Replications in Development Economics

Sandip Sukhtankar*

January 6, 2017

PRELIMINARY DRAFT, PLEASE DO NOT CITE

Abstract

I examine replications of empirical papers in development economics (JEL code "O") published in the top-5 and next-5 general interest journals between the years 2000 through 2015 (inclusive). Of the 1056 papers, 57 papers (5.4%) were replicated in another published paper or working paper. The strongest predictor of whether a paper is replicated or not is the paper's Google Scholar citation count, followed by year of publication (older papers are replicated less often). Papers based on randomized control trials (RCTs) appear to be replicated at a higher rate (12.5%). Most of the replications involve verification of the original paper's analysis, as well as some robustness checks and extensions. It is likely that many replications are not actually published or put out as working papers, given the many top graduate programs that require students to replicate analysis as part of a class.

JEL codes: B41, C81, O10
Keywords: replications, development economics

*University of Virginia, Monroe Hall, Charlottesville, VA 22904; sandip.sukhtankar@virginia.edu. I thank Lucas Coffman and Brian Nosek for helpful discussions, and Hanna Charankevich and Mia Giurato for excellent research assistance.

Replications in the social sciences have garnered considerable attention lately. In psychology, the "Many Labs" and Open Science Collaboration projects involved 270 researches attempting to replicate 100 studies published in the top three journals (Open Science Collaboration, 2015). In economics, Camerer et al. (2016) attempted to replicate 18 papers in experimental economics. To aid efforts in replication, all of the "Top 5" journals in economics have now adopted a Data Access Policy, requiring authors to make available data and code to all researchers for the purposes of replication.[1]

In this paper, I calculate the replication rate for papers in the field of development economics published since the year 2000 in one of the top 5 or next 5[2] general interest economics journals, and also examine the correlates of replicated papers. The issue of replications in development economics is particularly interesting for a number of reasons. First, there has been a marked increase in the number of empirical papers published in development, driven by advances in empirical methods, data availability, as well as the proliferation of Randomized Control Trials (RCTs) (Figures 1 and 2). Second, development economists work in a variety of contexts, with different institutional and market structures; whether results generalize across these various contexts is thus important to determine. Finally, the increase in the ease of conducting RCTs offers the hope that direct replication of interventions and tests of theories across contexts is possible.

In order to determine the replication rate, one has to define what is meant by "replication." Social psychologists have proposed a conceptual definition: "direct replication is the attempt to duplicate the conditions and procedure that existing theory and evidence anticipate as necessary for obtaining the effect" (Nosek and Lakens, 2014). Economists have come up with various working definitions: for example, Clemens (2015) separates replications from robustness tests, classifying "verification" (replication) as those using the same sample, population, and empirical specification; "reproduction" (replication) as those using different samples from the same population but using the same specification; "reanalysis" (robustness) using the same sample and population but different specifications; and "extension" (robustness) using different samples and populations but the same specification. Hamermesh (2007) separates "pure replications" (same methods, same data) from "statis-

[1]The *American Economic Review* was the first to adopt this policy in 2005. It states that "It is the policy of the American Economic Association to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication." The *Quarterly Journal of Economics* and the *Journal of Political Economy* have recently adopted this wording verbatim; *Econometrica* uses similar wording. The *Review of Economic Studies* has a DAP but it does not make clear whether data need to be made available only to the journal or to other researchers too.

[2]These are the *American Economic Journal: Applied Economics (AEJAE)*, the *American Economic Journal: Economic Policy (AEJEP)*, the *Economic Journal (EJ)*, the *Journal of the European Economic Association (JEEA)*, and the *Review of Economics and Statistics (ReStat)*.

tical replications" (alternative data or methods) and "scientific" replications (alternative theoretical or conceptual approaches). While seemingly incongruent, both seem to agree on what is "pure replication" or "verification"; and the "statistical replication" of Hamermesh (2007) seems to correspond to the "robustness" categories defined by Clemens (2015).

