AEA Papers and Proceedings 2020, 110: 42–48
https://doi.org/10.1257/pandp.20201058

EMPIRICAL PRACTICE IN ECONOMICS: CHALLENGES AND OPPORTUNITIES‡

Technology and Big Data Are Changing Economics: Mining Text to Track Methods†

By Janet Currie, Henrik Kleven, and Esmée Zwiers*

The last 40 years have seen huge innovations in computing technology and the availability of all types of data. The torrent of new data is sweeping the field of economics in specific directions that we attempt to document. We argue that new data often requires new methods, which in turn can inspire new data collection.

I. Data and Methods

Our data come from two sources: the first is all papers in the National Bureau of Economic Research (NBER) working paper series between January 1, 1980 and June 30, 2018, and the second is all papers published in the "top five" economics journals (*American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*) between January 1, 2004 and August 2019. Because our focus is on the ways that new data and methods are changing economics, we focus on applied microeconomics. For top-five papers, we use the JEL codes corresponding to applied microeconomic fields as suggested by Card and DellaVigna (2013) with the addition of category I3 for Welfare, Well-Being, and Poverty and category Q for Environmental Economics. For the NBER working papers, we include papers in the following programs: Aging, Children, Development, Education, Health Care, Health Economics, Industrial Organization, Labor Studies, Political Economy, Public Economics, International Trade, and Environment and Energy. We end up with a sample of 2,830 top-five papers and 10,324 NBER working papers. See the online data Appendix for more details.

Figure 1 shows that the number of top-five papers has increased over time, especially since 2008, and that the number of applied microeconomics papers has increased even faster. As a result, the fraction of papers published in major general interest journals that are classified

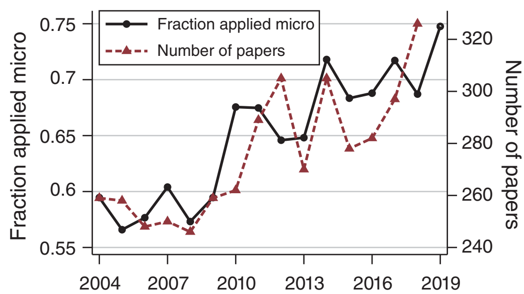

Figure 1. Applied Microeconomics Articles in Top-Five Journals

Note: This figure shows the fraction of papers in top-five journals that report an applied microeconomics JEL code (left axis) and the total number of papers in the top-five journals (right axis).

‡*Discussants:* Lawrence Katz, Harvard University; Guido Imbens, Stanford University; Benjamin F. Jones, Northwestern University.

* Currie: Department of Economics, Princeton University (email: jcurrie@princeton.edu); Kleven: Department of Economics, Princeton University (email: kleven@princeton.edu); Zwiers: Center for Health and Wellbeing, Princeton University (email: ezwiers@princeton.edu). We are grateful to Lawrence Katz for helpful comments. We thank Dana Scott for outstanding research assistance and Tilmann Herchenroder for excellent research assistance in the early stages of the project.

† Go to https://doi.org/10.1257/pandp.20201058 to visit the article page for additional materials and author disclosure statement(s).

