations from economics journals to economics fields using the top-6 weighting scheme described above and in Angrist et al. (2017).

Our investigation of extramural citations to economics fields includes citation rates for groups of related disciplines, like the social sciences and the group of business disciplines. These are computed as unweighted

[19] The denominator for s^t_{jf} is the *WoS* count of journal j's cites in year t. The sum across fields of journal j's cites to fields f, that is, the sum of s^t_{jf} over f, may be less than the share of journal j's cites to economics as whole ($s^t_{j,\text{econ}}$) because the data underlying our analysis of cites to entire disciplines includes articles without JEL codes and articles not matched to *Econlit*.

averages of the discipline-level shares, s^t_{df}, across the disciplines in the group:

$$s^t_{Gf} \equiv \frac{1}{|G|} \sum_{d \in G} s^t_{df}. \tag{3}$$

Note that unweighted averages of this type are affected little by citation patterns in disciplines in which economics gets a small share of citations.

Finally, because we're interested in the *relative* importance of economics fields, rather than how often sociology, say, cites labor economics, the field-level measures described by equations (2) and (3) are normalized to sum to one over fields. It's these field *shares* that appear in our figures. Specifically, we gauge the relative extramural influence of economics fields using

$$\tilde{s}^t_{df} \equiv \frac{s^t_{df}}{\sum_{f'} s^t_{df'}}, \tag{4}$$

where f' indexes fields in the normalizing sum in the denominator. These normalized shares sum to one across fields by construction. Plots of the extramural influence of economics fields on discipline groups similarly show normalized group-level shares,

$$\tilde{s}^t_{Gf} \equiv \frac{s^t_{Gf}}{\sum_{f'} s^t_{Gf'}}. \tag{5}$$

As with the formula in (3), this quantity is most affected by the disciplines in group G that cite economics most heavily.

As a benchmark for the distribution of extramural citation shares to fields, Figure 7 reports the field distribution of economics intramural citation shares. This figure plots $\tilde{s}^t_{\text{econ},f}$ as defined in equation (4), with s^t_{df} computed using AER and top-6 weights. The two weighting schemes generate a broadly similar picture of the current distribution of field influence. We see, for example, that microeconomics has the largest weighted citation share, macroeconomics is roughly tied with the group of miscellaneous fields for second place, labor economics is fourth, and, at the other end, international economics and development are the least cited fields.

Trends in intramural field influence are affected somewhat by the choice of weights. The AER-weighted series suggests microeconomics became increasingly influential from the early 1980's through the late 2000s, while top-6 weighting paints a picture in which the microeconomics citation share is larger in general, but peaked in the early 1990s. Viewed through the lens of either weighting scheme, the collection of miscellaneous fields appears to have become increasingly influential. Development has also become markedly more influential in the past ten years, while the AER-weighted series shows macroeconomics, labor economics, and industrial organization with generally declining citation shares. Top-6 weighting moderates declining citations to labor and macro, but the decline in citations to IO remains pronounced using either weighting scheme. Citations to econometrics rise and fall in both versions, peaking earlier in the AER-weighted series.

5.3 Extramural Influence by Field

Different disciplines find different parts of economics relevant or useful. This is apparent in Figures 8 and 9, which plot field shares (formula (5)) for four discipline groups. Specifically, these figures show trunk-journal-weighted citations to the five most highly cited fields plus other fields for which the average extramural citation rate exceeds 5%.

As can be seen in the left panel of Figure 8, social science disciplines (political science, sociology, anthropology, and psychology) cite labor, microeconomics, and the group of miscellaneous fields most heavily, but social scientists also reference macro and econometrics, and, increasingly, development and public finance. Social scientists' citation to the group of miscellaneous fields have also increased markedly since the mid-1990s, while the citation share going to microeconomics has fallen. Increased citations to the miscellaneous group reflect, in part, increased citations to political economy.

Not surprisingly, the group of business disciplines (finance, accounting, marketing, and management) cite miscellaneous fields heavily, since the latter includes finance papers published in economics journals. This can be seen in the right panel of Figure 8. Notably, however, the bulk of extramural citations from business disciplines is to papers from other parts of economics, with substantial citation shares going to microeconomics, IO, macroeconomics, and econometrics. It is also interesting to note that the share of business-discipline citations going to IO, which increased considerably in the 1980s and early 1990s, has since fallen to a little over half its late-1990s peak. On the other hand, following a modest decline in the 1980s, the share of business-discipline citations going to macroeconomics has been increasing for the past twenty years.

Mathematical disciplines (operations research, statistics, computer science, and mathematics) increasingly cite microeconomics, a strong trend visible in the left panel of Figure 9. In fact, micro has recently displaced econometrics as the most cited field for this group of disciplines. We see especially steep growth in micro cites after 2000, a period in which the influence of economics as a whole on mathematical disciplines has been increasing. Until recently, industrial organization was the third most influential economics field in the math discipline group, but the IO citation share has plummeted since about 2008.

The right panel Figure 9 traces field influences on the discipline group containing public health, medicine, physics, and multidisciplinary science journals. As in the plot for math disciplines, this figure shows evidence of an interesting swap. Here, public finance replaces labor as the most cited field in the mid-1990s. This probably reflects the growing importance of health economics within the larger public finance field, as well as health-related disciplines' growing interest in econometric methods. But there also seems to have been a secular decline in this discipline group's interest in labor, even as the attention paid to other areas, including micro and development, was on the upswing.

A more detailed picture of field influence emerges from an examination of specific disciplines. Figure 10 presents citation data for the four disciplines in which economics looms especially large, in the sense of claiming a 10% or higher citation share recently. These "Group A disciplines" include finance, accounting, OR, and political science. Not surprisingly, the miscellaneous field, which includes finance, garners a large share of cites from the finance discipline. Macroeconomics' citation share has been rising since the late 90s and macro is now the second-most important influence on finance with a share recently approaching one-quarter.

Other panels in this figure show how the diversity of economics contributes to extramural influence. The marked increase in the microeconomics share of OR citations since 2000 coincides with the substantial increase in overall in citations to economics from OR, suggesting interest in micro is driving this growth. At the same time, microeconomics' influence on accounting and political science has fallen. Citations from political science to microeconomics appear to have been replaced by citations to the miscellaneous group of fields, which includes political economy, and, recently, to development economics. While political scientists' overall interest in economics has increased (a pattern documented in Figure 2), the economics fields capturing political scientists' attention have shifted.

Figure 11 reports field citation shares for four other disciplines in which economics is influential (sociology, statistics, marketing and management). Economics recently gets 5-10% of citations from these "Group B disciplines." Labor economics has been and remains the dominant field influencing sociology. As expected, statistics is most influenced by the econometrics field, which receives a large and steadily increasing share of extramural cites from this discipline. IO has long been the dominant influence on marketing, although marketing cites to IO have fallen steeply since 2000, with some substitution towards micro and econometrics. IO also has the largest citation share in management for much of the sample period, but again this share has fallen since 2000. Citations from management to labor fell dramatically in the 1980s and 1990s. It should be noted, however, that early 80s trends may be influenced by the absence of one of our management trunk journals in this period.

Finally, Figure 12 describes the field interests of five disciplines where economics is not (yet) highly influential, but where the share of citations to economics has more than doubled over the past 25 years (these "Group C" disciplines are CS, psychology, public health, medicine, and multidisciplinary science).[20] The surge in citations from CS to economics is attributable to growing interest in microeconomics (with cites going mostly to papers on game theory and mechanism design). Psychologists are mostly attentive to micro as well, especially since the 1990s.

In public health and medicine, increasing interest in economics is driven mostly by citations to public finance, which includes health economics. Consistent with Figure 9, public finance appears to have replaced labor as the main recipient of extramural citations from health-related disciplines outside economics. Also noteworthy is the fact that, starting from zero in 1980, econometrics has in recent decades garnered over 10% of these two disciplines' extramural citations to economics. Development has similarly emerged from virtual invisibility in 1980 to claim a significant share of public health references. The data for multidisciplinary science are especially noisy, reflecting the overall low citation rates to economics from this area. Still, Figures 5 and 12 suggest an uptick in multidisciplinary science journals' interest in economics, driven by references to

[20] Math and physics, whose citations to economics have also grown, are omitted because these disciplines' level of interest in economics remains substantially below that of disciplines included in Group C.

microeconomics and the group of miscellaneous fields (which includes political economy and lab experiments).

Importantly, no single field appears to monopolize economics' extramural influence. In the business disciplines for which economics has long been important, finance is most influential, but other fields also get attention. A few disciplines focus on a particular field, but the subjects of this focus are diverse: econometrics is read in statistics, labor in sociology, microeconomics in computer science and OR, and IO in marketing and management. Political science now focuses on the fields we've grouped in the miscellaneous category. The recent growth in references from public health and multidisciplinary sciences is driven by papers in public finance, while recent growth in interest in psychology and computer science is directed towards microeconomics. Far from being a monolithic structure dominated by finance, economics has long been and remains a diverse and evolving enterprise. Economics' extramural influence reflects this dynamic diversity.

6 Sources of Impact: Empirical vs Theoretical Economics

Empirical economics has flowered in recent decades, a development documented by Panhans and Singleton (2015) and Hamermesh (2013, 2018), among others. Angrist and Pischke (2010) argued that this change in economics research style reflects the proliferation of "design-based" empirical methods that yield more credible results than did earlier empirical economic research. This argument motivates us to evaluate the changing role of empirical work in the growth of economics' extramural influence. As in our classification of economics articles into fields, articles are classified into research styles using data from *Econlit*. The cited paper universe again contains papers published since 1970, though style classification is more accurate for articles published since the mid-1980s when *Econlit* began to include abstracts. Following a report on the style distribution of economics publications since 1980, therefore, the influence of economic styles is examined for citing papers published since 1990.

6.1 Classifying Economics Research Styles

We used machine learning techniques to classify papers published in the journals on our economics journal list as empirical or theoretical. This classification aims to distinguish research that produces data-based estimates of economically meaningful parameters from economic research of a purely theoretical or methodological nature. Papers that address methodological or theoretical issues while also producing estimates that might be seen as substantively meaningful were mostly classified as empirical. On the other hand, because methodological econometric research seems distinct from both economic theory and empirical work, papers classified in the econometrics field (using the process described above) were classified as falling into a distinct econometrics style category. Our style analysis therefore distinguishes papers in three categories: empirical, theoretical, and econometrics.

The machine learning algorithm used here starts with a training sample of 5,469 English-language papers, of which 1,503 were hand-classified by Ellison (2002). We updated the Ellison (2002) training sample by sampling from top journals and by drawing a random sample from all journals on the economics journal list. These additional training papers were classified by our (trained) research assistants. The goal was to classify papers as empirical if they report econometric estimates of substantive interest, constructed using real world data (as opposed to made-up or simulated data). Remaining papers are classified as theoretical unless previously classified in the econometrics field.

We fit a dummy variable indicating empirical papers using a logistic ridge algorithm that takes as predictors JEL codes and keywords, article titles, and features of abstracts (where available). Although JEL codes are topic-based rather than style-based, in practice they help predict style. Other predictive article features include the (initial) economics field coded earlier and publication decade. In a random sample of 100 articles, the classification algorithm predicts the style classifications made by two raters about 80% of the time (the raters themselves agree on style 82% of the time). The style classification process is detailed in the appendix.

A hint of the dramatic change in economics research styles over the past half century emerges from Appendix Table A2, which lists the top 10 most cited papers published in each decade since 1970, along with their fields and styles. This table also gives a sense of how our style classification strategy works in practice.[21] Kahneman and Tversky (1979), classified as theoretical, tops the 1970s list. Heckman (1979) and Hausman (1978), classified as econometrics, come next. Hall (1978) is the most highly cited empirical paper of this era, and the only paper classified as empirical to make either the 1970s or 1980s top ten. By contrast, the 1990s top ten list includes three empirical papers: Katz and Murphy (1992); Berry, Levinsohn and Pakes (1995); and Hall and Jones (1999). After 2000, empirical papers surge ahead, with six empirical in the top ten lists for both the 2000s and the 2010s. Note also that our algorithm classifies papers like Eaton and Kortum (2002) and Christiano, Eichenbaum and Evans (2005) as empirical, even though they combine theory with empirical work.

The distribution of articles reported in Table 1 shows a strong interaction between fields and styles. Specifically, Table 1 cross tabulates the field-by-style distribution of the roughly 137,000 economics papers published since 1970 found in both *Econlit* and the *Web of Science*. This is the set of papers used for our analysis of citations to economics fields and styles. Papers in the microeconomics field are mostly (though not entirely) classified as theoretical, while papers in what are now often thought of as "applied micro" fields (labor, development, and public finance) are mostly empirical. On the other hand, papers in IO, also an applied micro field, tilt towards theory. Macro and international are about evenly split. Smaller fields grouped under the miscellaneous heading (environmental, finance, lab experiments, history, law and economics, political economy, productivity, urban, and unclassified) cover a set of papers that is two thirds empirical.

[21] Papers in this table are ranked by the average-over-post-publication-years of the top-6 weighted share of all citations each article receives from the journals on our economics journal list (this is the weighted citation rate, c_i, defined in equation (6), below, divided by 2015 minus the publication year to produce an annualized measure.)

6.2 Intramural Style Changes

As in the discussion of fields, we begin the exploration of styles with an intramural benchmark, looking at the style distribution of economics publications and citations. The citation rates used to trace both the intramural and extramural influence of styles are constructed like those for fields, modified here by replacing citation rates for nine fields in formulas (2) and (3) with analogous rates for three research styles. Also paralleling the analysis of fields, we focus on the normalized share distribution over styles, computed as in (4) and (5); normalized style shares sum to one over styles. Economics journal output and intramural citations to styles are weighted to reflect journal importance using top-6 as well as AER (trunk journal) weights. Extramural citations to styles are computed using trunk journal weights only.

Figure 13 traces the style mix of economics journal output since 1980. Unweighted publication counts, plotted in the left panel, show that the empirical share in economics output has increased from about 50% to about 60%, with the increase coming mostly since 2000. We also see some growth in econometrics publications. The trend towards empirical output becomes more pronounced and starts earlier when tabulated using AER weights than when unweighted. Top 6 weighting yields an even larger increase in the empirical share of influential journal publications, from a low of just over one-third in the mid-1980s to around 56% today. This reflects both within-journal increases in the proportion of empirical papers published and increased top-6 weighting of more applied journals.

Paralleling the increase in empirical output, weighted empirical citation shares, plotted in Figure 14, show strong and steady growth since 1990. The AER weighted series, plotted in the left panel of the figure, shows an empirical share increasing from just under 30% in 1990 to 50% in 2015. The top-6 weighted series traces a slightly more modest increase from a moderately higher base. Both weighting schemes show declining citations to econometrics, as well as to theoretical work.

Are individual empirical papers increasingly cited, or are there just more of them? Figure 14 shows smooth and steady growth in empirical shares since 1990, with a ratio of the share empirical at the endpoints exceeding 1.7 using AER weights and close to 1.6 using top-6 weights. Citation growth in the first case exceeds the corresponding change in the ratio of empirical to theoretical output, but the second is about the same.