For the purposes of this paper, I consider all papers conforming to any of the four classifications by Clemens (2015), and the "pure" and "statistical" replications as defined by Hamermesh (2007), as replications. In addition, when calculating the "replication rate", I only consider whether a paper was replicated or not; I do not consider whether the paper *successfully* replicated the analysis.

1 Data

Using Econlit, I searched for every paper published in the above-referenced ten journals that were published between the years 2000 and 2015 (inclusive) and had an "O" Journal of Economic Literature classification.[3] I dropped papers that were themselves replications or comments. I classified the remaining papers into those that were purely theoretical and those that contained some empirical analysis, and the empirical papers into those that involved an RCT or not. This yielded a total of 1,265 papers, with 209 pure theory papers, 120 RCTs, and 936 empirical papers not involving RCTs (Table 1).

Figures 1 and 2 show the evolution of numbers of papers in each of the categories over the period I study. Contrary to perceptions, theoretical papers have not been crowded out by empirical papers and RCTs; the number of theory papers is more or less constant over the years.[4] The number of empirical papers does seem to have increased significantly, driven largely by the increase in papers published based on RCTs. The introduction of the *American Economic Journals* in 2009 also seems to have driven much of this change.

For the empirical papers in my sample, I searched for published or working papers that attempted to replicate the empirical analysis or intervention. The main method involved doing a reverse citation search on Google Scholar, and searching within the ensuing list for "replication" or alternative cognates. I supplemented this method by examining websites of various projects dedicated to replications in economics. The Online Appendix describes data and methods in detail.

[3]The current numbers do not include papers with solely O4 or O5 categorizations; the analysis for these papers is complete and does not qualitatively change any of the results highlighted here.

[4]Note that my sample does not include the *American Economic Journal: Microeconomics*, which mainly publishes theoretical papers; hence it is possible that the number of theoretical papers published has also increased.

2 Results

Of the 1,056 empirical papers, I found that 57 (5.4%) were replicated in a published or working paper. The number drops to 29 (2.7%) if considering only already published papers. RCTs seem to be replicated at a higher rate, with 15 of the 120 RCTs (12.5%) being replicated.[5]

The majority of the replications are verification tests as well as reanalyses in the same paper, or both pure and statistical replications. A few are extensions in the sense that a particular intervention is tried in a different context. A common theme is a reanalysis using a new econometric technique.

At face value the overall rate of replication seems low when compared to other fields like psychology, although it is not clear what the comparison is for other fields within economics. Whether scarce researcher resources should be devoted to increased replications is of course an entirely different although important question. In the rest of this section I restrict myself to analyzing the correlates of the current set of replicated papers.

A reasonable basis for determining whether a paper should be replicated or not might consider the impact of the paper as well as the uncertainty of the empirical analysis; i.e., influential papers with large standard errors should be replicated. While it is a task of considerable magnitude to determine the "uncertainty" of the empirical analysis in over a thousand papers, measuring influence via citations is a much more attainable task. I find that papers that are replicated are far more highly cited than papers that are not; on average, a replicated paper has 4 times as many Google Scholar citations as a non-replicated paper (897 vs 226). On this basis, the "right" papers are being replicated.

I run some very simple regressions in order to assess the statistical relevance of the above data, and examine the correlates of replications in more detail. Table 6 confirms that having a higher citation count as well as being an RCT are both strongly correlated with being replicated. It also reveals that the linear term for year of publication is significant, suggesting that more recent papers are more likely to be replicated. Being published in a "Top 5" journal by itself does not seem to be predictive after taking into account citation counts. The RCT and citation count results survive the inclusion of year and journal dummies, as well as a logistic rather than a linear probability specification.

[5]Five of the replicated RCTs are replicated by a single paper that uses an alternative statistical model for analysis; even without including these five, the rate of replication for RCTs is still higher, and statistically different from that of non-RCTs.