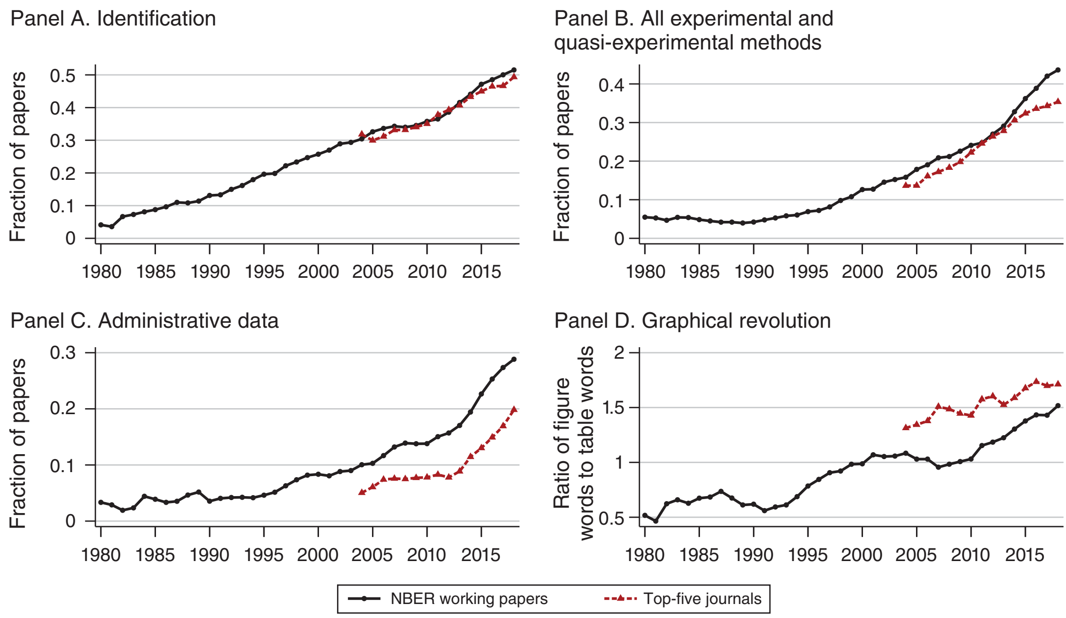

FIGURE 2. THE CREDIBILITY REVOLUTION

Notes: This figure shows different dimensions of the "credibility revolution" in economics: identification (panel A), all experimental and quasi-experimental methods (panel B), administrative data (panel C), and the graphical revolution (panel D). Panel D shows the ratio of the number of "figure" terms to the number of "table" terms mentioned. See Table A.I for a list of terms. The series show five-year moving averages.

as applied microeconomics has risen from 55–60 percent to about 75 percent.

For each paper, we use a plain text version excluding references. We use regular expression (regex) searches in Python to find keywords and phrases. See online Appendix B for detail. Table A.I lists the search categories and the specific trigger phrases within each category. For most categories, we search for any instance of each trigger phrase. For example, for "event study" we search for the trigger phrases "event stud" and "event-stud" with a wild card at the end. The wild card ensures that we capture permutations such as "event studies" or "event-study specification." This search is not case sensitive, but for other categories and trigger phrases (for example, the phrase "DiD" in the difference-in-difference category), it is important that the search is case sensitive.

For some categories, we condition on using the word "data." That is, papers must mention the word "data" or a permutation of "data" at least once. An example would be "clustering." Conditioning on "data" reduces the likelihood of picking up false positives. Some categories involve more complex search instructions. To illustrate, for "survey data," we look for instances in which the words "survey" and "data" are both mentioned within two full stops. The search for "identification" is more involved, because we want to avoid counting papers that use permutations of the word "identification" in ways unrelated to causal research designs. Our search algorithms have been designed by trial and error, with the aim of minimizing the prevalence of type I and type II errors. However, our main focus is on trends rather than levels. Our methods are similar to those in Kleven (2018) and Brice and Montesinos-Yufa (2019).

II. Results

We document changes in applied micro methods by plotting time series of methods-related words and phrases since 1980 (for NBER working papers) and 2004 (for top-five papers). Figure 2 highlights different dimensions of the "credibility revolution" in economics (see