A regression analysis of citations per article isolates dimensions of increasing empirical impact. We measure citations to individual papers using an AER-weighted citation measure for individual papers similar to those used for fields and styles. This measure is

$$c_i \equiv \sum_t \sum_{\{j|D(j)=\text{econ}\}} w_j^t s_{ji}^t, \tag{6}$$

where s_{ji}^t is the share of journal j's year t citations made to paper i.[22]

Using Poisson regression, the conditional mean of c_i is modeled as a time-varying exponential function of style dummies (EMP_i and MET_i), a vector of article-level covariates (denoted X_i), and, in some speci-

[22] As in the analysis of economics' citations to fields and styles, the denominator here is the set of papers in journals on the economics journal list. Consequently, shares do not sum to one in the sample of papers classified into fields and styles.

fications, a battery of year-specific field and journal indicators, indexed by $f(i)$ and $j(i)$. Baseline controls include a cubic polynomial in article page length and indicator variables for the number of authors. The model of interest can be written for year t as

$$E\left[c_i|X_i, EMP_i, MET_i, f(i), j(i), t(i)\right] = e^{\beta_1^t EMP_i + \beta_2^t MET_i + \beta_3^t X_i + \delta_{j(i)}^t + \gamma_{f(i)}^t}. \tag{7}$$

Because many papers are never cited and the citation distribution is highly skewed, this exponential model fits the conditional mean function of interest better than a linear model (37% of the papers in the sample are never cited by other papers in the sample). The coefficient β_1^t captures a time-varying covariate-adjusted log ratio of empirical to theoretical citations per paper.[23]

Theoretical articles published in the 1980s and 1990s were cited far more often than empirical work of the same period. This can be seen in panel A of Figure 15, which plots the time series of estimates of β_1^t from a model omitting field and journal effects. Starting from around $-.7$, the empirical citation deficit began to shrink in the late 1980s, and by around 1995 citation rates to empirical papers had attained a rough parity with citation rates for theoretical work.

Estimates of β_1^t in a model with field and journal controls, reported in panel B of Figure 15, show that some of the early theoretical citation advantage can be attributed to differences in the distribution of paper styles across fields and journals. In particular, theoretical papers have tended to appear in more highly cited journals. Controlling for field and journal dummies—that is, looking within fields and journals—the empirical citation deficit shrinks to around $-.5$ in the 1980s and essentially disappears in the early 1990s. Since around 2000, empirical papers have been cited more often that theoretical papers in the same field, published in the same journal and year. The increasing attention paid by the economics discipline to empirical work therefore reflects more than improved journal placement.

6.3 Extramural Influence by Style

The empirical share of extramural social science citations has grown steadily since around 2000, with nearly 70% of references from non-economics social sciences going to empirical work by 2015. This can be seen in the left panel of Figure 16, which, like Figure 8, describes (trunk-journal weighted) economics citations from non-economics social sciences as a group. The right panel of Figure 16 plots citations from business-related disciplines. These disciplines cite empirical economics in a proportion similar to that for economics itself, and also at an increasing rate. Interestingly, growth in the share of social science and business disciplines' cites to empirical papers seems to have lagged growth in the empirical share for economics.[24]

The left side of Figure 17 suggests that mathematical disciplines are more heavily influenced by theoretical and econometric papers than by empirical papers. In recent years, about half of extramural citations from

[23] Each observation in the sample used to estimate β_1^t is one economics paper (n= 137,162). Observations are unweighted. Regressions are run separately for each publication year. See Angrist et al. (2017) for estimates of a model like equation (7) using top-6 weighted shares; these are similar to those reported here.

[24] The online appendix includes a list of most cited papers for groups of non-economics disciplines, reported in a format analogous to that of Appendix Table A2.

this discipline group have been to theoretical papers and about a quarter to econometrics. Yet the empirical share of citations from math disciplines has increased modestly, from about 20% to 27% over the sample period. By contrast, the theoretical share has nearly held steady, so the shift towards empirical work has been mostly at the expense of econometrics.

Most extramural citations from our "other sciences" discipline group go to empirical work, a pattern documented in the in the right panel of Figure 17. Results for this group primarily reflect citation patterns in public health and multidisciplinary science (since these cite economics much more often than do other disciplines in the group). The empirical share for other-science citations runs in the mid-70s at the beginning and end of our sample period, while the citation share from this discipline group to econometrics holds steady at around 10%. We also see a modest shift towards theory in the late 1990s and early 2000s, but the expansion in references to theory faded in the late 2000s.

Figure 18 looks at citation shares to styles from the individual disciplines where economics is most influential (we dubbed these "Group A" disciplines in the discussion of fields). Finance and accounting now cite empirical work in about the same share as does economics. Accounting was considerably more theory-influenced in the early 1990s, so the shift toward empirical work is larger here. OR remains heavily influenced by economic theory, but we also see a modest increase in OR citations to empirical work. Like finance and accounting, political science has moved decisively to favor empirical papers, with an empirical citation share increasing from around 40% in the 1990s to over 60% in 2015.

The style story for disciplines where economics has somewhat less influence is more mixed (these disciplines were labelled "Group B" in the fields discussion). As can be seen in Figure 19, sociology has long focused on empirical work. But the empirical share in extramural citations from sociology increased steadily after the early 1990s, so that 80% of sociology references now go to empirical papers. At the same time, sociology's early emphasis on empirical economics suggests the discipline's engagement with economics has long been substantive as well as critical of economists emphasis of rational choice in human behavior.

Figure 19 also documents statistics' long-standing and growing interest in econometrics, a result discussed above in the context of Figure 11 for fields. The same figure shows an empirical share in citations from marketing which ends up below that in economics, while management directs more attention to empirical work. Extramural citations from marketing tilt more towards empirical work at the end of the sample period than at the beginning, but the changes here aren't dramatic. Starting from a very low base in 1990, Management has recently begun to reference econometrics.

The changing mix of research styles cited among "Group C" disciplines where economics' influence is growing is documented in Figure 20. Bucking the trend towards empirical work in extramural citations from other disciplines, growth in extramural citations from computer science is decidedly attributable to increasing references to theory. By contrast, psychology's accelerating interest in economics seems to reflect increased interest in empirical work. Since around 2000, citations from multidisciplinary science have also increasingly tilted empirical. The extramural citation share going to empirical papers has crossed the 50% line for both of these disciplines. At the same time, medicine and public health have long favored applied economics; this empirical emphasis is unchanged.

Figure 21 concludes our extramural investigation with a per-paper analysis in the mode of Figure 15, turning here to the changing style preferences of disciplines outside economics. Panel A of Figure 21 reports the empirical effect in weighted citation rates from the group of non-economics social sciences. Similarly, Panel B bundles non-social science disciplines. The regressions generating these estimates include field and journal controls, as in Panel B of Figure 15.[25] The resulting estimates suggest an early-80s empirical citation penalty, smaller and more noisily estimated for the social sciences than for economics. The corresponding estimates are larger but also noisy for non-social science disciplines. In the results from both discipline groups, however, the empirical disadvantage becomes a substantial and enduring empirical premium by the late 1990s.

Just as economists have moved to read and reference more of their own empirical scholarship, so too have most of the outsiders who follow our work. The shift towards empirical work in extramural citations per paper seem to have came around the same time as the shift in per-paper citations from economics. This timing is consistent with the Angrist and Pischke (2010) claim that empirical economics has evolved since the 1980s to be more credible and increasingly worth attending to. But economic theory remains important both inside and outside economics. The theory share in citations from economics today runs around 40%, and the group of mathematical disciplines still cites theory more than empirical work.

7 Summary and Conclusions

Is economics scholarship an inside job or an enterprise with deep impact? The value of any scholarly enterprise is necessarily subjective, and a discipline's practitioners may provide a biased view. We find it significant, therefore, that many sophisticated non-economists find economic scholarship to be worth referencing. Economics is the most influential social science in 7 out of the 16 extramural disciplines we've examined, and we're recently tied for first in two more (psychology and CS). And in many disciplines, our extramural influence is growing; *only* in business disciplines has our extramural impact fallen. Some disciplines' growing interest in economics started in the 1980s, while for others the increase is more recent. We've also seen that a variety of fields contribute to economics' extramural influence.

Consistent with the "credibility revolution" hypothesis advanced by Angrist and Pischke (2010), empirical work has been drawing a greater share of attention from most of the disciplines where economics is important. This mirrors the growing importance of empirical work within economics, a sustained shift that is visible within fields. At the same time, theoretical scholarship retains a large—and in the case of Computer Science, growing—share of our extramural readership from more mathematically oriented disciplines.

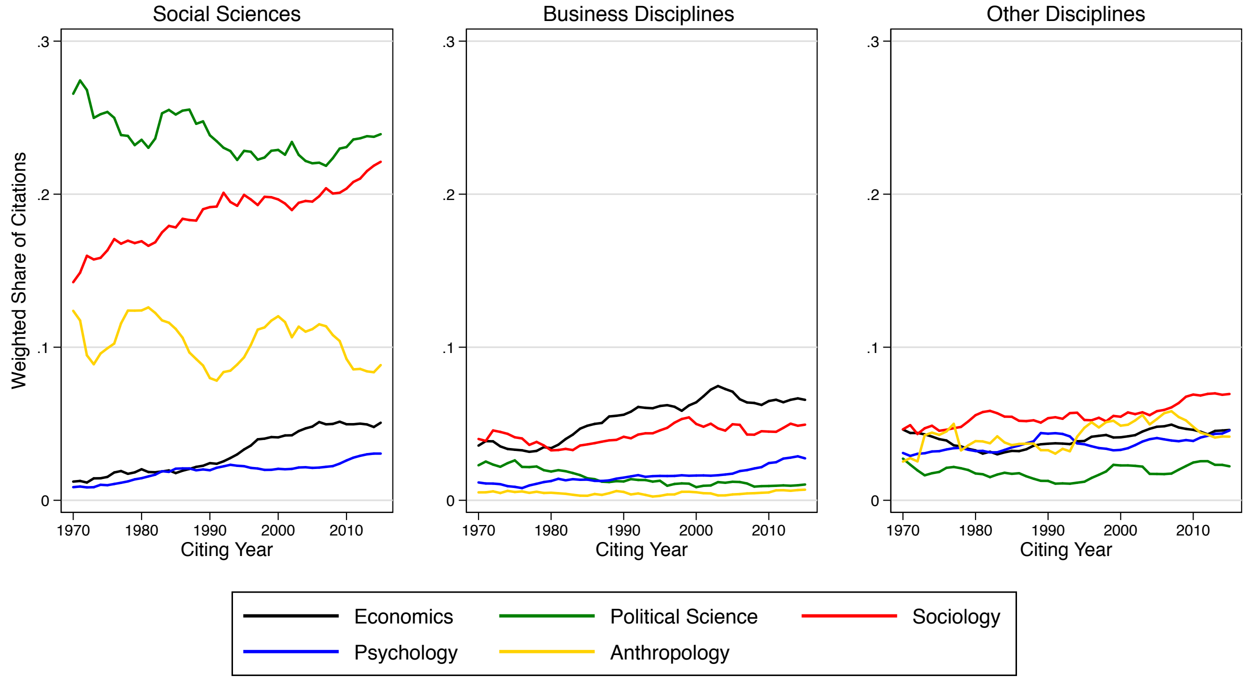
The role of empirical economics as a *cause* of increasing extramural influence probably varies by discipline. In finance, accounting, and political science, the influence of economics had reached or approached its peak

[25] Also as in Figure 15, each observation in the sample used to compute these estimates is one economics paper. The dependent variable in this case is the trunk-weighted citation share from non-economics journals over the life of the article. Shares are averaged across disciplines in each citing group.

by the mid-1980s, a period when empirical work was still playing second fiddle to economic theory. But these long-attentive disciplines have moved since the late 1990s to focus more on empirical work, a factor likely contributing to their sustained interest. The timing of economics' increased influence on psychology, public health, medicine, and multidisciplinary science, which dates roughly from 2000, is consistent with empirical work as a causal factor driving overall citation growth. On the other hand, increased interest in economics in OR and CS, which also starts around 2000, seems likely attributable to theory. This reinforces our observation that the diversity of economics scholarship is one of its strengths.

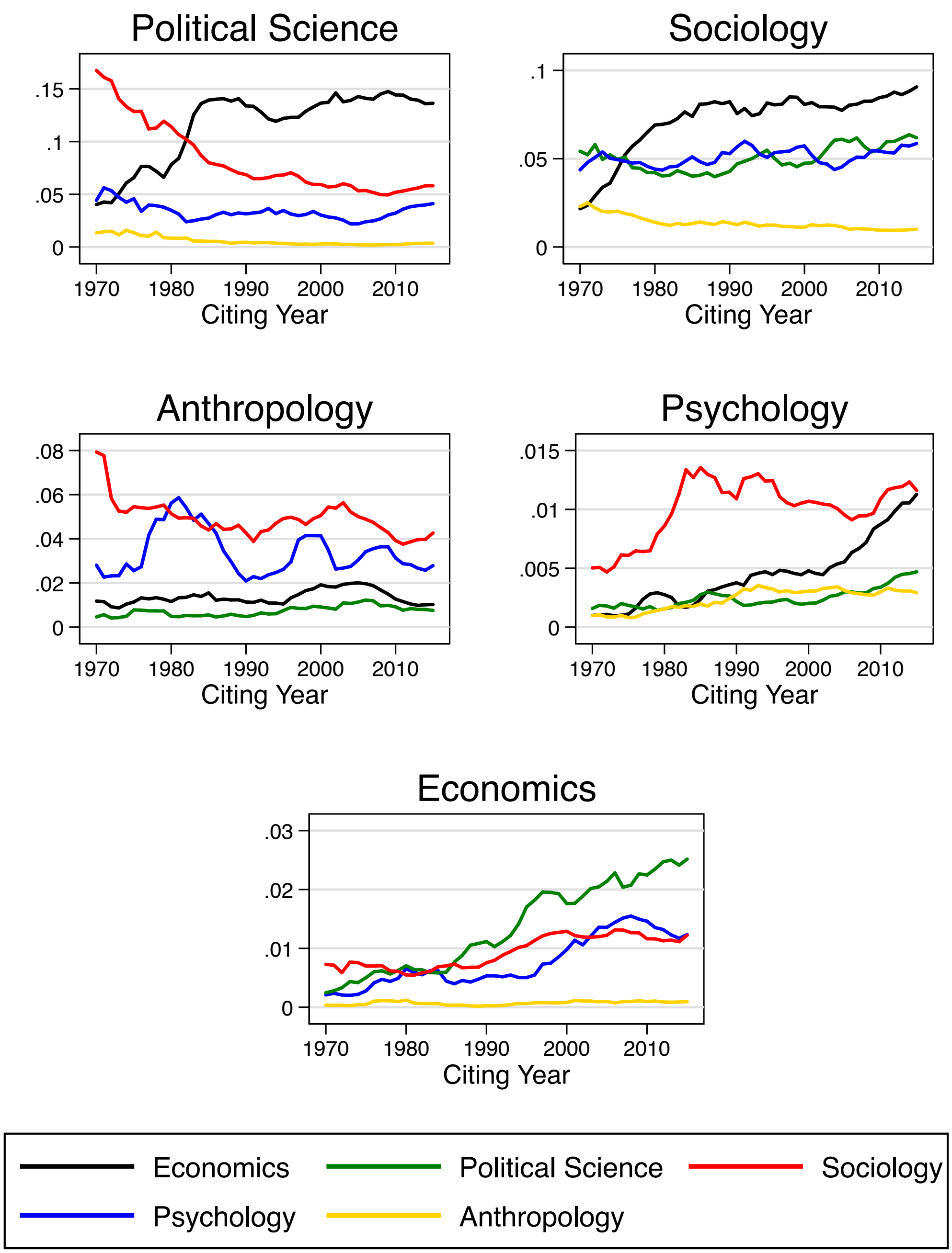
Finally, we return to the fact that economists are also increasingly likely to read other social sciences. This expansion of horizons is generating an extramural citation rate from economics to other social sciences that now exceeds the extramural citation rate from psychology. Since 1990, economics has been especially and increasingly attentive to political science. We see little in citation statistics to support the notion that economics is intellectually isolated. Rather, the growing links between economic research and a wide range of other disciplines reinforce our view that economic scholarship has never been more exciting or useful than it is today.