3 Discussion

While the reason for replications of papers with high citation counts is clear, why RCTs as well as more recent papers tend to be replicated more is less clear. One possible explanation is data availability. While the AEA adopted its DAP in 2005, other top journals have only adopted these policies as recently as 2015. For example, one paper that replicates multiple RCTs notes: "Due to the policies of the two journals that published these papers - the AEJ:Applied and Science - all the microdata from these RCTs is freely available online" Maeger (2016). In general, the donors that fund RCTs require authors to make data available for others to use, not simply for replication but to maximize their bang for the buck.

Other reasons for why newer papers are replicated more may simply have to do with recency bias. For example, the econometric theory papers that replicate and reanalyze empirical papers in order to make a methodological point seem to just pick a recent paper published in the AER that satisfies the criterion they need.[6]

As surmised in the introduction, RCTs do seem to allow for direct replication of interventions in different contexts. For example, the *AEJAE* dedicated an issue (January 2015) to six RCTs (five in developing countries) that each tested the impact of expansion of microfinance in various contexts. While the interventions were not identical, the basic theory being tested was the same. An example in which the identical intervention was tested in six countries is Banerjee et al. (2015), who examined an intervention that assisted the extreme poor to "graduate" out of poverty via sustainable self-employment opportunities.

What might be done to encourage more replications? Other papers in this session and elsewhere directly examine incentives to replicate. However, one source of data that is currently available but untapped is the set of replications undertaken by PhD students for coursework everywhere. Many top PhD program in economics have classes in which graduate students replicate papers and run robustness tests or extensions.[7] A recent blog post describes one such exercise in a macro course at the New Economic School, in which the professor asked students to replicate and test robustness for papers published in the AER.[8]

Indeed, Fecher, Frassdorf and Wagner (2016) argue that replication exercises should be a mandatory part of PhD coursework in economics. Such replications can at the very least serve as verification tests on the data and code used, but currently there is no way to access

[6]For example, Sloczynski (2014) replicates Martinez-Bravo (2014) in a paper titled "New Evidence on Linear Regression and Treatment Effect Heterogeneity," where the criteria for selecting this paper appear to be as simple as the ones outlined here.

[7]For example, I was able to precisely replicate the results in Abadie and Gardeazabal (2003) in a graduate course at Harvard in 2006.

[8]See `https://replicationnetwork.com/2016/12/27/campbell-is-the-aer-replicable-and-is-it-robust-eviden` for details.

them. A repository of these replications - for example through the Center for Open Science (`https://cos.io/`) - might help increase our confidence in currently published papers. Such a repository could provide incentives to graduate students to complete and write up replication efforts as papers, as well as incentives for authors to provide data more easily if these replications can assist in validating their results.

References

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*, 93(1): 113–132.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science*, 348(6236).

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu. 2016. "Evaluating replicability of laboratory experiments in economics." *Science*.

Clemens, Michael A. 2015. "The Meaning of Failed Replications: A Review and Proposal." *Journal of Economic Surveys*.

Fecher, Benedikt, Mathis Frassdorf, and Gert Wagner. 2016. "Perceptions and Practices of Replication by Social and Behavioral Scientists: Making Replications a Mandatory Element of Curricula Would Be Useful." IZA Discussion Paper 9896.

Hamermesh, Daniel S. 2007. "Viewpoint: Replication in economics." *Canadian Journal of Economics*, 40(3): 715–733.

Maeger, Rachel. 2016. "Understanding the Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of 7 Randomised Experiments." MIT.

Martinez-Bravo, Monica. 2014. "The Role of Local Officials in New Democracies: Evidence from Indonesia." *American Economic Review*, 104(4): 1244–87.

Nosek, Brian, and Daniel Lakens. 2014. "Registered reports: A method to increase the credibility of published results." *Social Psychology*, 45(3): 137–141.

Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science." *Science*, 349(6251).