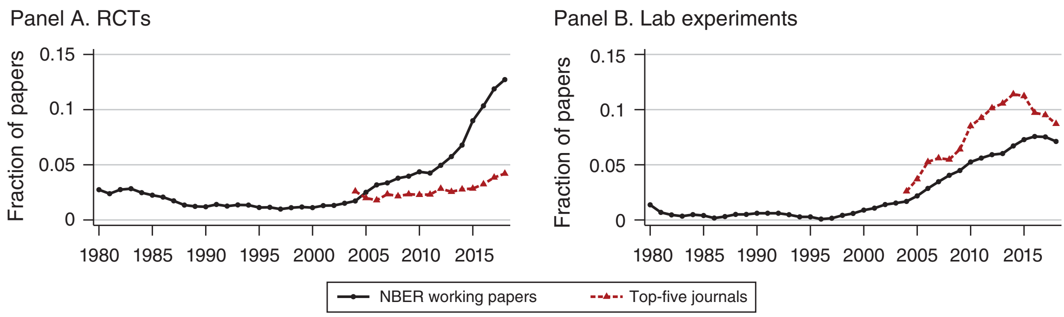

Figure 3. Experimental Methods

Notes: This figure shows the fraction of papers referring to each type of experiment. See Table A.I for a list of terms. The series show five-year moving averages.

Angrist and Pischke 2010). Panel A illustrates a linear rise in the fraction of papers making explicit reference to identification in both the NBER and top-five series. This fraction has risen from around 4 percent to 50 percent. Figure A.I in the online Appendix shows trends in specific identification concerns: omitted variables, selection biases, reverse causation, and simultaneity.

With this focus on cleaner identification has come a somewhat slower rise in experimental and quasi-experimental methods, illustrated in panel B. Currently, over 40 percent of NBER papers and about 35 percent of top-five papers make reference to randomized controlled trials (RCTs), lab experiments, difference-in-differences, regression discontinuity, event studies, or bunching.

Panel C shows a similar pattern in references to administrative data. The NBER series starts increasing in the mid-1990s, rising to about 30 percent today. The top-five series shows a similar increase, with a lag of about three years. Appendix Figure A.II considers alternative data sources. References to survey data still appear in over 50 percent of NBER working papers. References to proprietary data, internet data, and big data have increased over time, although the exact timing varies across categories. The term "big data" suddenly skyrocketed after 2012, with a more recent uptick in the top five.

Panel D depicts what we have called the "graphical revolution" in applied economics. It tracks the ratio of words about figures (charts, graphs, figures, and plots) to words about tables. This ratio has increased substantially over time and in two phases. The 1990s saw the diffusion of new software such as STATA that made it easier to create impactful figures. The second phase has happened in the last 10–12 years and is ongoing. This is likely due to the increasing use of administrative datasets, which lend themselves to compelling graphical representations. The top-five journals are more graphically oriented than the NBER working paper series. This could reflect either the type of paper that is selected into top-five journals or editorial decisions that favor graphical over tabular evidence. However, a matched sample of 610 NBER working papers that were published in top-five journals did not show large differences between the two versions.

The next two figures unpack the rise in experimental and quasi-experimental methods. Panel A of Figure 3 shows a sharp rise in the fraction of NBER working papers discussing RCTs since 2005, and especially since 2010. Since RCTs are particularly common in the field of development economics, the later rise is likely an artifact of the creation of the NBER program in development in 2012. The top-five series shows a much gentler rise in RCTs, to about 5 percent of papers by 2019. Panel B shows that laboratory experiments have grown steadily in popularity since the late 1990s, which is connected to the rise of behavioral economics during this time period (see Kleven 2018). Interestingly, more than 10 percent of top-five publications mention lab experiments in their peak year.