Figure 1: Social Science Insularity



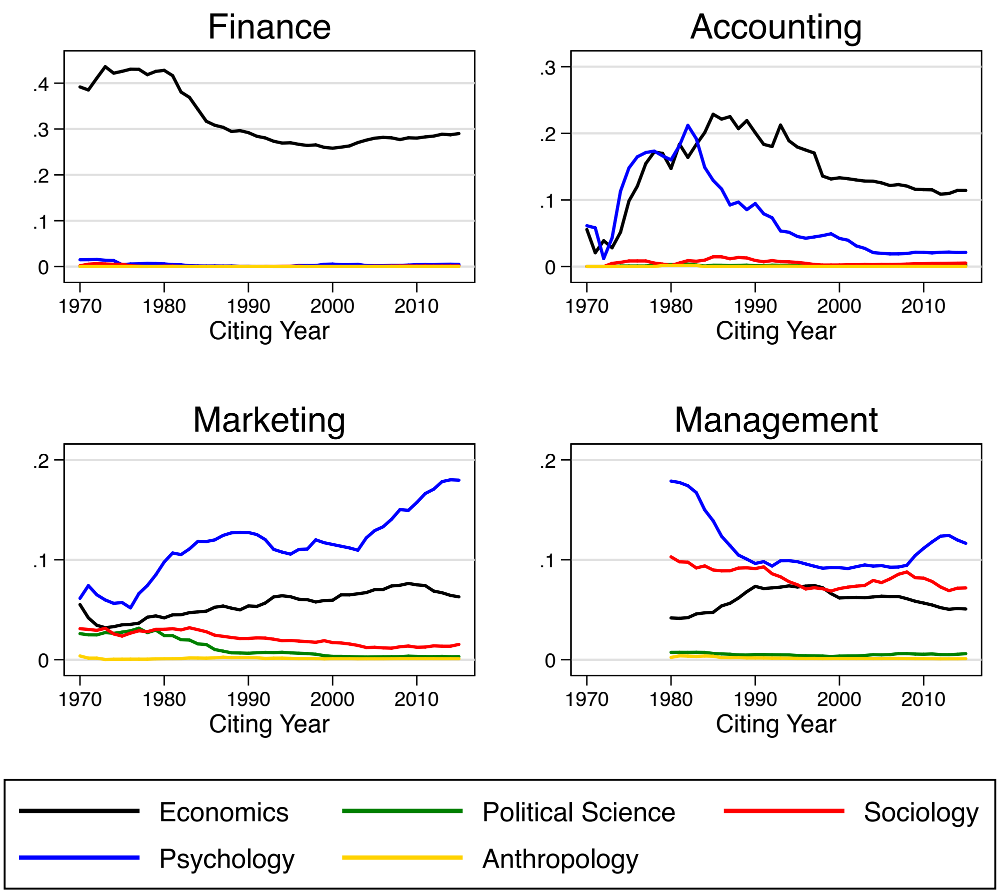
Note: The left panel of this figure shows citation rates from each social science to all other social sciences as a proportion of the total for each citing discipline. The center panel shows citation rates from each social science to four business disciplines. The right panel shows citations to seven other disciplines (this group includes all non-social-science and non-business disciplines, excepting multidisciplinary science). Plots are smoothed using five-year moving averages. Papers cited were published between 1955 and 2015.

Figure 2: Citation Rates between Social Science Disciplines



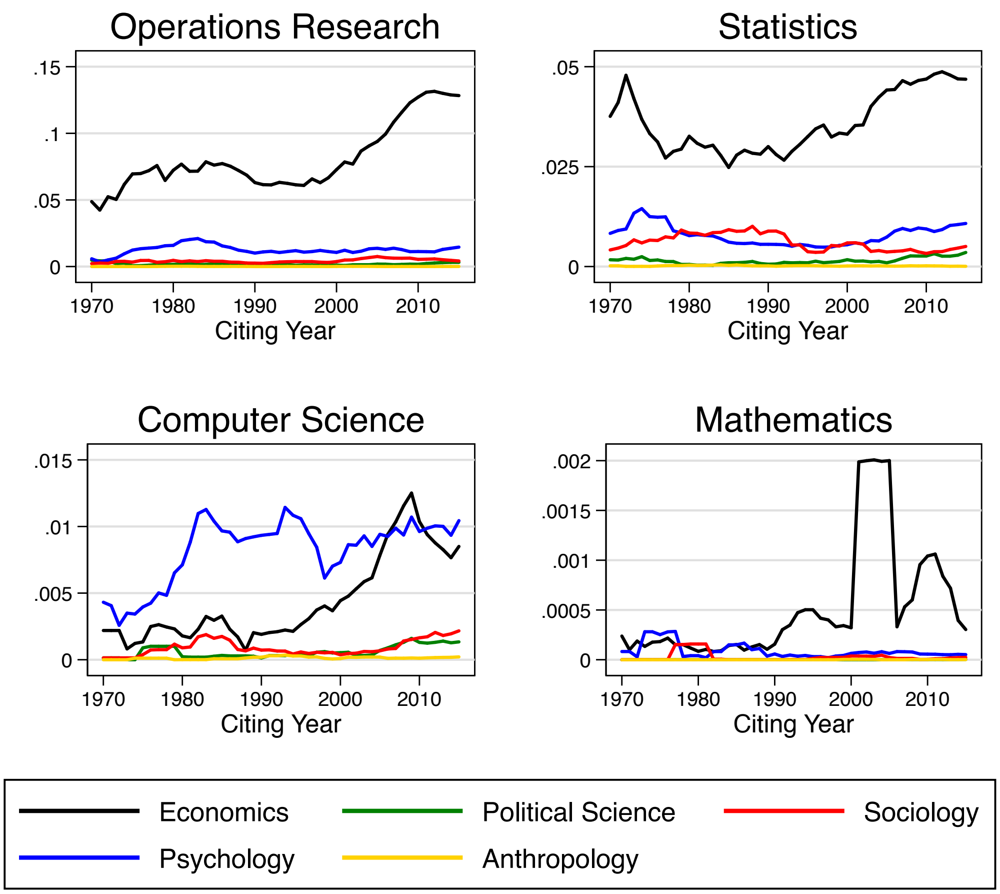
Note: This figure shows weighted citation rates from each of five social sciences to the other four. Plots are smoothed with five-year moving averages. Papers cited were published between 1955 and 2015.

Figure 3: Social Science Citation Rates from Business Disciplines



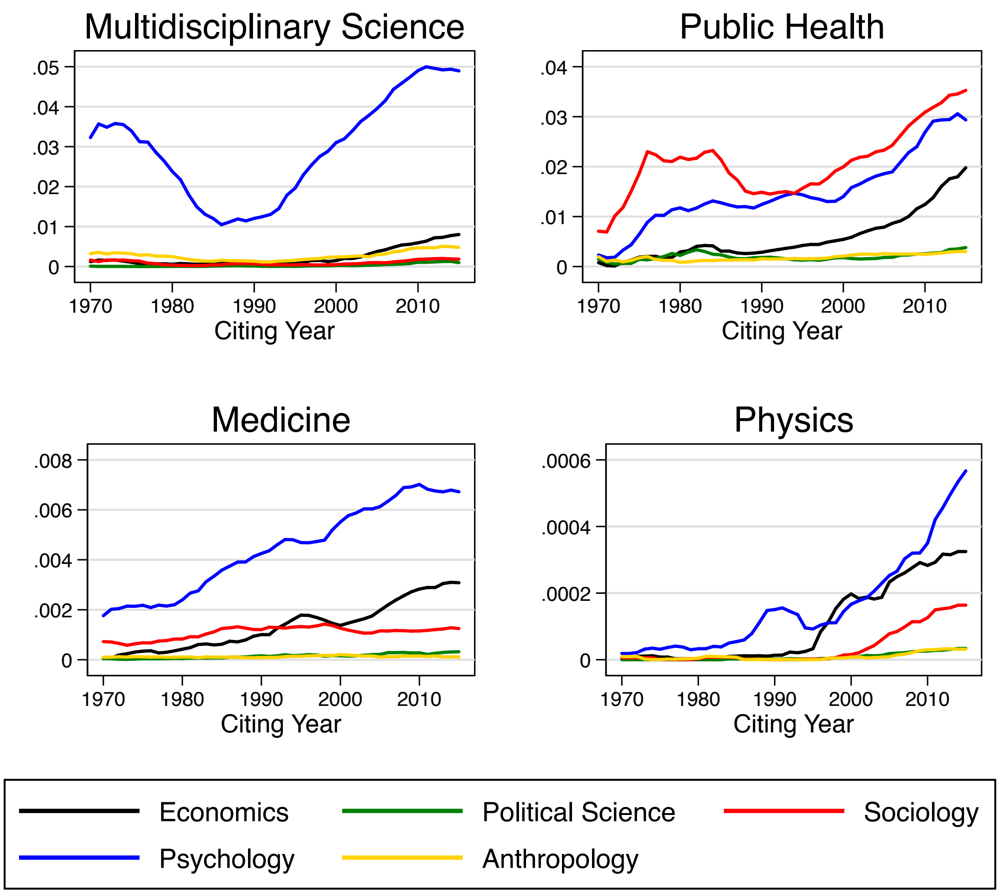
Note: This figure shows weighted citation rates from each of four business disciplines to five social science disciplines. Plots are smoothed with five-year moving averages. Papers cited were published between 1955 and 2015.

Figure 4: Social Science Citation Rates from Math Disciplines



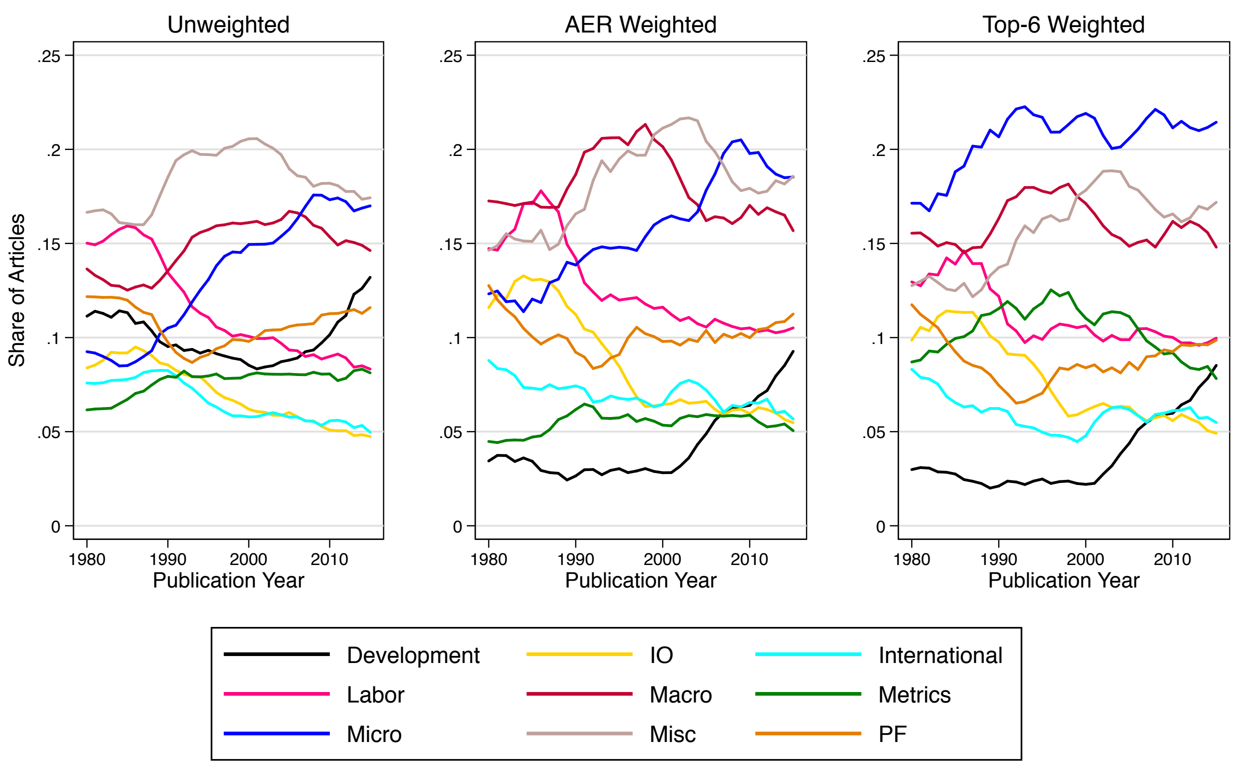
Note: This figure shows weighted citation rates from each of four mathematical disciplines to five social science disciplines. Plots are smoothed with five-year moving averages. Papers cited were published between 1955 and 2015.