Sloczynski, T. 2014. "New Evidence on Linear Regression and Treatment Effect Heterogeneity." IZA Discussion Paper 9491.

Online Appendix: Not for Publication

A Tables and Figures

Table 1: Breakdown of Paper Types

	RCT			
	-	No	Yes	Total
Empirical	0	936	120	1056
Theory	209	0	0	209
Total	209	936	120	1265

Table 2: Replication Status by RCT Status

	RCT					
	No		Yes		Total	
	#	%	#	%	#	%
Not Replicated	894	95.5	105	87.5	999	94.6
Replicated	42	4.5	15	12.5	57	5.4
Total	936	100.0	120	100.0	1,056	100.0

Table 3: "Cited By" Counts

	RCT		
	No	Yes	Total
Not Replicated	230.5	190.2	226.3
Replicated	1,085.0	372.1	897.4
Total	268.9	213.0	262.5

Table 4: Simple Replication OLS Regression

	(1)
	Replicated
year	0.00406*
	(0.00165)
Citation Count	0.000142***
	(0.0000143)
RCT	0.0755***
	(0.0213)
Paper is from top five journal	0.0136
	(0.0138)
Constant	-8.145*
	(3.309)
Observations	1056

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Replication OLS Regression with Year and Journal Dummies

	(1)
	Replicated
Citation Count	0.000143***
	(0.0000145)
RCT	0.0662**
	(0.0217)
Observations	1056

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Simple Replication Logit Regression

	(1) Replicated
year	0.0770
	(0.0424)
Citation Count	0.00149***
	(0.000271)
RCT	1.178***
	(0.345)
Paper is from top five journal	0.469
	(0.312)
Constant	-158.5
	(85.20)
Observations	1056

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Replication Logit Regression with Year and Journal Dummies

	(1) Replicated
Citation Count	0.00185***
	(0.000356)
RCT	1.029**
	(0.378)
Observations	963

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: Breakdown of Paper Types by Year