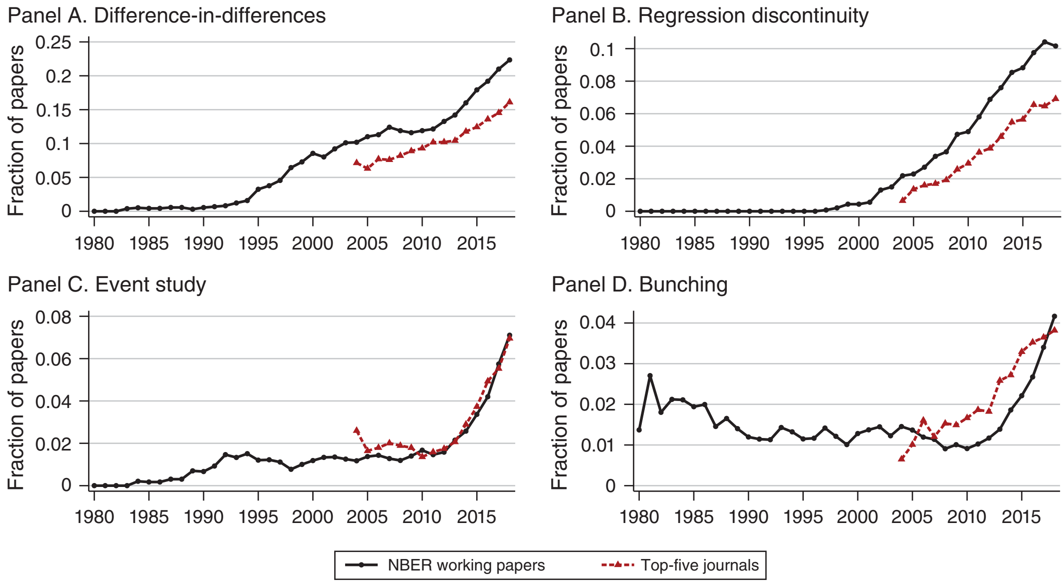

FIGURE 4. QUASI-EXPERIMENTAL METHODS

Notes: This figure shows the fraction of papers referring to each type of quasi-experimental approach. See Table A.I for a list of terms. The series show five-year moving averages.

The rise of experiments has generated its own critiques, such as concerns about external validity. Online Appendix Figure A.III shows that discussion of external validity began in the late 1990s and rose sharply in both the NBER and top-five series after 2005. This pattern mirrors the rise of lab and field experiments closely. One possible reaction to external validity concerns is to focus on mechanisms, thereby allowing readers to gauge whether estimates can be extrapolated to other settings. Online Appendix Figure A.IV shows an impressive rise in the fraction of applied micro papers discussing mechanisms, from about 20 to 60 percent in the NBER series. The fraction of top-five papers discussing mechanisms is even higher at more than 70 percent today, suggesting that editors either select papers that provide evidence on mechanisms or push authors to add such evidence as part of the editorial process.

Figure 4 drills down on specific quasi-experimental methods: difference-in-differences, regression discontinuity, event studies, and bunching. These methods have all become more popular over time, in roughly the order named. The use of difference-in-differences was virtually nonexistent until 1990 and then started growing. The first paper that mentions difference-in-difference estimators in our data is Ashenfelter and Card (1985), which appeared as a NBER working paper in 1984. It is quite striking that today almost 25 percent of all NBER working papers in applied micro make references to difference-in-differences. Regression discontinuity approaches started gaining popularity around 2000, following the early contributions reviewed in Hahn, Todd, and Van der Klaauw (2001).

Event studies and bunching approaches are more recent, having taken off during the last decade. Both of these approaches are closely linked to the new sources of large-scale data, since they are data intensive. Over time, it has become rare to use difference-in-differences without showing an event study graph, and conversely it is rare to show event studies without a control group. As a result, the sharp rise in the use of event studies over the last ten years goes hand in hand with the increased slope of the difference-in-difference series. The modern bunching approach starts with Saez (2010), although the NBER working paper version of that paper appeared more than ten years prior.