Figure 5: Social Science Citation Rates from Other Sciences



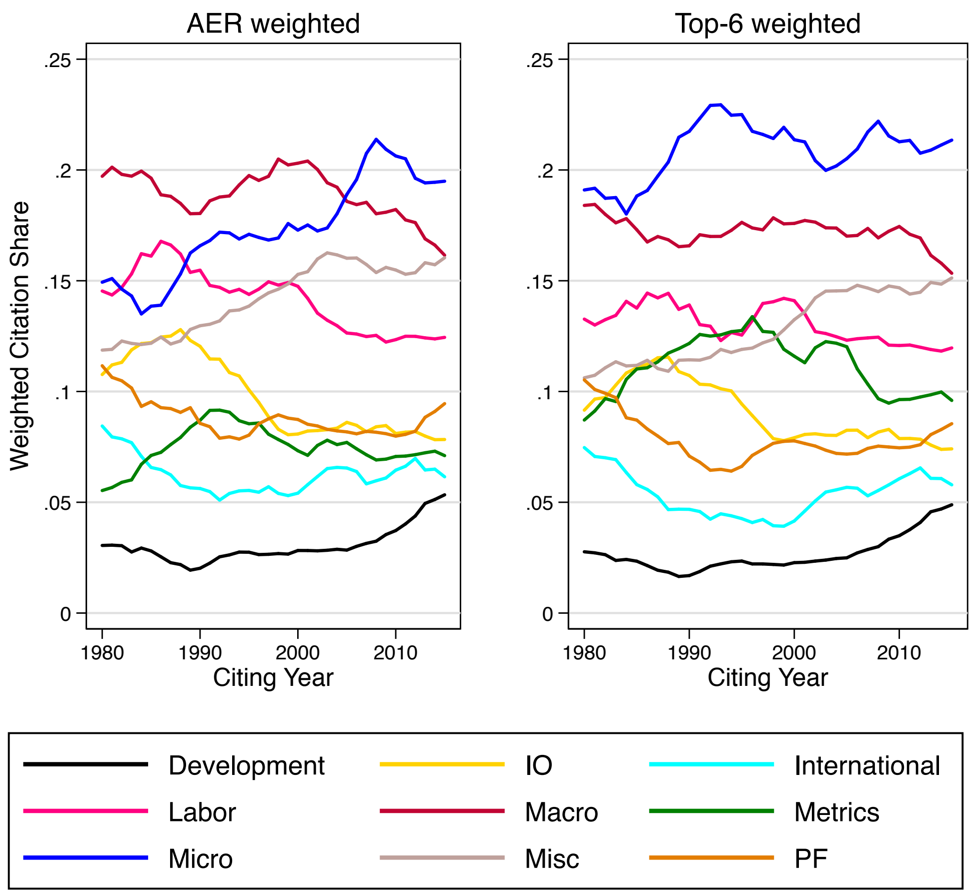
Note: This figure shows weighted citation rates from each of four other science disciplines to five social science disciplines. Plots are smoothed with five-year moving averages. Papers cited were published between 1955 and 2015.

Figure 6: Economics Publications by Field



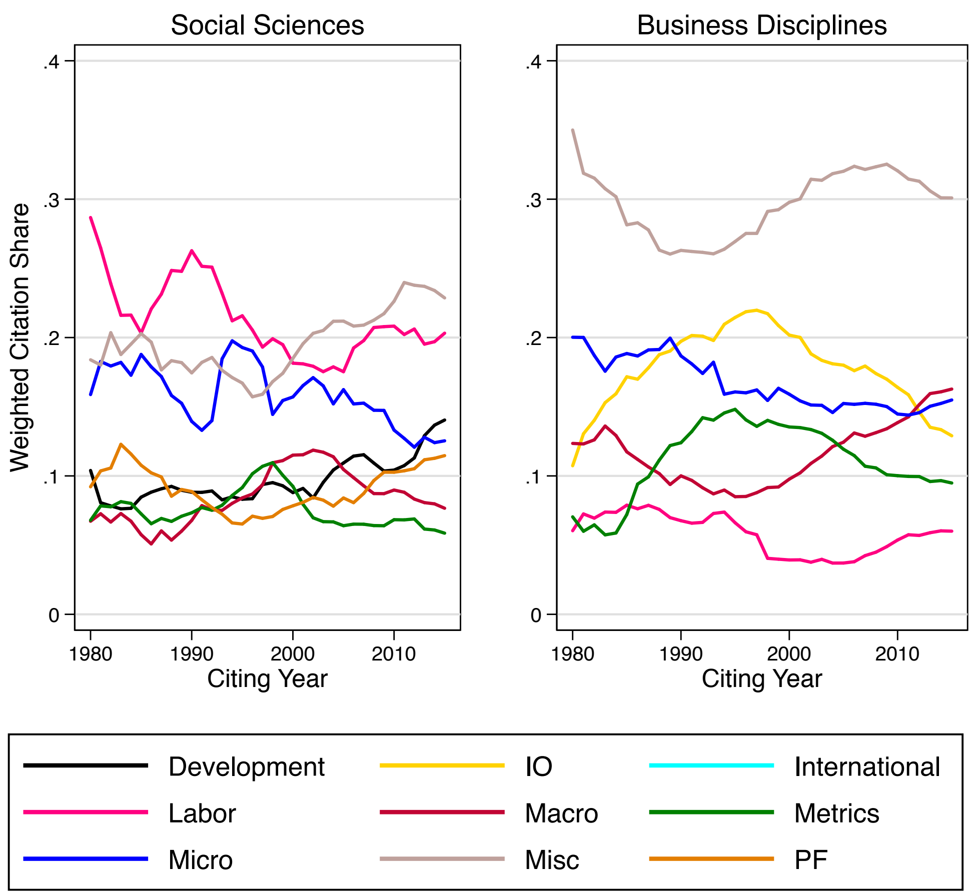
Note: This figure shows publication shares of economics papers in each field. Unweighted shares are presented in the left panel, and shares weighted by the importance of the publishing journal are plotted in the center (AER weights) and right panels (Top-6 weights). Plots are smoothed with five-year moving averages.

Figure 7: Economics Citation Shares to Fields



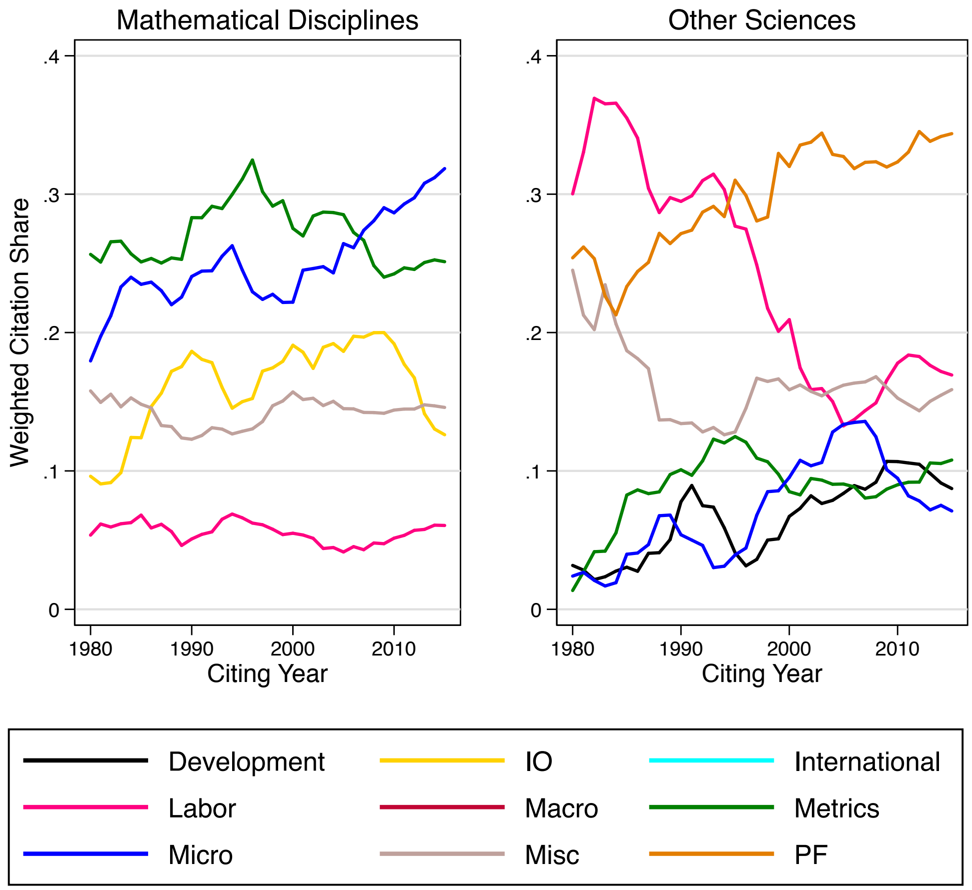
Note: This figure shows weighted citation shares of economics papers to economics fields. Citations are weighted by importance of the citing journal in the left (AER weights) and right panels (Top-6 weights). Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 8: Aggregate Extramural Citation Shares to Fields, Social Science and Business



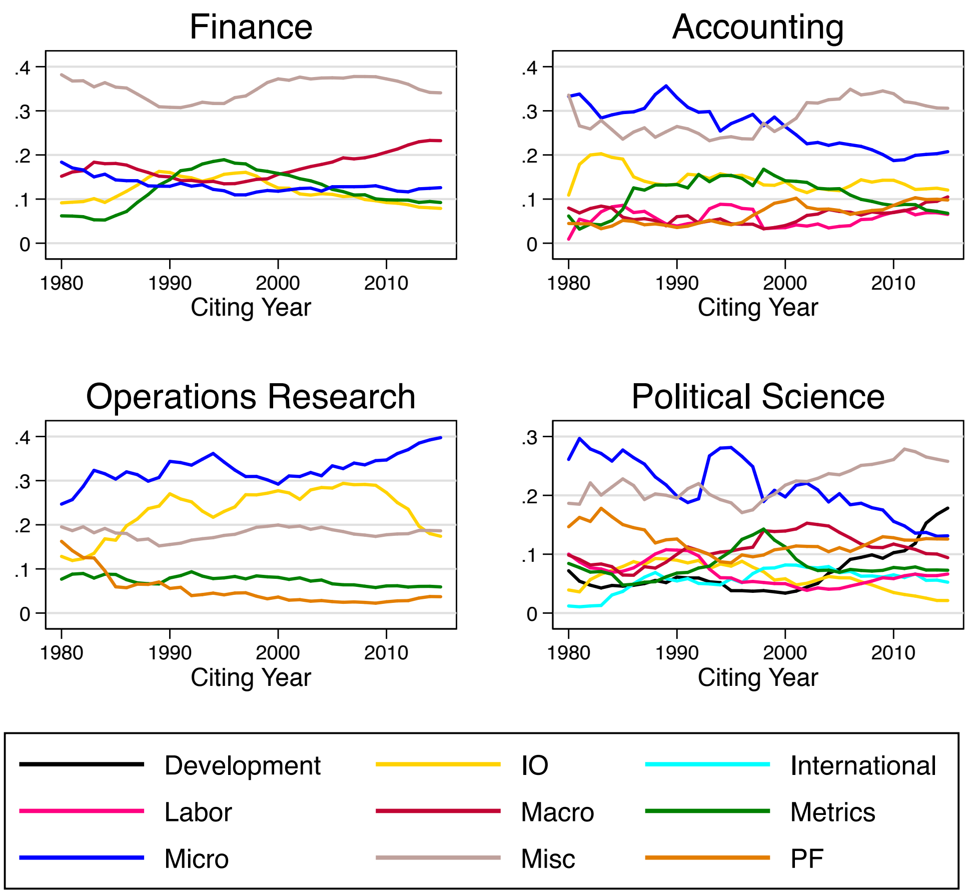
Note: This figure shows aggregated weighted citation shares from social science disciplines (psychology, sociology, political science, anthropology) and business disciplines (management, finance, accounting, marketing) to economics fields. Shares are plotted for the top 5 fields most cited, as well as for any field with at least a 5% average share across years. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 9: Aggregate Extramural Citation Shares to Fields, Math and Other Sciences



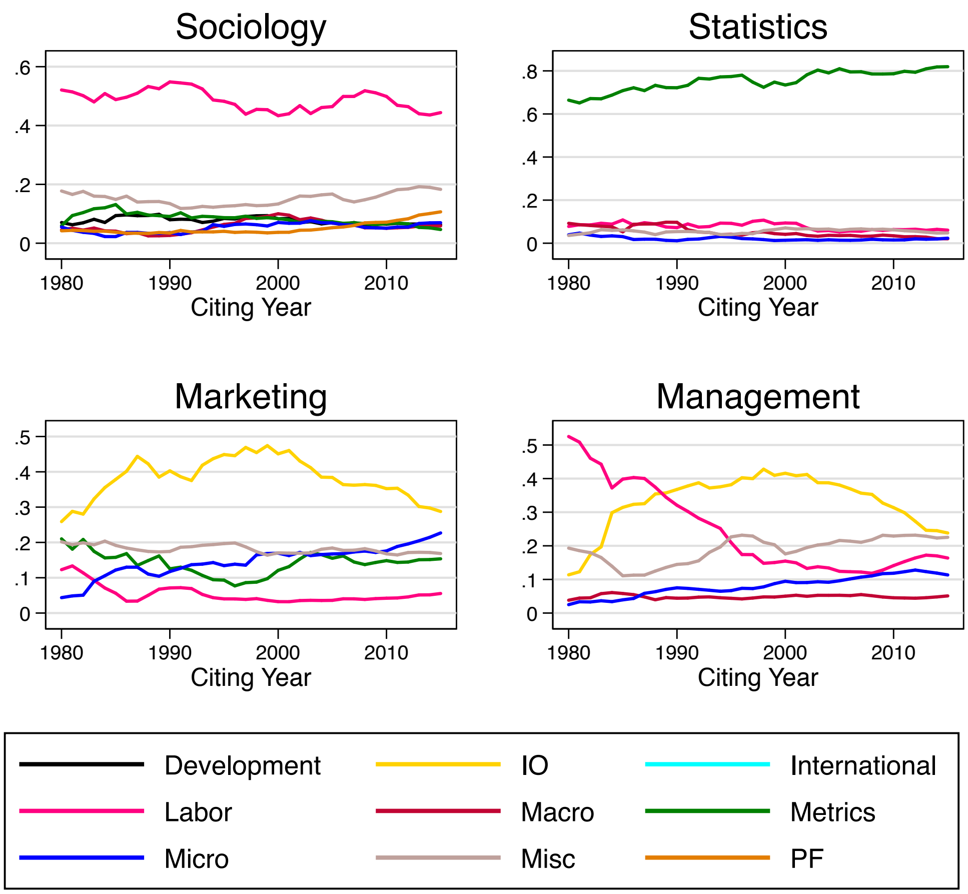
Note: This figure shows aggregated weighted citation shares from applied mathematical disciplines (statistics, OR, computer science, math) and other sciences (medicine, public health, physics, and multidisciplinary science) from to economics fields. Shares are plotted for the top 5 fields most cited, as well as for any field with at least a 5% average share across years. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 10: Citations from Discipline Group A to Economics Fields