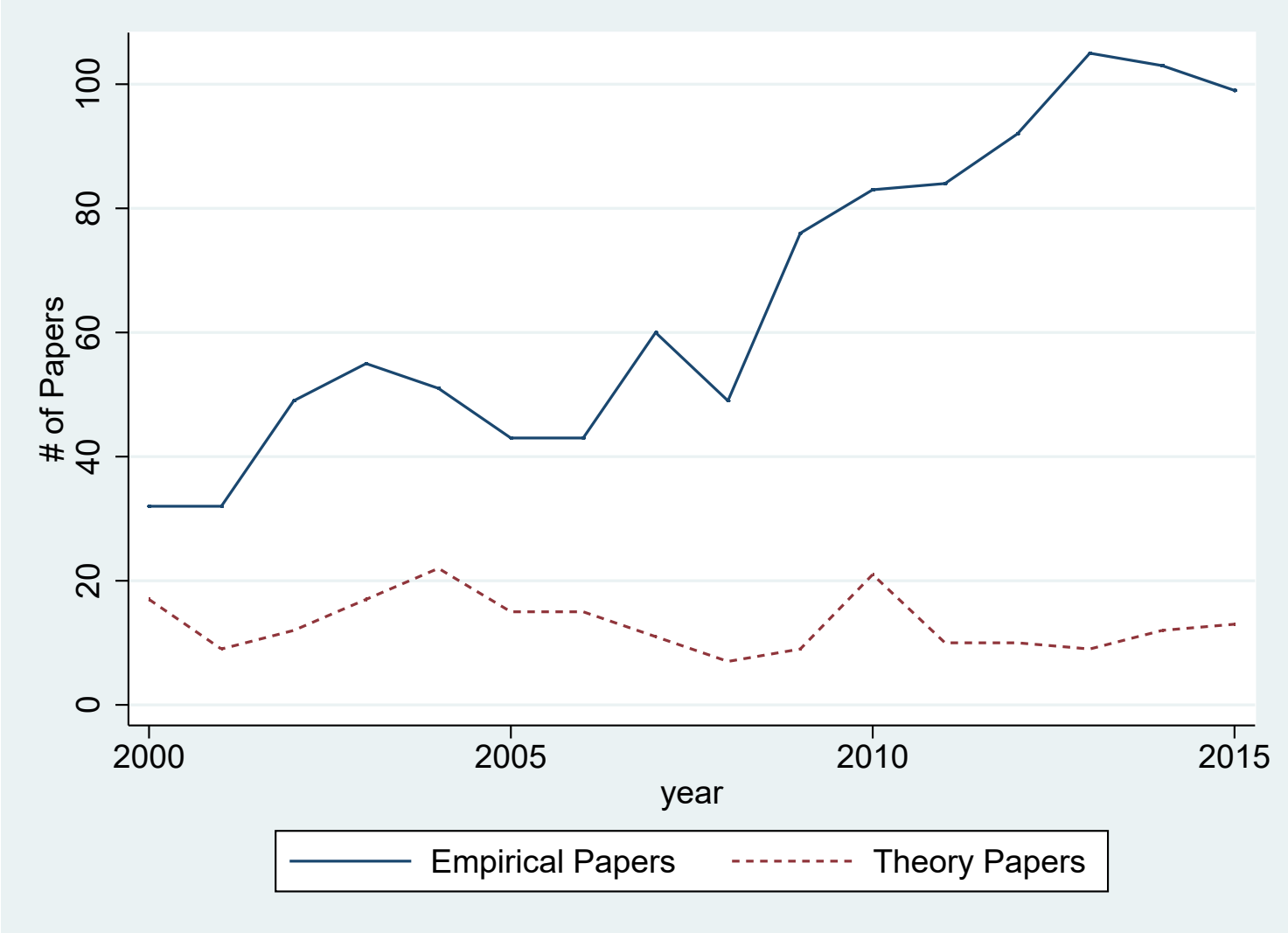


Figure 2: Quantity and Share of Empirical Papers which are RCTs by Year