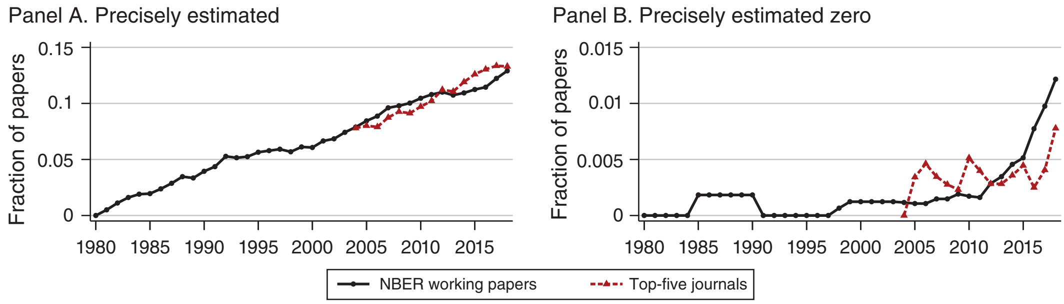

Figure 5. Precision

Notes: This figure shows the fraction of papers referring to each term. See Table A.I for a list of terms. The series show five-year moving averages.

One might wonder whether these new methods have grown at the expense of older empirical methods such as instrumental variables and fixed effects. However, online Appendix Figure A.V suggests that this is not the case. Mentions of instrumental variables and fixed effects have both grown continuously since the 1980s, while mentions of matching methods have been growing since the mid-1990s. Synthetic control methods are a relative newcomer, showing rapid growth in the NBER working paper series since 2010. The fact that old and new methods appear to be complements rather than substitutes suggests that another outgrowth of the credibility revolution is the rise of the "collage" approach to empirical work. Authors no longer hang their hats on a single method or dataset but attempt to make a case based on a more multipronged approach.

Figure 5 documents another implication of the rise in "big data" and more credible research designs, namely that authors have become increasingly concerned with whether their estimates are precisely estimated and not merely with whether they are significantly different from zero in a statistical sense. The focus on precision has grown continuously since the beginning of our series and is almost identical in the NBER and top-five series.[1] Panel B of Figure 5 highlights a specific dimension of this change: the concept of a "precisely estimated zero." This concept, virtually nonexistent until around 2010, has seen a sharp rise within the last decade. We view this as a positive development, because it holds the promise to reduce the publication bias that arises in a world where empiricists hunt for "large effects" and drop projects that show "no effects." We still have a ways to go: the idea that a precisely estimated zero could be as useful as a precisely estimated nonzero is only mooted in a little over 1 percent of papers.

Figure 6 considers a number of recent developments and innovations. Binscatter plots have become a more popular way of visualizing big data since 2010. Discussions of preanalysis plans have trended up sharply since 2012, when the American Economic Association voted to create a registry for them. Machine learning is the most popular among these brand-new methods, with mentions in 2.5 percent of NBER working papers. Text analysis like we do here has also become more common, with mentions in about 1 percent of NBER working papers in 2019.

Finally, Figures A.VIII–A.X in the online Appendix consider time trends in references to structural methods. Specifically, we focus on words and phrases related to structural models, general equilibrium, and specific functional forms (see Table A.I for details). These figures provide an example of a case where it is important to consider heterogeneity across subfields in applied micro. While there is no clear time trend in applied micro as a whole (see panel A of each figure), this masks relatively strong secular trends within specific subfields

[1] Related to this focus on precision, Figure A.VI in the online Appendix shows a sharp rise in references to confidence intervals since the mid-1990s. Figure A.VII shows that, after the year 2000, there has also been a massive increase in attention paid to the clustering of standard errors.