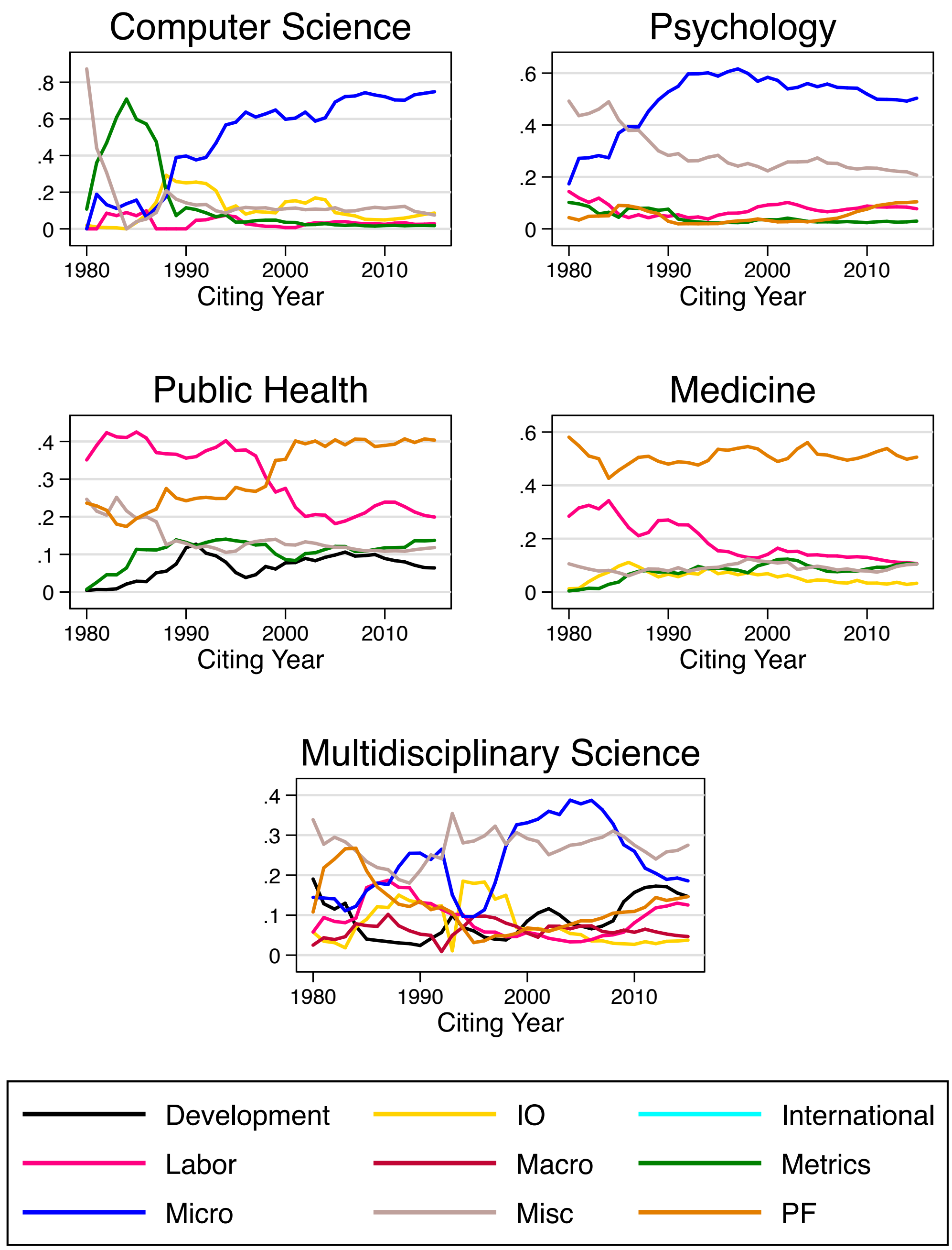
Note: This figure shows weighted citation rates from disciplines where economics is very influential (those where economics has a 10+% citation share) to economics fields. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 11: Citations from Discipline Group B to Economics Fields



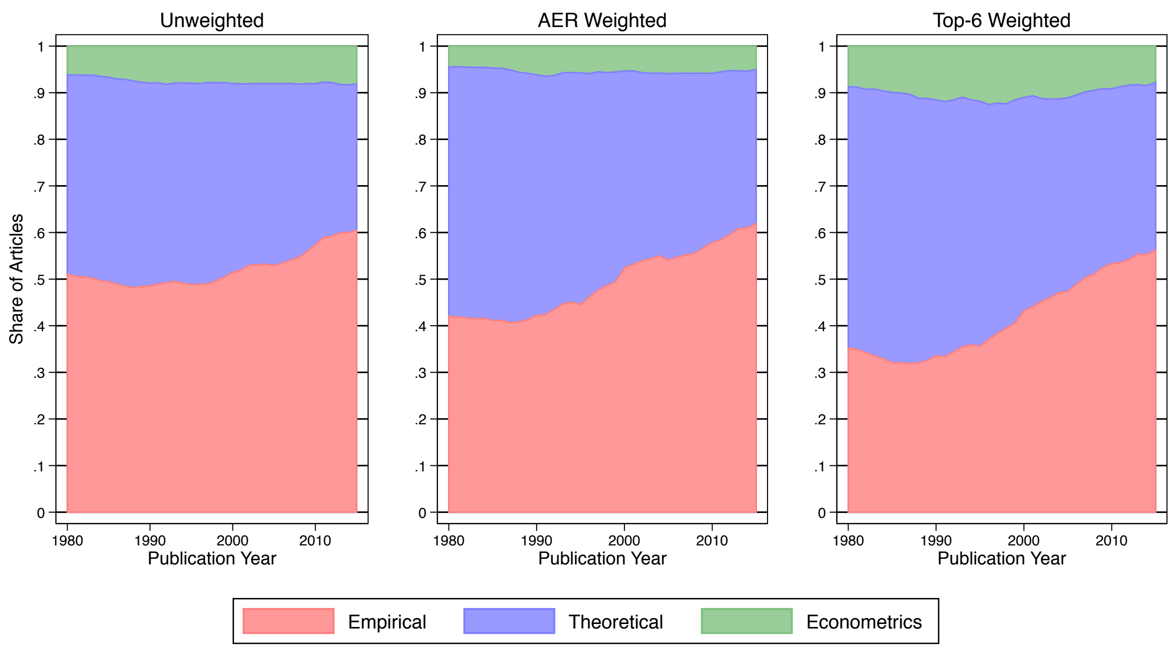
Note: This figure shows weighted citation rates from disciplines where economics is influential (those where economics has a 5-10% citation share) to economics fields. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 12: Citations from Discipline Group C to Economics Fields



Note: This figure shows weighted citation rates from disciplines where the influence of economics is growing to economics fields. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 13: Economics Publications by Style

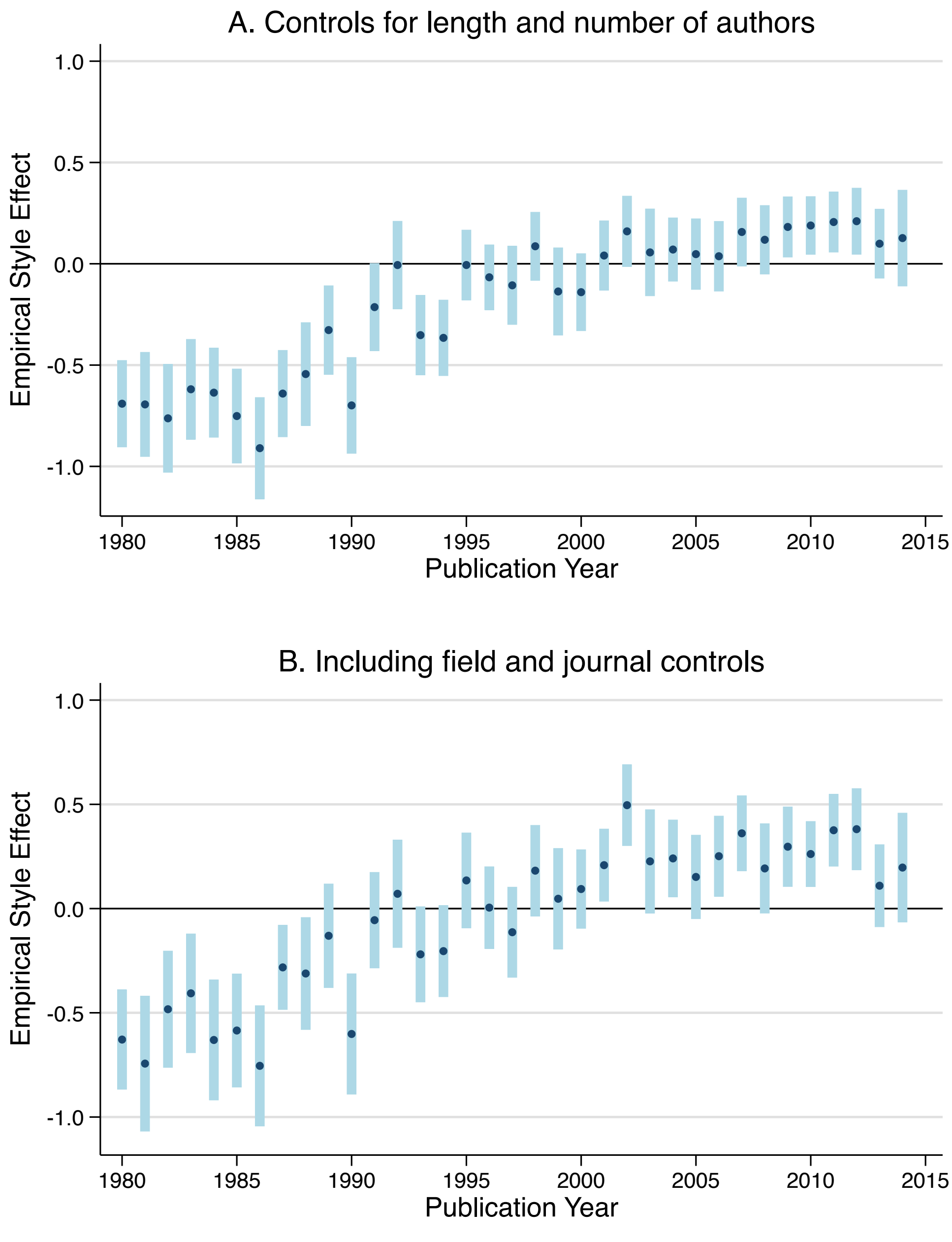


Note: This figure shows publication shares of economics papers in each style. Unweighted shares are presented in the left panel, and shares weighted by the importance of the publishing journal are plotted in the center (AER weights) and right panels (Top-6 weights). Plots are smoothed with five-year moving averages.

Figure 14: Economics Citation Shares to Styles

Note: This figure shows weighted citation shares of economics papers to economics styles. Citations are weighted by importance of the citing journal in the left (AER weights) and right panels (Top-6 weights). Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 15: The Empirical Effect in Economics Citations per Paper

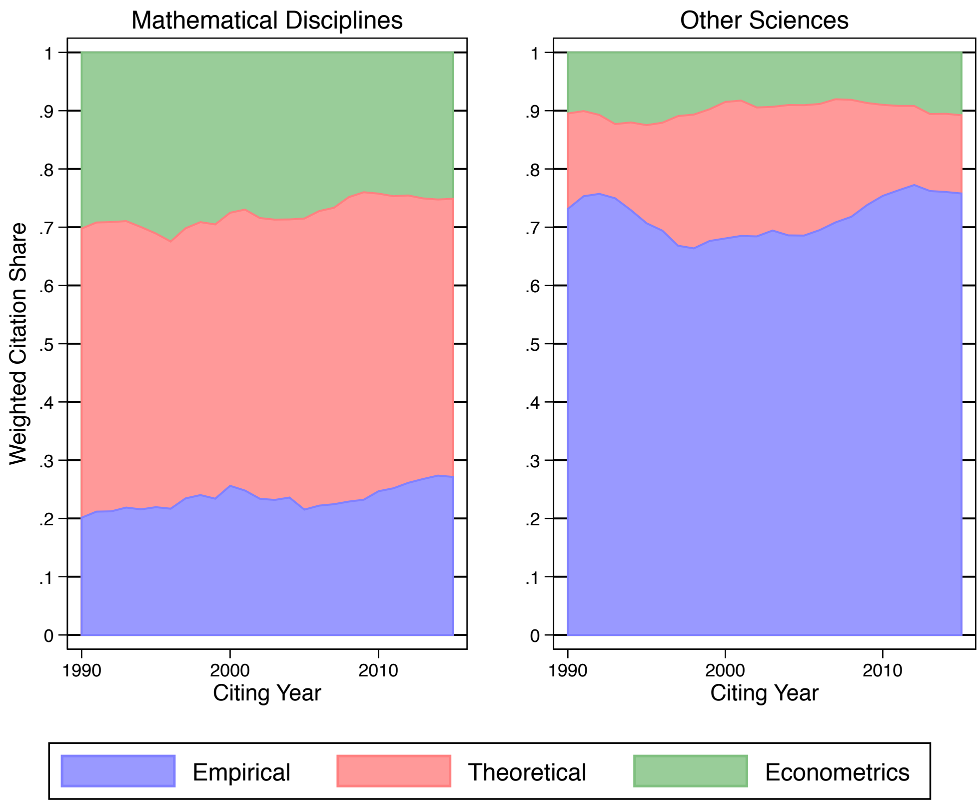


Note: This figure plots Poisson regression estimates of the empirical effect on weighted citations per paper. Panel A estimates are from models estimated separately by year, with flexible controls for paper length and number of authors. Estimates in panel B add field and journal controls. Confidence bands use robust standard errors

Figure 16: Extramural Citation Shares to Styles, Social Science and Business Disciplines

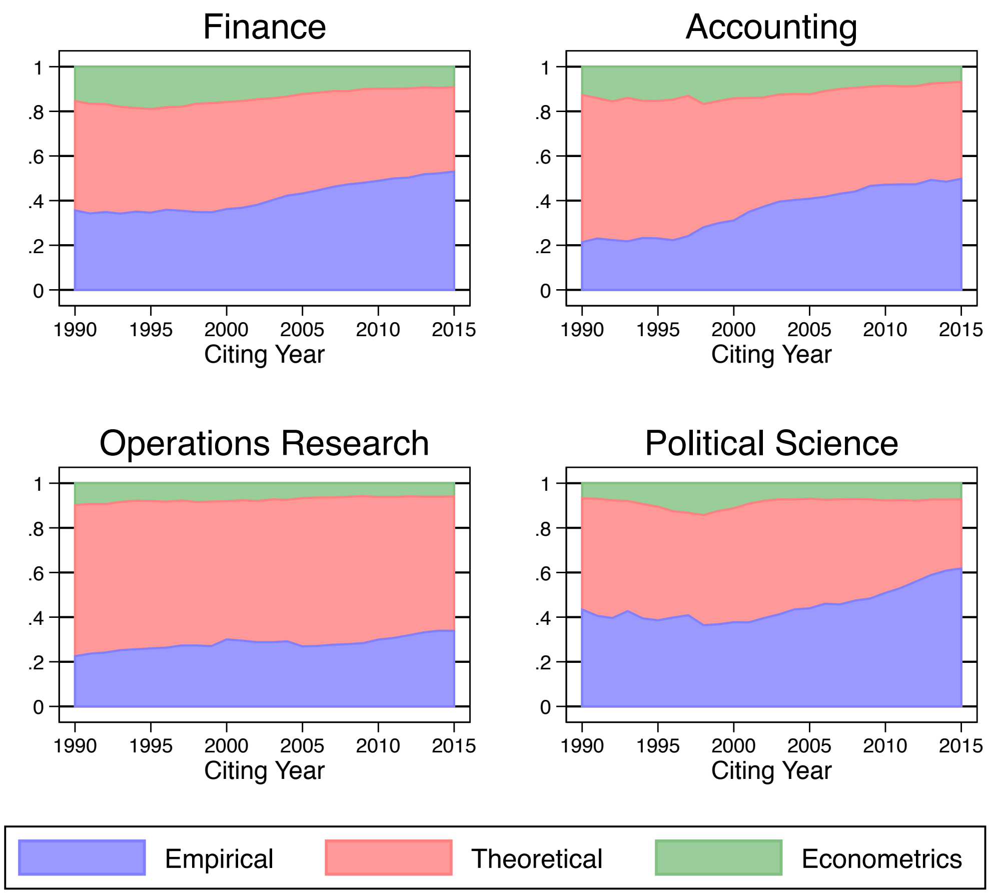
Note: This figure plots aggregated weighted citation shares from social science disciplines (psychology, sociology, political science, anthropology) and business disciplines (management, finance, accounting, marketing) to economics styles. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 17: Extramural Citation Shares to Styles, Math and Other Science Disciplines