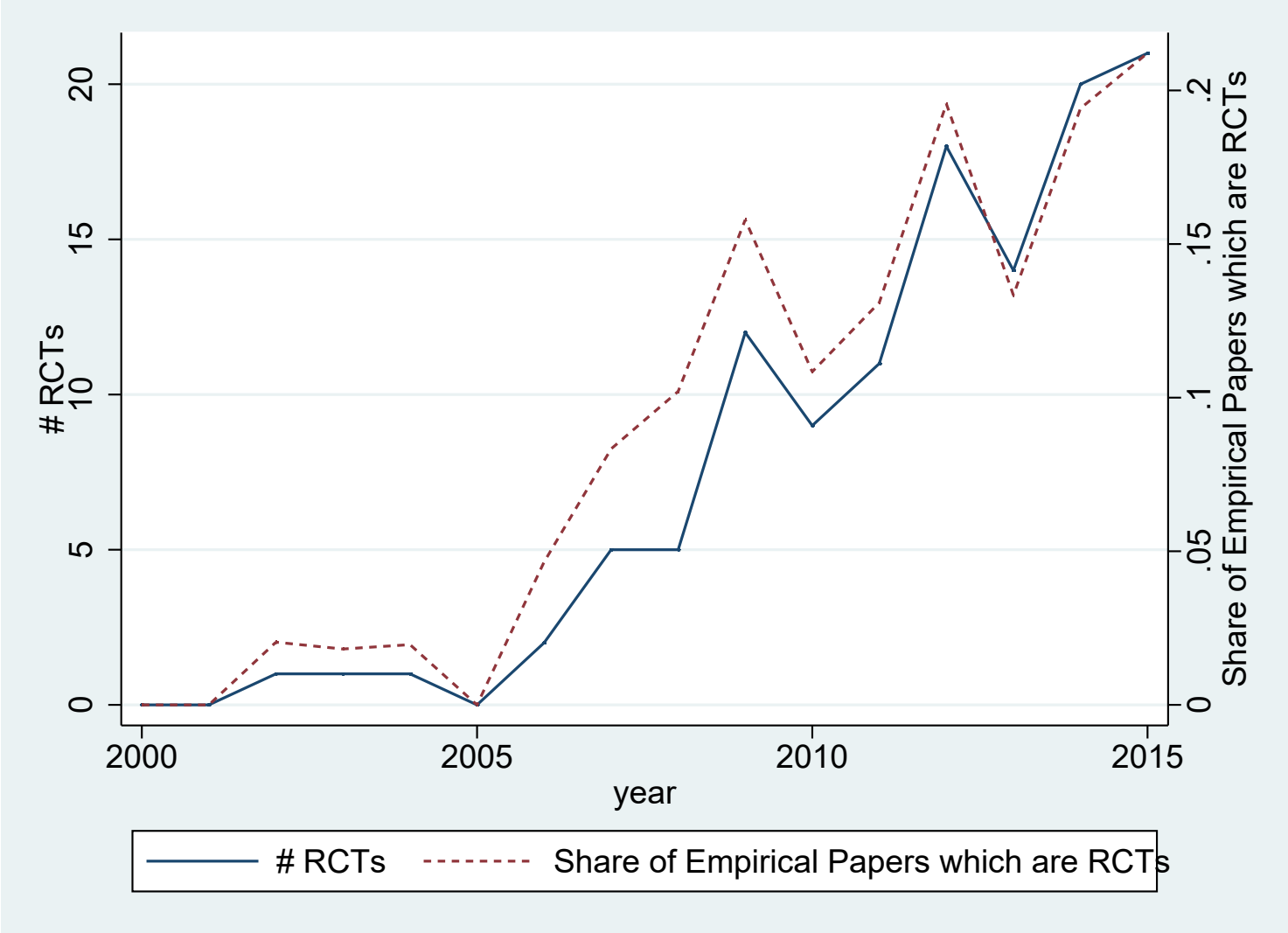


Figure 3: Share of Empirical Papers which are