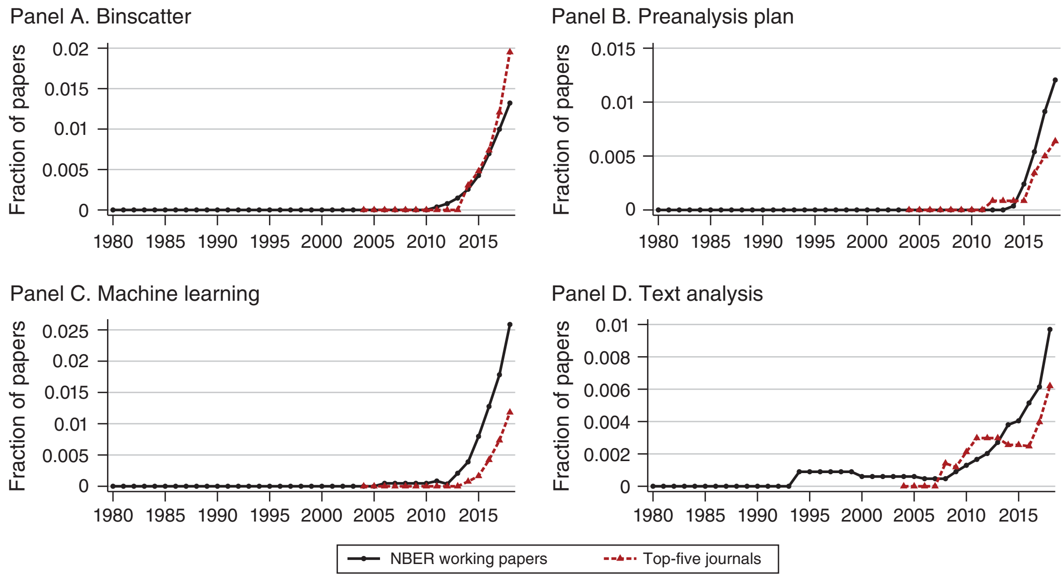

Figure 6. What's New?

Notes: This figure shows the fraction of papers referring to each method. See Table A.I for a list of terms. The series show five-year moving averages.

(see panels B–D of each figure). For example, structural models have been on the decline in labor economics, whereas they have been on the rise in public economics and industrial organization.

III. Conclusion

In the late 1960s and 1970s, new computing methods and data sources reshaped economics and made it a more applied field (Moffitt 1999). Fields like labor and public economics shifted their focus from theory and discussions of institutions and case studies toward estimating quantities such as returns to schooling and behavioral elasticities. Micro data from surveys in developing countries made it possible to focus on the determinants of household and individual well-being. Large-scale social experiments such as the negative income tax experiments and the RAND Health Insurance Experiment were conducted and evaluated for the first time. In turn, these new data prompted a flowering of interests in new methods such as selection and panel data approaches.

We may be at a similar turning point in the field today, with a proliferation of new data and methods. If history is any guide, some methods will stick and others will prove to be a flash in the pan. However, the larger trend toward demanding greater credibility and transparency from researchers in applied economics and a "collage" approach to assembling evidence will likely continue.

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." *Journal of Economic Perspectives* 24 (2): 3–30.

Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." *Review of Economics and Statistics* 67 (4): 648–60.

Brice, Brandon D., and Hugo M. Montesinos-Yufa. 2019. "The Era of Evidence." https://www.researchgate.net/publication/318600096_The_Era_of_Evidence.

Card, David, and Stefano DellaVigna. 2013. "Nine Facts about Top Journals in Economics." *Journal of Economic Literature* 51 (1): 144–61.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–09.

Kleven, Henrik Jacobsen. 2018. "Language Trends in Public Economics." https://www.henrikkleven.com/uploads/3/7/3/1/37310663/languagetrends_slides_kleven.pdf.

Moffitt, Robert A. 1999. "New Developments in Econometric Methods for Labor Market Analysis." In *Handbook of Labor Economics*, Vol. 3A, edited by Orley C. Ashenfelter and David Card, 1367–97. Amsterdam: North-Holland.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.

This article has been cited by:

1. Dmitry Arkhangelsky, Susan Athey, David A. Hirshberg, Guido W. Imbens, Stefan Wager. 2021. Synthetic Difference-in-Differences. *American Economic Review* **111**:12, 4088-4118. [Abstract] [View PDF article] [PDF with links]