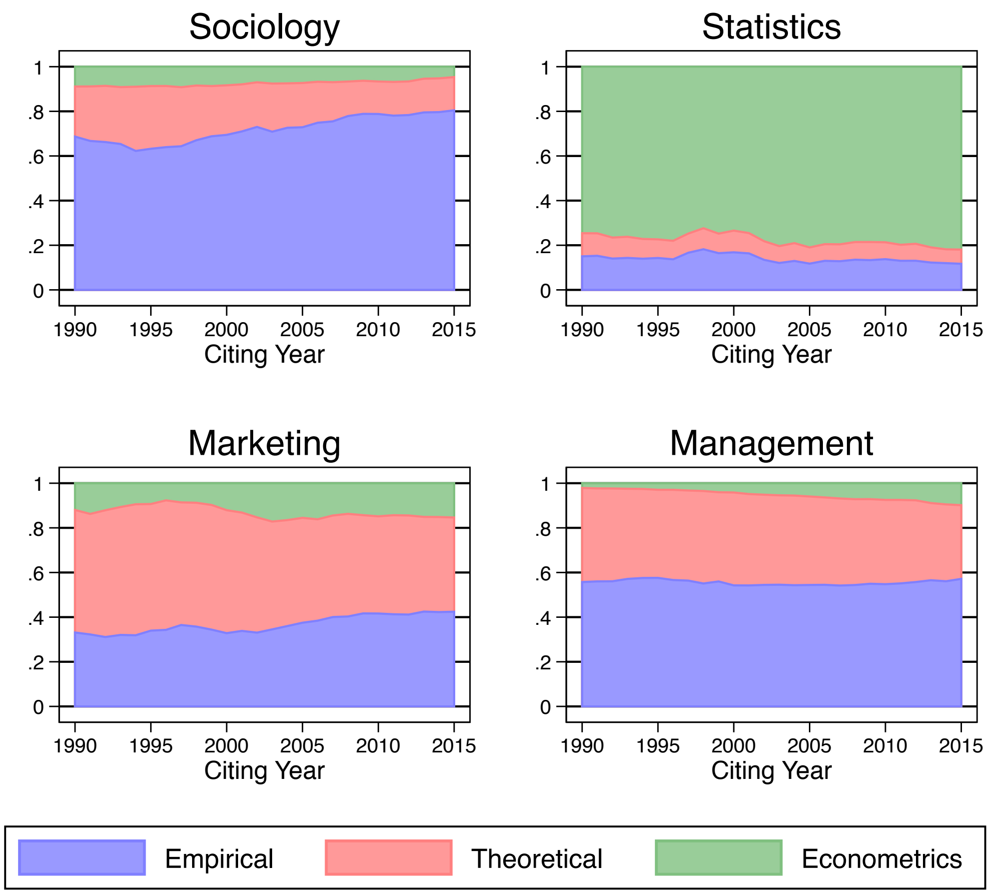
Note: This figure plots aggregated weighted citation shares from math disciplines (statistics, OR, computer science, and math) and other sciences (medicine, public health, physics, and multidisciplinary science) to styles. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 18: Citations from Discipline Group A to Economics Styles



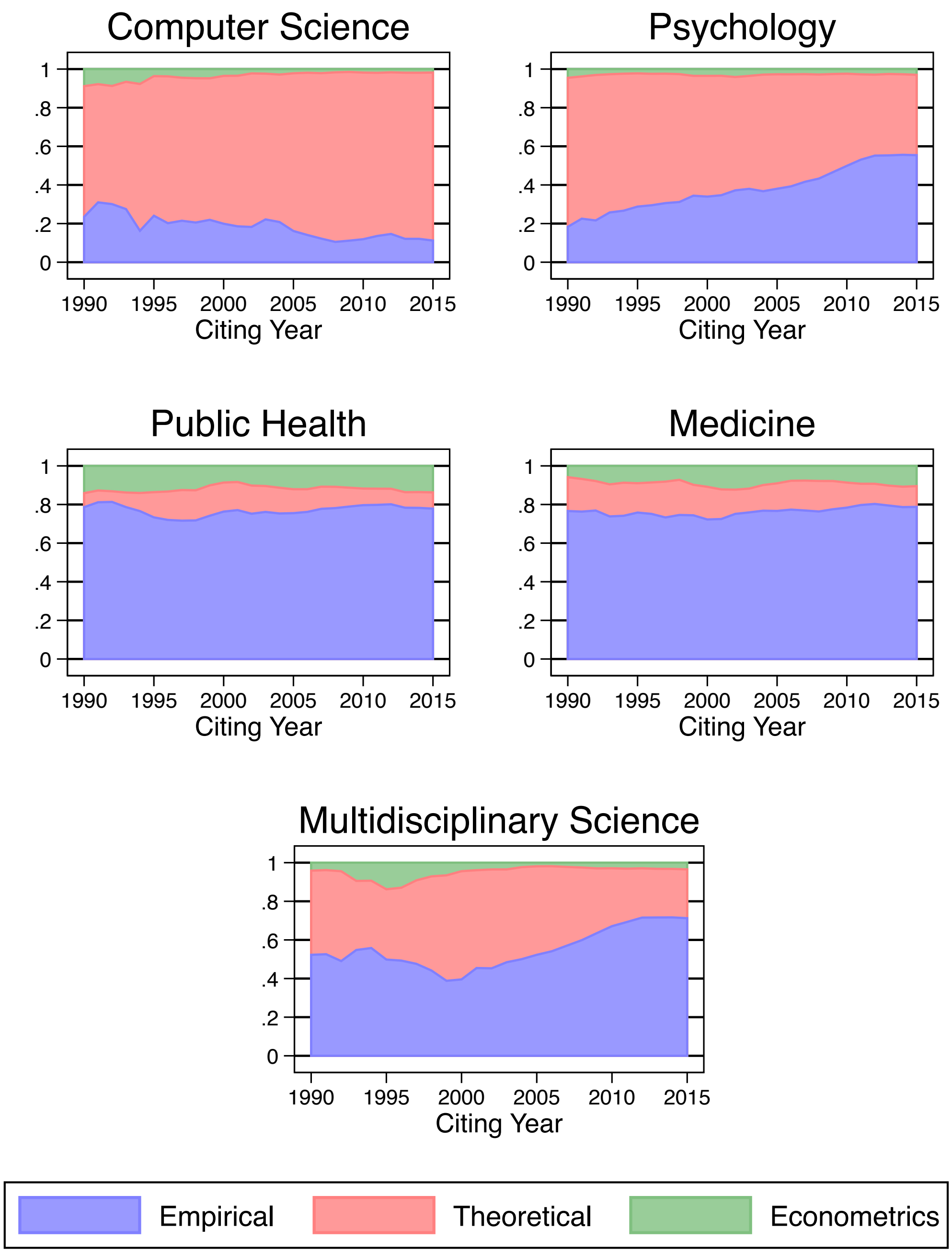
Note: This figure shows weighted citation rates from disciplines where economics is very influential (those where economics has a 10+% citation share) to economics styles. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 19: Citations from Discipline Group B to Economics Styles



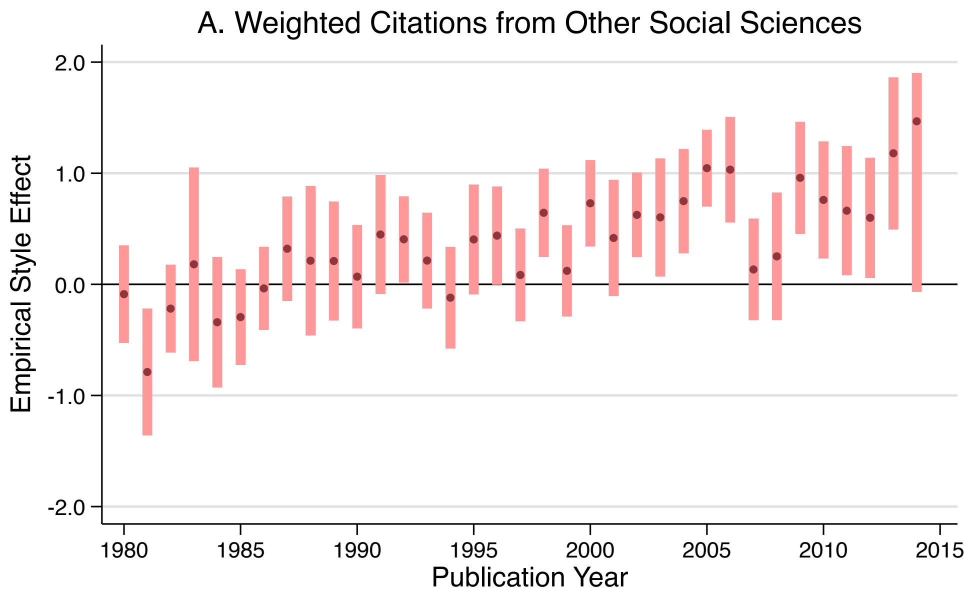
Note: This figure shows weighted citation rates from disciplines where economics is influential to economics styles (those where economics has a 5+% citation share). Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

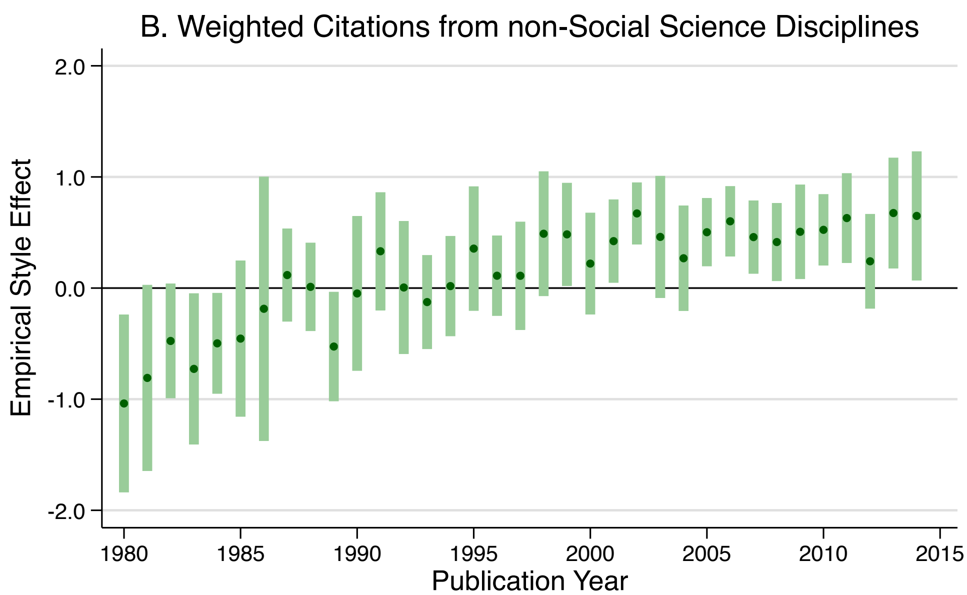
Figure 20: Citations from Discipline Group C to Economics Styles



Note: This figure shows weighted citation rates from disciplines where the influence of economics is growing to economics styles. Plots are smoothed with five-year moving averages. Papers cited were published between 1970 and 2015.

Figure 21: The Empirical Effect in Extramural Citations per Paper





Note: This figure plots Poisson regression estimates of the empirical effect on weighted citations per paper. Estimates are from models estimated separately by year, with flexible controls for paper length, number of authors, and field and journal controls. Panel A is weighted citations from non-Economics social sciences, and Panel B is from all other disciplines. Confidence bands use robust standard errors.

Table 1: The Distribution of Economics Fields and Styles

Field	Distribution by Initial Field				Distribution by Final Field			
	Empirical	Metrics	Theoretical	Total	Empirical	Metrics	Theoretical	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Development Economics	11,784	98	2,951	14,833	11,062		2,779	13,841
Econometrics	513	8,796	737	10,046		10,072		10,072
Industrial Organization	3,780	69	4,757	8,606	4,201		5,314	9,515
International Economics	4,247	69	4,402	8,718	4,608		4,502	9,110
Labor Economics	12,887	129	2,552	15,568	13,471		2,716	16,187
Macroeconomics	10,364	265	9,295	19,924	10,573		9,559	20,132
Microeconomics	2,511	118	15,502	18,131	2,787		15,407	18,194
Public Finance	8,927	117	5,758	14,802	8,830		5,816	14,646
Miscellaneous					16,854		8,611	25,465
Economic History	3,759	25	254	4,038				
Environmental Economics	2,259	37	1,896	4,192				
Experimental Economics	1,714	18	366	2,098				
Finance	1,668	163	1,800	3,631				
Law and Economics	897	13	875	1,785				
Political Economy	214	6	394	614				
Productivity	395	27	327	749				
Urban Economics	2,996	49	834	3,879				
Unclassified	3,471	73	2,004	5,548				
Total	72,386	10,072	54,704	137,162	72,386	10,072	54,704	137,162

Note: This table reports the number of economics articles published 1970-2015, indexed in both the Web of Science and EconLit, classified by economics field and research style. Initial fields follow the classification scheme used by Ellison (2002), with modifications discussed in the text and appendix. Final fields are produced by applying k-means clustering as described in the appendix. Styles are classified by machine learning based on a sample of hand classified articles.

Appendix A   Interdisciplinary Citation Rates

Extramural citation rates are constructed by classifying articles in the *Web of Science* into disciplines, and then computing the fraction of references from articles in discipline d to articles in d' in year t. Article discipline is determined by citation rates to and from the trunk journals listed in Appendix Table A**??**. Our citation rates are weighted averages, placing more weight on citations from each disciplines' most important journals. The importance of journal j to discipline d is measured by the rate at which discipline d's trunk journal cites papers in journal j. The importance of economics journals is also measured by citation rates from a top-6 composite journal, constructed as described in Angrist et al. (2017).