Replicated by Year

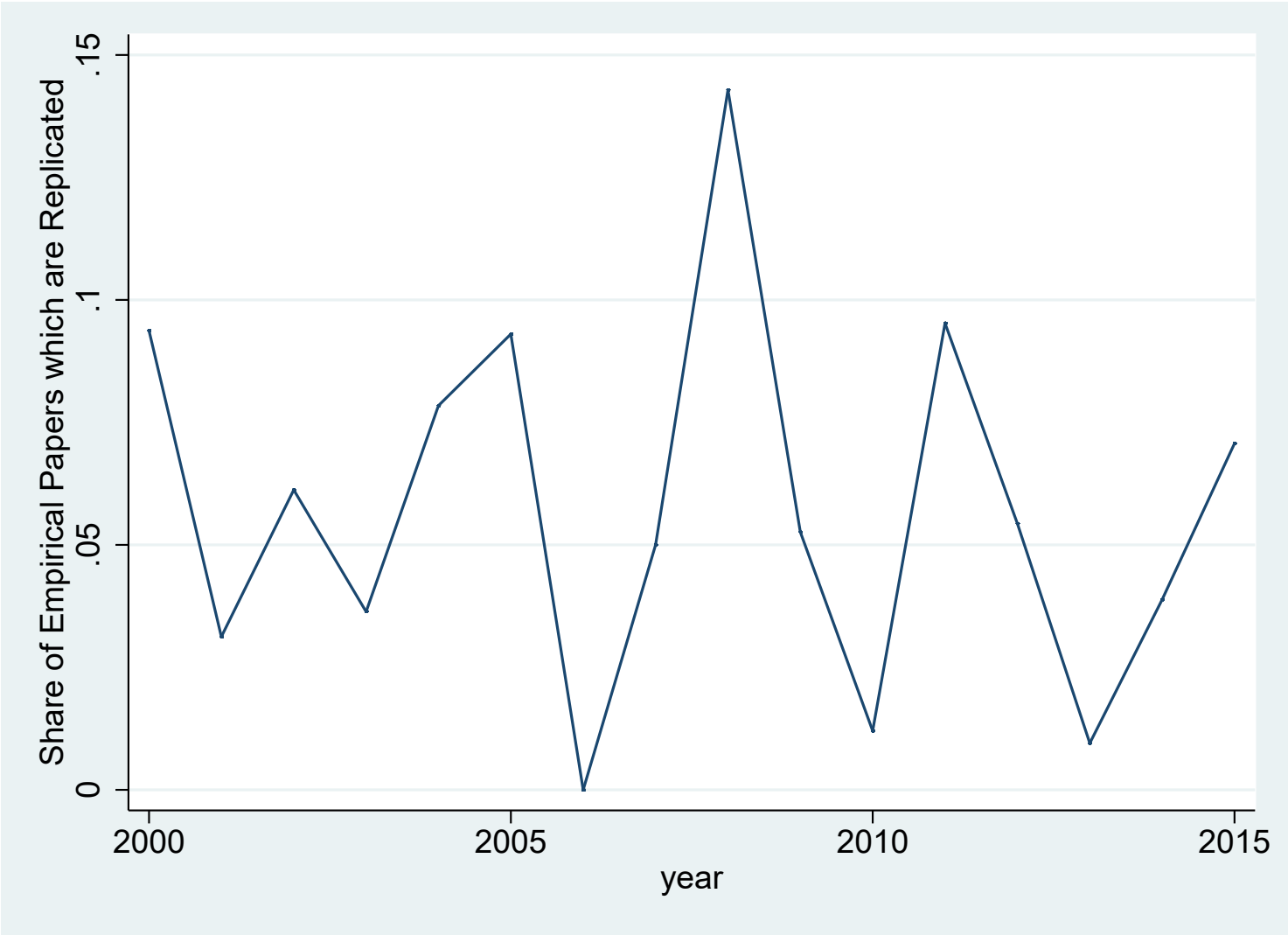
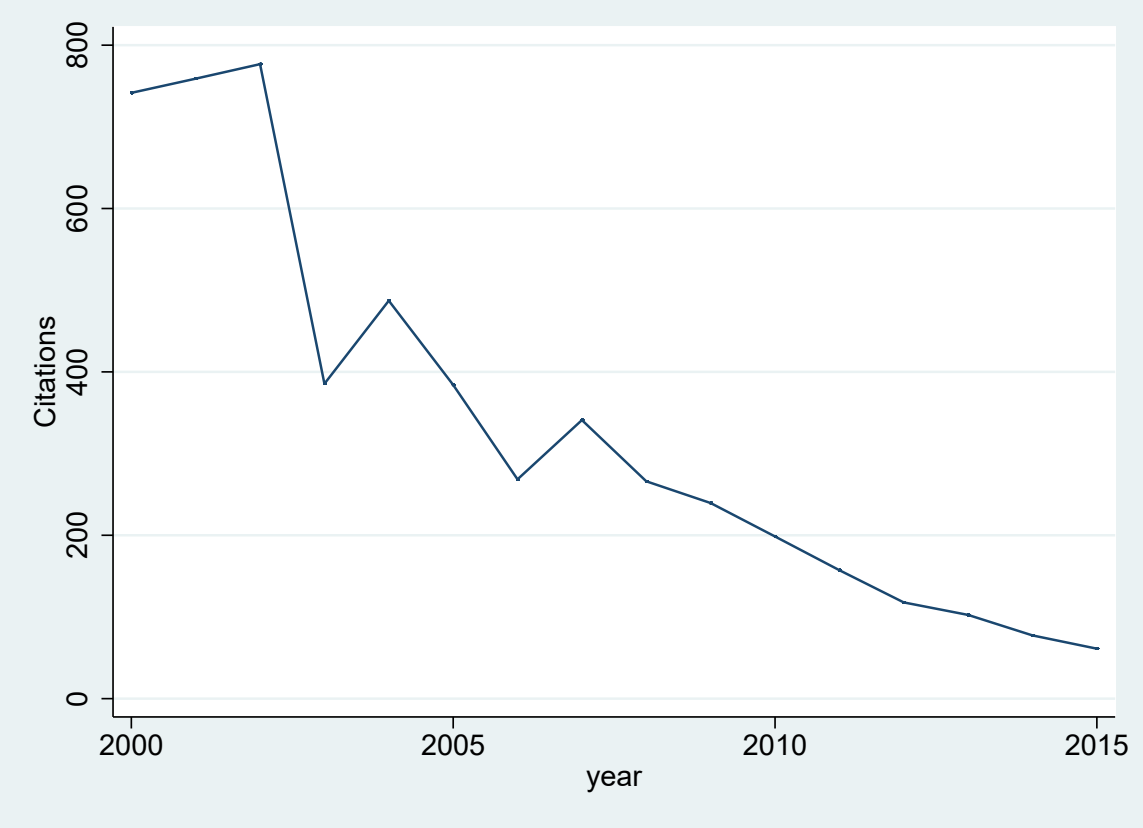


