

Why Economics Is Having a Replication Crisis

[bloomberg.com/view/articles/2018-09-17/economics-gets-it-wrong-because-research-is-hard-to-replicate](https://www.bloomberg.com/view/articles/2018-09-17/economics-gets-it-wrong-because-research-is-hard-to-replicate)

Noah Smith



Lab rats can't help.

Photographer: BSIP/Universal Images Group/Getty Images

Noah Smith is a Bloomberg Opinion columnist. He was an assistant professor of finance at Stony Brook University, and he blogs at Noahpinion.

[Read more opinion](#) [Follow @Noahpinion on Twitter](#)

[LISTEN TO ARTICLE](#)

5:33

By now, most people have heard of the replication crisis in psychology. When researchers try to recreate the experiments that led to published findings, only slightly more than half of the results tend to turn out the same as before. Biology and medicine are probably riddled with similar issues.

But what about economics? Experimental econ is akin to psychology, and has similar issues. But most of the economics research you read about doesn't involve experiments — it's empirical, meaning it relies on gathering data from the real world and analyzing it statistically. Statistical calculations suggest that there are probably a lot of unreliable empirical results getting published and publicized. Those results in turn drive policy.

And because the research involved is empirical rather than experimental, replicating it is hard — after all, you can't just rewind the world and run it again and collect new data. Instead, you can do one of three things:

First, you can go collect different data and try to do a similar analysis. This is useful, but it's not true replication, because conditions have changed. Suppose one author finds that a \$3 minimum wage hike in Los Angeles doesn't raise unemployment. If you look at a \$3 minimum wage hike in Houston and find that it does put some people out of work, that doesn't necessarily mean the Los Angeles study was badly done — it could just mean that the two cities are different.

A second type of pseudo-replication involves modifying authors' analysis in simple, reasonable ways, and checking whether the qualitative results still hold true. One famous case of this involved a 2000 paper by economist Caroline Hoxby, who found that in cities where rivers created more natural boundaries for school districts, students did better. She reasoned that competition among public schools caused the schools to improve — a finding with obvious implications for the school choice debate. But five years later, Jesse Rothstein tried to reproduce Hoxby's findings, and found that if he slightly changed the definition of how large a stream has to be to count as a river, the original finding vanished.

Another example is John Donohue and Steven Levitt's 2001 finding that abortion reduces crime. In 2005, Christopher Foote and Christopher Goetz used a different definition of per-capita crime rates, accounted for different state-level trends and corrected an error in Donohue and Levitt's code. The relationship between abortion and crime disappeared.

In both of these cases, the original authors vigorously defended their findings, leading to years of back-and-forth arguments. But both episodes show that many economics findings are dependent on murky research methods that readers rarely see.

A third kind of replication is simply to check whether the authors' own analysis can be repeated. This requires getting the data from the authors to see if you can get the same results by performing the exact same analysis. This is a pretty low bar, since if you use the same data and run the same statistical procedures, you should always get the same numbers.

Incredibly, this often doesn't work. In 2016, government economists Andrew Chang and Phillip Li tried to reproduce the results of 65 econ papers published in good journals. They got the original data, and even contacted the authors for help in following their footsteps. Yet still they only managed to reproduce 49 percent of the published findings — less than the fraction of successful replications reported in psychology or other disciplines, even though the number ought to have been much higher.

This strongly suggests that there is a lot of mystery meat that goes into economists' analysis. Scholars may add or remove different control variables, cut up their data into different slices, add or delete data, or modify their statistical analyses in a number of subtle ways. Any one of these can change the results.

The most dramatic demonstration of this was in 2013, when a paper by influential macroeconomists Carmen Reinhart and Kenneth Rogoff, alleging a correlation between high government debt and low growth, was challenged by a team of economists who discovered a spreadsheet error and questionable data-censoring practices. The study, which had been used to encourage austerity in the wake of the recession, is now widely viewed as discredited.

Fortunately, there are plenty of ideas for addressing the replication crisis in empirical economics. Economists Garret Christensen and Edward Miguel have a raft of suggestions for how economists can improve their research practices — pre-registering research plans, full and open sharing of data and code, more stringent statistical tests and the publication of “null” results. Jan Höffler and Thomas Kneib of the Institute for New Economic Thinking suggest that replicating papers should be an important part of graduate students’ education. That idea seems especially promising — not only will it harness a vast unexploited reservoir of talent toward the task of replication, but it will be an effective way of teaching students how to do their own research.

These cultural changes will take many years to become standard practice. In the meantime, economics writers and their readers are faced with the daunting task of deciding how much confidence to place in the results coming out of research in the field. The best strategy, as I see it, is strength in numbers — if a finding is confirmed by multiple teams using multiple data sets and methods of analysis, it’s inherently more reliable than if it relies on one paper only. Instead of treating empirical findings as breakthroughs, we should treat them as pieces of evidence that go into building an overall case.

That doesn’t mean that single results aren’t worth reporting or taking into account, but a single finding shouldn’t be enough to generate certainty about how the world works. In a universe filled with uncertainty, social science can’t progress by leaps and bounds — it must crawl forward, feeling its way inch by inch toward a little more truth.