A.1   Constructing the Journal List

Our analysis covers 17 disciplines: five social sciences (anthropology, economics, political science, psychology, and sociology) and 12 other disciplines. We chose one or two trunk journals for each discipline, usually an association journal or journals. For example, the political science trunk is the *American Political Science Review* (APSR), published by the American Political Science Association.

Appendix Table A**??** lists the 17 disciplines covered here, along with the trunk journal(s) and professional associations that generate this list. The journal list for each discipline starts as the set of 50 journals most cited by the discipline's trunk journal(s) in decades defined as 1970-1979, 1980-1989, 1990-1999, 2000-2009, and 2010-2015. For disciplines with two trunks, citations are added. We define highly cited journals by decade in order to capture all journals that were ever important in a discipline, even if their influence has changed.

Journals in the top 50 for more than one discipline-decade were assigned to a single discipline-decade as follows. First, journals were assigned to the discipline whose trunk cited them most in a given decade. For example, between 2010-15, the *Quarterly Journal of Economics* was the second most-cited journal by the AER and the ninth most cited by the APSR. This puts the QJE in economics for 2010-2015.[26]

The list was then adjusted to take account of what each journal *cites*. This adjustment defines a set of "core journals" for each discipline-decade. Core journals are those in the smallest set initially assigned to each discipline-decade accounting for 30% of the citations from the relevant trunk journal(s).[27] We then counted the fraction of each journal's citations in each decade to core journals in each discipline. A journal-decade initially assigned to discipline d_0 was moved to discipline d_1 if the journal made at least 50% more citations to discipline d_1's core journals than to discipline d_0's core journals in that decade and if citations to discipline d_1's core journals comprised at least 5% of the total citations from this journal in that decade. For example, this rule moved *The Journal of Economic History* from Political Science to Economics in 2010-2015.

The journal reassignment process also produced a collection of unclassified journals, for which a discipline was not clearly identified. Specifically, journals were deemed "unclassified" when fewer than 3% of their

[26] An exception here is *Science*, which was assigned to multidisciplinary science even when ranked more highly by another discipline's trunk.

[27] The core journals for public health are its trunks. This was motivated by the large number of citations from public health to medicine, a different discipline in our taxonomy.

outgoing citations were to core journals in their originally assigned discipline and no other discipline cleared the thresholds above for a new assignment. These reassignment rules were applied with an exception for the public health discipline. Because public health papers cite many in medicine, the 5% moving threshold was raised to 10% and the the 3% "unclassified" threshold lowered to 2% for journals initially in public health.

The procedure detailed above assigns journals to disciplines for each decade. Time-invariant discipline assignments were produced by assigning each journal to the discipline to which it was assigned in the largest number of decades. Journals without a unique modal assignment, or for which "unclassified" was the modal assignment, were designated (or remained) unclassified.

Finally, we modified a few algorithmic assignments that seemed incorrect. These modifications sent *the Harvard Business Review* and *Organizational Behavior and Human Decision Processes* to management, *the Annual Review of Public Health* to public health, *Nature Medicine* to medicine, and *the Journal of Econometrics*, *Econometric Theory*, *the Bell Journal of Economics and Management Science*, and *the Journal of Business Economics and Statistics* to economics. We found that the process of algorithmic assignment led the multidisciplinary science discipline to collect journals that are not really multidisciplinary. Many of these journals, like *The Journal of Cell Biology*, cover life science topics that are neither medicine nor public health, and, like those from chemistry, almost never cite the social sciences. We therefore restricted the multidisciplinary list to three prominent and well-cited multidisciplinary journals: *Science, Nature*, and *PNAS*. The final journal lists for all disciplines can be found in the online appendix.

Appendix B Field Classification

B.1 Overview and Data[28]

Our field classification starts by classifying articles into one of 17 "initial fields," using each article's Journal of Economic Literature classification (JEL) codes reported in *EconLit*. Many papers have multiple JEL codes. We therefore use machine learning to assign a single initial field to papers with more than one code. The second field classification step uses each paper's initial field classification and the initial field of the papers each paper cites to form 9 clusters. These clusters, constructed using the k-means algorithm, become our "final fields".

We classify *EconLit* papers published in journals on the economics journal list in the period 1970-2015. The sample for field (and style) classification is limited to papers matched to the *WoS* database because our analysis relies on the citation network unique to *WoS*. *EconLit* provides bibliographic information, JEL codes, keywords, and abstracts for most of these papers. Our copy of *Econlit* indexes 214,868 articles published between 1886 and 2016. Restricting this file to papers published in journals on our economics journal list from 1970-2015 leaves a database containing 145,680 papers.

Information on cited papers comes from the *WoS*. The potential *WoS* sample includes 192,091 articles published in journals on our journal list published from 1970-2015. This is a larger set of papers than the

[28]Appendices B and C were drafted by Suhas Vijaykumar.

set found in *EconLit* for the same journals and years because *WoS* indexes a wider variety of document types. For example, *WoS* indexes each book review separately, while *EconLit* largely ignores these. Other documents found only in *WoS* are editor's notes, conference announcements and notes, and econometric problems and solutions. Since most of these missing publications rarely cite or are cited by other articles, their omission is unlikely to matter.[29]

There is no unique identifier common to *WoS* and *EconLit*. We therefore started by matching each article's journal issn, publication year, volume, issue, start page number, and end page number. This generates 127,484 matches. An additional 8,474 papers are matched on title and author (after removing capitalization, punctuation, common speech articles and author first names). Finally we execute a Stata `reclink` fuzzy merge using issn, year, volume, issue, start page, end page, and author last names. We evaluate these fuzzy matches manually based on the match score and title. The final matched sample contains 138,079 papers, or 94.7% of potential *EconLit* matches.

We omit articles that do not contain at least one JEL code, since this feature is used to classify fields. Almost all of the articles without any JEL code were published in 1990, a year in which only 75% of articles published in 1990 have codes (this probably resulted from the transition to a new JEL system). The final classification sample for fields and styles contains 137,162 articles.

B.2 Classification into Initial Fields

Our 17 initial fields are microeconomics, macroeconomics, public finance, labor, industrial organization, development, urban economics, environmental, econometrics, finance, international, experimental (lab), economic history, political economy, productivity, law and economics, and other. Each JEL code is mapped to one of these fields using the scheme in Ellison (2002). Each article is then assigned a unique initial field using machine learning (ML) as described below.

B.2.1 Training Data

We assembled a training dataset that exploits the fact that before 2004, JEL codes typically appear in *EconLit* in order of importance rather than alphabetically. We therefore assigned fields using the first JEL code for papers published in these years. Our ML algorithm treats fields assigned in this manner as a dependent variable, to be predicted using the full set of up to 7 (unordered) JEL codes as well as article titles and keywords.

These training data are supplemented with a set of field assignments for articles in widely recognized field journals (like the *Journal of Labor Economics*). Regardless of the JEL codes listed for these articles, the field journal's field becomes the dependent variable for articles in these journals.

Articles with a single JEL code were omitted from the training data because our scheme makes the set of JEL codes for these articles perfectly informative about fields. Training data with these articles included

[29] For example, the *WoS* indexes 477 entries in a report of the World Congress of the Econometric Society published in *Econometrica* issue no. 4 in 1971, while *EconLit* omits these.

over-represents the prevalence of single-code fields, generating a misleadingly high success rate. Although single-JEL papers are not in the training data, they were classified by the ML model. The ML algorithm reclassified a few of these papers using information in titles and keywords.

B.2.2 Development and Political Economy Training Supplement

Fields that have shifted research focus since the 1970s and 1980s proved hard to classify. We especially struggled with development and political economy; many recent development papers were initially classified as labor or public finance, while our ML routine classified many studies that are now considered political economy as macro or public finance. We believe this problem arises from the evolution of topics within these fields. Development economics has moved from studying growth and institutions in developing countries to a much broader set of topics. Modern development authors cite earlier development papers little, instead citing methodologically similar studies in labor and public finance. Development authors today often assign JEL codes from these other fields as well. Political economy has also seen a sea change towards empirical papers that often make little with earlier work in the field. We therefore supplemented the training data with 481 articles that were randomly selected from the set of papers that had at least one development or political economy JEL code published after 1990.

The random sampling procedure for this purpose weighted papers based on the share of AER citations that the article's journal received in the publishing year. Papers in top journals therefore make up the bulk of this training supplement. These papers were hand classified into fields by trained research assistants and added to the training data set. Although these papers contained at least one development or political economy JEL code, most of them were classified in other fields, with development and political economy classifications given to 18% and 20% of the supplement respectively.

B.2.3 Field Classification Algorithm

The training data set was used to train a random forest classifier for multi-JEL papers (Breiman, 2001). Predictors include (up to 7) fields for (up to 7) JEL codes, dummies for words occurring in the title, and dummies for keywords.[30] Words occurring in the titles and keywords of more than 50% of articles or fewer than .5% of articles were excluded. Titles were preprocessed using standard procedures in the Python Natural Language Toolkit (NLTK) (Bird, Klein and Loper, 2009), including stemming words (e.g. "regressing" is reduced to "regress"). Geopolitical entities were tagged and numbers were replaced by a word indicating their type (e.g. year, decimal, fraction, percentage, integer). Finally we marked papers that had the name of a non-OECD country in the title to further address the challenge of identifying modern development papers.

We classified papers into fields using a random forest algorithm because this worked well in cross-validation comparisons with other schemes.[31] Our classifier consists of 500 trees with 30% of covariates sampled for each tree, with each tree trained to classify a sample of articles drawn randomly (with replace-

[30] Classification and coding uses the Python "Scikit-learn" package (Pedregosa et al., 2011).
[31] See Morales (2017) for more on relative algorithm performance in the task of economics field classification.

ment) from the training data set. The number of covariates per tree was chosen to minimize classification error in a 90-10 split-sample test. Also in a 90-10 split sample test, the algorithm with these parameters classified 78.7% of training articles correctly.

B.3 Clustering into Final Fields

Nine final fields were constructed by clustering the 17 initial fields using a k-means algorithm that looks at each paper's initial field and the initial fields of the papers it cites[32]. This process allows us to focus on larger fields and moves papers partly on the basis of articles authors choose to cite[33]

Our application of k-means uses a weighting scheme to balance the influence of papers' own initial field and the initial fields of cited articles. Specifically, each article, i, is assigned dummies for initial field, denoted D_{fi} for field f, and 17 variables that count the number of cited articles on article i's reference list for each field, denoted N_{fi} for field f. We then weight these variables as follows.

First, a reference list weight is defined:

$$w_i^{ref} = w^a \cdot (1 - w^b(1 - x_i))$$

where x_i is the percentage of reference list citations that were classified using *EconLit* data. Since our classified set of papers covers only 70 journals and 45 years, many reference list papers are not classified. We down-weight the influence of reference list fields for papers that have a low percentage of classified references. We found that the reference list fields were more informative for papers published in later decades, so we increased the weights linearly across years. The weights w^a and w^b were preselected after inspection of a range of values; we used $w^b = 0.3$ and a year specific $w^a = 0.635 + \frac{\text{year}-1970}{1000}$.

Next we define the own-field weight:

$$w_i^{own} = 1 - w_i^{ref}.$$

Finally, we create 17 variables own_{fi} and 17 variables refr_{fi}

$$\text{own}_{fi} = D_{fi} \cdot ({w_i^{own}}/{17})$$
$$\text{refr}_{fi} = (\text{share}_{fi} - \overline{\text{share}_f}) \cdot ({w_i^{ref}}/{17}),$$

where $\text{share}_{fi} = \frac{N_{fi}}{\sum_g N_{gi}}$ is the fraction of articles in field f on article i's reference list, and $\overline{\text{share}_f}$ is the average over all articles for field f. The variables own_{fi} and refr_{fi} are used as features in the k-means clustering algorithm. A set of 16,887 articles with no references to other papers in our merged sample are manually assigned to clusters using their initial own-field classification.

[32]See Bishop (2006) for more on `kmeans`, a Matlab package used for this purpose

[33]For example, Kamenica and Gentzkow (2011) develop a model of persuasion with applications to litigators, lobbyists, and salespeople. This paper gets law and economics as an initial field by virtue of the paper's JEL codes and microeconomics as a final field by virtue of the fact that 72% of the papers it cites are initially classified as micro.

Appendix C Style Classification

Economics articles were classified into three styles: empirical, theoretical, and econometrics. Papers are first classified as empirical. Among those not classified as empirical, those not in the econometrics field are classified theoretical. As with classification into fields, style classification uses supervised machine learning. Specifically, style classification uses logistic ridge regression with inputs (explanatory variables) derived from article titles, journal identifiers, initial fields, keywords, publication decade, and abstracts (where available). Also as in the field classification procedure, this algorithm was chosen after comparing several alternatives.[34]

Roughly 30% of articles to be classified have no abstract. Not surprisingly, classification is more accurate with an abstract. We therefore first classified the full sample without using abstracts, then separately classified the subset of papers with abstracts using information from abstracts as additional features. The final classification gives precedence to the with-abstract classification result where available.

C.1 Training Data

The training sample for style classification contains 5,469 hand-classified articles over-representing top journals. The training data include:

1. Articles originally classified as empirical or theoretical by Ellison (2002). These papers are from top-6 economics journals and published from 1971-1998: 1,503 articles
2. Articles from entire issues of the AER, JPE, and Econometrica, as follows
   - AER, 1992-2004: 485 articles
   - Econometrica, 1998-2013: 822 articles
   - JPE, 1987-2014: 931 articles
3. Fifteen randomly chosen articles from each journal in our economics list published 1980-1989: 678 articles
4. Fifteen randomly selected articles per economics journal per decade (1990-1999, 2000-2013) for top-20 journals based on cites from the AER. Five randomly selected articles per journal per decade for all other journals: 1,050 articles

C.2 Text Processing

We pre-process the text contained in titles, keywords, and abstracts to produce informative features for ML. This reduces dimensionality of text data and takes advantage of semantic similarities between documents.

The title and keywords are turned into a word-document matrix, where the rows represent documents and the columns represent unique words. The entries of this matrix count word frequency in a document.

[34] Algorithms compared include logistic regression (with L1 and L2 penalty), support vector machines (with L1 and L2 penalty), binary classification trees, the naive-Bayes algorithm, k-nearest-neighbor classification (with both standard and `word2vec` embeddings), and classification using a shallow convolutional neural network (Kim, 2014).

We drop words that occur in less than .001% or more than 50% of articles. We then fit a topic model to these title and keyword data using Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003). This reduces dimensionality by forming topics containing groups of words that commonly appear in the same documents. Each document is then represented as a distribution over topics. Since titles contain only 10-15 words drawn from a vocabulary of about 20,000, they are highly sparse, and many informative words never appear in the training data. LDA is a popular dimension-reduction tool used in this scenario to capture similarity between documents (in this case, titles). We fit a model of 10, 30, 50, 70, 90, 110, 130, and 200 topics, following past work in the natural language processing literature on the classification of short text (Chen, Jin and Shen, 2011). The resulting topic data set was used in classification both with and without abstracts.