Figure 4: Average Number of Citations per Year for All Empirical Papers



B Data

A list of all papers satisfying the following conditions was compiled using EconLit:

1. The paper was published between 2000 and 2015 (inclusive)
2. The paper was published in one of the following journals
   (a) "Top five" journals
      - the American Economic Review (AER)[9]
      - Econometrica
      - the Journal of Political Economy (JPE)
      - the Quarterly Journal of Economics (QJE)
      - the Review of Economic Studies (ReStud)
   (b) Five other highly regarded general interest journals
      - the American Economic Journal: Applied Economics (AEJAE)
      - the American Economic Journal: Economic Policy (AEJEP)
      - the Economic Journal (EJ)
      - the Journal of the European Economic Association (JEEA)
      - the Review of Economics and Statistics (ReStat)
3. The paper had an "O" EconLit classification code[10]

Papers that were themselves comments or replications were dropped. Papers were then manually labeled as an empirical paper or a theory paper, and the empirical papers as RCT or non-RCT.

C Method for determining whether papers were replicated

I used the following methods to determine whether a paper classified as "empirical" in the list above was replicated, and to find the corresponding replication paper:

[9] All "Papers and Proceedings" papers are excluded.

[10] The current version does not include those with classification O4 (Economic Growth and Aggregate Productivity) and O5 (Economywide Country Studies) and their subcategories. These classifications was excluded since the initial idea in this paper was to focus on microeconomic studies of development. However, since a non-trivial number of papers have classifications that are both in O4 (or O5) and another O category, and these papers are similar to those in solely the O4 (or O5) category, I have decided to include all O papers, and am working on finding replications for the final draft of the paper.

1. I searched for each paper separately in Google Scholar, and then searched for the terms *replicate OR replicates OR replicated OR replication OR replicating* within the set of papers that cited this paper. The abstracts and introductions of papers satisfying both criteria were then examined to determine whether the citing paper did indeed attempt a replication of the original paper. The majority of replications were found using this method.

2. I supplemented the above method using the following websites to search for replications:

   (a) `http://replication.uni-goettingen.de/wiki/index.php/Main_Page`. The website allows you to search by JEL category, and I examined search results for every single "O" classification.

   (b) `https://replicationnetwork.com/`

   (c) `http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/`

Both published papers as well as unpublished working papers were counted as replication papers. If multiple versions of the same paper were found, the published version was included in the data.