We process words in abstracts (where available) using term-frequency minus inverse-document-frequency (TF-IDF). Here, we restrict the word-document matrix to words appearing in $.1-50\%$ of abstracts. TF-IDF is a metric computed by dividing the number of times a word appears in a specific document by the number of times that same word appears in all documents to be classified (Wu et al., 2008). This process puts more weight on words that are unique to papers, causing the ML procedure to respond to the most informative text.[35]

C.3 Classification

The full set of features used for style classification are the LDA and TF-IDF weights described above, an indicator for titles containing a question mark, fields assigned by the field classification procedure, journal names, and journal-decade interactions.

Using these predictors, articles were classified as empirical using ridge logistic regression, with regularization parameter $\lambda = .0013$ for classification with abstract data (respectively $\lambda = .0015$ without abstract data). The regularization parameter was chosen to maximize accuracy in a 90-10 split sample; the split was repeated 5 times for each potential choice of regularization parameter λ. In split-sample tests, classification accuracy was 81.7% without abstracts and 87.5% with abstracts.

As noted in the text, two raters classified fields and styles in a random sample of 100 papers. Rater styles agree with ML styles about 80% of the type and with each other 82% of the time. Inter-rater and ML agreement are both lower for fields than for styles, at 76% and 74%, respectively.

[35] We compared the performance of a number of data representations including TF-IDF, dummies for each word, and sums of `word2vec` embeddings before settling on our chosen representation. Comparisons were performed using a 90-10 split-sample test, as elsewhere.

Table A1: Trunk Journals and Professional Associations

Discipline	Journal	ISSN	First Year Indexed	Association
Accounting	Accounting Review	0001-4826	1927	American Accounting Association
Anthropology	American Anthropologist	0002-7294	1901	American Anthropological Association
Computer Science	Journal of the ACM	0004-5411	1954	Association for Computing Machinery
Economics	American Economic Review	0002-8282	1911	American Economic Association
Finance	Journal of Finance	0022-1082	1951	American Finance Association
Management	Academy of Management Review	0363-7425	1983	Academy of Management
	Academy of Management Journal	0001-4273	1958	Academy of Management
Marketing	Journal of Marketing	0022-2429	1936	American Marketing Association
Mathematics	Annals of Mathematics	0003-486X	1884	Princeton University
Medicine	New England Journal of Medicine	0028-4793	1928	Massachusetts Medical Society
	Journal of the American Medical Association	0098-7484	1945	American Medical Association
Multidisciplinary Science	Science	0036-8075	1901	American Association for the Advancement of Science
	Proceedings of the National Academy of Sciences	0027-8424	1915	National Academy of Sciences
Operations Research	Operations Research	0030-364X	1956	Institute for Operations Research and the Management Sciences
Physics	Physical Review Letters	0031-9007	1958	American Physical Society
Political Science	American Political Science Review	0003-0554	1906	American Political Science Association
Psychology	Psychological Review	0033-295X	1901	American Psychological Association
	Psychological Science	0956-7976	1990	Association for Psychological Science
Public Health	American Journal of Public Health	0090-0036	1912	American Public Health Association
	American Journal of Epidemiology	0002-9262	1965	Society for Epidemiologic Research
Sociology	American Sociological Review	0003-1224	1936	American Sociological Association
Statistics	Journal of the American Statistical Association	0162-1459	1901	American Statistical Association

Note: Disciplines with more than one leading professional association (Medicine, Multidisciplinary Science, Public Health) or whose leading association has two flagship journals (Management) are assigned two trunk journals.

Table A2: Highly Cited Economics Articles' Fields and Styles

Author	Year	Journal	Title	Field	Style	Weighted Citation Share	Raw Citations
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Kahneman, Tversky	1979	Econometrica	Prospect Theory: An Analysis of Decision under Risk	Misc	Theoretical	0.071	1053
Heckman	1979	Econometrica	Sample Selection Bias as a Specification Error	Metrics	Metrics	0.064	969
Hausman	1978	Econometrica	Specification Tests in Econometrics	Metrics	Metrics	0.053	743
Lucas	1978	Econometrica	Asset Prices in an Exchange Economy	Misc	Theoretical	0.050	416
Dixit, Stiglitz	1977	American Economic Review	Monopolistic Competition and Optimum Product Diversity	Misc	Theoretical	0.048	742
Holmstrom	1979	Bell Journal Of Economics	Moral Hazard and Observability	Micro	Theoretical	0.043	447
Hall	1978	Journal Of Political Economy	Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence	Macro	Empirical	0.040	357
Mirrlees	1971	Review Of Economic Studies	An Exploration in the Theory of Optimum Income Taxation	PF	Theoretical	0.038	493
Akerlof	1970	Quarterly Journal Of Economics	The Market for 'Lemons': Quality Uncertainty and the Market Mechanism	Micro	Theoretical	0.038	551
Kydland, Prescott	1977	Journal Of Political Economy	Rules Rather Than Discretion: The Inconsistency of Optimal Plans	Macro	Theoretical	0.038	529
Hansen	1982	Econometrica	Large Sample Properties of Generalized Method of Moments Estimators	Metrics	Metrics	0.136	1013
Newey, West	1987	Econometrica	A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix	Metrics	Metrics	0.116	926
Lucas	1988	Journal Of Monetary Economics	On the Mechanics of Economic Development	Macro	Theoretical	0.113	985
White	1980	Econometrica	A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity	Metrics	Metrics	0.091	1112
Romer	1986	Journal Of Political Economy	Increasing Returns and Long-run Growth	Misc	Theoretical	0.090	849
Grossman, Hart	1986	Journal Of Political Economy	The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration	IO	Theoretical	0.084	665
Engle, Granger	1987	Econometrica	Co-integration and Error Correction: Representation, Estimation, and Testing	Metrics	Metrics	0.081	988
Cho, Kreps	1987	Quarterly Journal Of Economics	Signaling Games and Stable Equilibria	Micro	Theoretical	0.076	488
Rubinstein	1982	Econometrica	Perfect Equilibrium in a Bargaining Model	Micro	Theoretical	0.075	558
Milgrom, Weber	1982	Econometrica	A Theory of Auctions and Competitive Bidding	Micro	Theoretical	0.074	524
Fehr, Schmidt	1999	Quarterly Journal Of Economics	A Theory of Fairness, Competition, and Cooperation	Micro	Theoretical	0.090	791
Katz, Murphy	1992	Quarterly Journal Of Economics	Changes in Relative Wages, 1963-1987: Supply and Demand Factors	Labor	Empirical	0.087	530
Andrews	1991	Econometrica	Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation	Metrics	Metrics	0.085	503
Laibson	1997	Quarterly Journal Of Economics	Golden Eggs and Hyperbolic Discounting	Macro	Theoretical	0.083	458
Berry, Levinsohn, Pakes	1995	Econometrica	Automobile Prices in Market Equilibrium	IO	Empirical	0.073	410
Romer	1990	Journal Of Political Economy	Endogenous Technological Change	Dev	Theoretical	0.069	486
Hall, Jones	1999	Quarterly Journal Of Economics	Why Do Some Countries Produce So Much More Output Per Worker Than Others?	Macro	Empirical	0.069	506
Staiger, Stock	1997	Econometrica	Instrumental Variables Regression with Weak Instruments	Metrics	Metrics	0.069	594
Imbens, Angrist	1994	Econometrica	Identification and Estimation of Local Average Treatment Effects	Metrics	Metrics	0.069	351
Holmstrom, Milgrom	1991	Journal Of Law, Eco., And Org.	Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design	Micro	Theoretical	0.068	478
Melitz	2003	Econometrica	The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity	Intl	Theoretical	0.126	694
Fischbacher	2007	Experimental Economics	Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments	Misc	Empirical	0.099	840
Bertrand, Duflo, Mullainathan	2004	Quarterly Journal Of Economics	How Much Should We Trust Differences-in-Differences Estimates?	Metrics	Metrics	0.097	609
Christiano, Eichenbaum, Evans	2005	Journal Of Political Economy	Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy	Macro	Empirical	0.089	581
Acemoglu, Johnson, Robinson	2001	American Economic Review	The Colonial Origins of Comparative Development: An Empirical Investigation	Misc	Empirical	0.077	518
Smets, Wouters	2007	American Economic Review	Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach	Macro	Empirical	0.067	372
Bolton, Ockenfels	2000	American Economic Review	ERC: A Theory of Equity, Reciprocity, and Competition	Micro	Theoretical	0.065	536
Eaton, Kortum	2002	Econometrica	Technology, Geography, and Trade	Intl	Empirical	0.064	287
Kling, Liebman, Katz	2007	Econometrica	Experimental Analysis of Neighborhood Effects	Misc	Empirical	0.063	178
Chernozhukov, Hong, Tamer	2007	Econometrica	Estimation and Confidence Regions for Parameter Sets in Econometric Models	Metrics	Metrics	0.063	116
Bloom et al.	2013	Quarterly Journal Of Economics	Does Management Matter? Evidence from India	Dev	Empirical	0.074	36
Pavan, Segal, Toikka	2014	Econometrica	Dynamic Mechanism Design: A Myersonian Approach	Micro	Theoretical	0.059	36
Grubb, Osborne	2015	American Economic Review	Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock	Micro	Empirical	0.055	14
Lee, Lemieux	2010	Journal Of Economic Literature	Regression Discontinuity Designs in Economics	Metrics	Metrics	0.054	143
Arkolakis, Costinot, Rodriguez-Clare	2012	American Economic Review	New Trade Models, Same Old Gains?	Intl	Empirical	0.052	73
Mcrae	2015	American Economic Review	Infrastructure Quality and the Subsidy Trap	Misc	Empirical	0.046	8
Manzini, Mariotti	2014	Econometrica	Stochastic Choice and Consideration Sets	Micro	Theoretical	0.045	14
Maestas, Mullen, Strand	2013	American Economic Review	Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt	PF	Empirical	0.045	29
Andrews, Soares	2010	Econometrica	Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection	Metrics	Metrics	0.045	59
Ito	2014	American Economic Review	Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing	IO	Empirical	0.044	31

Note: This table lists the 10 most cited papers among those published in each decade based on top-6 weighted citation rates. Weighted citation rates (reported here as percentages) can be interpreted as the average across post-publication years of the weighted share of all citations from the journals on our economics journal list to each paper. Columns 5 and 6 show each article's field and style classification. Column 8 shows the raw citation count to papers on the list.

References

Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." The Journal of Economic Perspectives, 24(2): 3–30.

Angrist, Joshua, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. 2017. "Economic Research Evolves: Fields and Styles." American Economic Review, 107(5): 293–297.

Backhouse, Roger, and Béatrice Cherrier. 2014. "Becoming Applied: The Transformation of Economics After 1970." The Center for the History of Political Economy Working Paper Series, No. 2014-15.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." Econometrica, 63(4): 841–890.

Biddle, Jeff E., and Daniel S. Hamermesh. 2017. "Theory and Measurement: Emergence, Consolidation and Erosion of a Consensus." History of Political Economy, forthcoming.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python. Sebastopol, CA:O'Reilly Media, Inc.

Bishop, Christopher M. 2006. Pattern Recognition and Machine Learning. Springer.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." Journal of Machine Learning Research, 3(Jan): 993–1022.

Breiman, Leo. 2001. "Random Forests." Machine Learning, 45(1): 5–32.

Card, David, and Stefano DellaVigna. 2013. "Nine Facts About Top Journals in Economics." Journal of Economic Literature, 51(1): 144–161.

Chen, Mengen, Xiaoming Jin, and Dou Shen. 2011. "Short Text Classification Improved by Learning Multi-Granularity Topics." Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 3: 1776–1781.

Cherrier, Beatrice. 2017. "Classifying Economics: A History of the JEL Codes." Journal of Economic Literature, 55(2): 545–579.

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." Journal of Political Economy, 113(1): 1–45.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." Econometrica, 70(5): 1741–1779.

Ellison, Glenn. 2002. "The Slowdown of the Economics Publishing Process." Journal of Political Economy, 110(5): 947–993.

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The Superiority of Economists." Journal of Economic Perspectives, 29(1): 89–114.

Hall, Robert E. 1978. "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence." Journal of Political Economy, 86(6): 971–987.

Hall, Robert E, and Charles I Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" The Quarterly Journal of Economics, 114(1): 83–116.

Hamermesh, Daniel S. 2013. "Six Decades of Top Economics Publishing: Who and How?" Journal of Economic Literature, 51(1): 162–172.

Hamermesh, Daniel S. 2018. "Citations in Economics: Measurement, Uses and Impacts." Journal of Economic Literature, forthcoming.

Hausman, Jerry A. 1978. "Specification Tests in Econometrics." Econometrica, 46(6): 1251–1271.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." Econometrica, 47(1): 153–161.

Hofbauer, Josef, and Karl Sigmund. 2003. "Evolutionary Game Dynamics." Bulletin of the American Mathematical Society, 40(4): 479–519.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." Econometrica, 47(2): 263–291.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." American Economic Review, 101(6): 2590–2615.

Katz, Lawrence F, and Kevin M Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." The Quarterly Journal of Economics, 107(1): 35–78.

Kelly, Michael A, and Stephen Bruestle. 2011. "Trend of Subjects Published in Economics Journals 1969–2007." Economic Inquiry, 49(3): 658–673.

Kim, Yoon. 2014. "Convolutional Neural Networks for Sentence Classification." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1746–1751.

Leydesdorff, Loet. 2004. "Top-down Decomposition of the Journal Citation Report of the Social Science Citation Index: Graph- and Factor-analytical Approaches." Scientometrics, 60(2): 159–180.

Lynn, Freda B. 2014. "Diffusing Through Disciplines: Insiders, Outsiders, and Socially Influenced Citation Behavior." Social Forces, 93(1): 355–382.

Moody, James, and Ryan Light. 2006. "A View from Above: The Evolving Sociological Landscape." The American Sociologist, 37(2): 67–86.

Morales, Jose Ignacio Velarde. 2017. "Classification of Economics Research Articles." Manuscript, MIT Department of Economics, August.

Panhans, Matthew T, and John D Singleton. 2015. "The Empirical Economist's Toolkit: From Models to Methods." Center for the History of Political Economy Working Paper No. 2015-03.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. 2011. "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12: 2825–2830.

Pieters, Rik, and Hans Baumgartner. 2002. "Who Talks to Whom? Intra- and Interdisciplinary Communication of Economics Journals." Journal of Economic Literature, 40(2): 483–509.

Piketty, Thomas. 2014. Capital in the 21st Century. Cambridge, MA:Harvard University Press.

Wu, Ho Chung, Robert Wing Pong Luk, Kam Fai Wong, and Kui Lam Kwok. 2008. "Interpreting TF-IDF Term Weights as Making Relevance Decisions." ACM Transactions on Information Systems (TOIS), 26(3): 13.

Zingales, Luigi. 2013. "Preventing Economists' Capture." In Preventing Regulatory Capture: Special Interest Influence and How to Limit It. , ed. Daniel Carpenter and David Moss, Chapter 6. Cambridge, England:Cambridge University Press.