This paper was commissioned for the Committee on Reproducibility and Replicability in Science, whose work was supported by the National Science Foundation and the Alfred P. Sloan Foundation. Opinions and statements included in the paper are solely those of the individual author, and are not necessarily adopted, endorsed, or verified as accurate by the Committee on Reproducibility and Replicability in Science or the National Academies of Sciences, Engineering, and Medicine.

Reproducibility and Replicability in Economics – FINAL DRAFT

White paper prepared for the National Academies' Committee on Reproducibility and Replication in Science

Lars Vilhuber[1]

2018-07-22

[1] Cornell University, lars.vilhuber@cornell.edu. The author is also affiliated with the U.S. Census Bureau and is the Data Editor of the American Economic Association. The opinions expressed herein are solely the author's, and do not represent the views or the policies of the U.S. Census Bureau or the American Economic Association.

1. Introduction

In this overview, I provide a summary description of the history and state of reproducibility and replicability in the academic field of economics. I will attempt to discuss not just the narrower definition of computational reproducibility, but also other correlates of intellectual reproducibility and transparency, such as the sharing of research findings outside of peer-reviewed publications ("grey publications"), and the importance of various types of data for empirical economics.

I start by defining reproducibility and replicability. Our focus is primarily on the journals that are the prime publication outlets of academic economists, and the role they have and can continue to play. Part of the reason for this focus is because it is much easier to measure replicability for published materials (even if what is being measured may change from study to study). Nevertheless, the informal and non-peer-reviewed sharing of documents, code, and data plays an important role in economics. I describe the historical context for journals and grey literature, review the historical roots of the use of pre-collected public and non-public data, and touch on the role of proprietary software in economics. This discussion frames the description of the state of reproducibility and replicability in modern economics, which here means within the last 30 years. I highlight the increasing importance of restricted-access data environments in economics and the interaction with reproducibility. In contrast, the role of replication, reproduction, and emulation in the teaching of economics is much harder to assess, though I will provide some indications as to its use in education. I then describe what is currently occurring in economics, touching on topics like big data and reproducibility in economics, or the search for the right method to surface reproductions and replications. Much of this is new, and the evidence on sustainability and impact is yet to be collected. Finally, I make an attempt at a conclusion

2. Definitions

In this text, we adopt the definitions of *reproducibility* and *replicability* articulated, inter alia, by Bollen et al (1). In economics, as in other sciences, a variety of usages and gradations of the terms are in use (2–5). At the most basic level, *reproducibility* refers to "to the ability […] to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator." "Use of the same procedures" may imply using the same computer code, or re-implementing the statistical procedures in a different software package, as made explicit in the notion of "*narrow replicability*" (5, 6). Reproducibility may be seen as analogous to a "unit test"[2] in software engineering.

Replicability, on the other hand, refers to "the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected," and *generalizability* refers to the extension of the scientific findings to other populations, contexts, and time frames. Because there is a grey zone between these two definitions, we will generally refer to either context as "replicability". Hamermesh (3) calls this "scientific replication." Robustness tests performed by researchers have aspects of self-replication, by identifying conditions under which the findings continue to hold when software or data are varied.

In this article, we will use the terms as defined above, even when authors use different terms.[3]

3. Historical Context

3.1. Replicability and Reproducibility in Early Economics

Publication of research articles specifically in economics can be traced back at least to the 1844 publication of the *Zeitschrift für die Gesamte Staatswissenschaft (8)*. The American Economic Association was founded in 1885, though initially most articles were not novel

[2] Unit testing is a concept from software engineering, where components of a larger piece of software are tested to ascertain that the software performs as intended.
[3] In fact, some define this terms in exactly the opposite way (7).

research reports (9). US-based journals that were founded at the time were Harvard's *Quarterly Journal of Economics* (1886) and the University of Chicago's *Journal of Political Economy* (1892), the Economic Journal (of the UK Royal Economic Society) in 1891 (8). However, publications by prominent economists had appeared in generalist academic journals prior to those initial issues (8). The modern-day *American Economic Review* followed in 1911 and the *Review of Economics and Statistics* in 1918. Of some significance in the context of replicability was the founding of *Econometrica* in 1933. As the first editor of *Econometrica*, Ragnar Frisch, noted, "the original data will, as a rule, be published, unless their volume is excessive […] to stimulate criticism, control, and further studies." (10). Most data at the time would have been published in paper form, and cited as such, as there would have been no distinction between "data" and "text" as we generally observe it today. However, editors in later years of *Econometrica*, as well as of the other journals, put rather less emphasis on this aspect of the publication process, whether by specialization - only 17.4% of articles in Econometrica in 1989-1990 had empirical content (8) – or for other reasons, is unknowable.

Much of economics was premised on the use of statistics generated by national statistical agencies as they emerged in the late 19th and early 20th century[4]. These were already so prevalent as a source of data, to be used in economic research and broadly sharable and shared, that the founding issue of the *Review of Economics and Statistics* explicitly precluded duplicating such collection and dissemination of data (12). At the same time, data sharing was easier: the same founding issue simply published tables of data as used by the author, both "original" and "computed" (13). There was, it can be argued, a greater similarity of "reproducibility" in theoretical economics (where proofs can be verified) and applied economics (where manual calculations, given the printed data, can be verified).

[4] For an interesting overview of the history of the U.S. Census Bureau, with reference to various other related agencies that produce public statistics, see (11)

The emphasis on statistical and empirical analyses increased, not just in the *Review of Economics and Statistics*, but also in the more technical *Econometrica* and the more generalist *American Economic Review* (9). By the late 1950s, the idea of even greater access to published and confidential government data by a large group of "data users", including economists, was well accepted (14, 15). This same period also saw the creation of archives, such as the Inter-University Consortium for Political Research (soon renamed as the inter-University Consortium of Political and Social Research, ICPSR), specifically designed to collect, convert, standardize and disseminate electronic records to academics, from surveys and other sources (16, 17).[5] Constraining wider dissemination was the ability to actually perform machine-based computations, as the Census Bureau and a few big universities were the only ones with sufficient compute power to actually leverage many of these data.

A key takeaway is that much of economic research relied on publicly available data. Initially, the data was just another form of (paper) publication, and thus easily identified and referenced by standard bibliographic citations. Later, with the advent of electronic records, a relatively small set of consortia and data providers were responsible for dissemination, by tapes, CD-ROMs and starting in the 1990s, FTP servers. While this should lead to relatively unambiguous data citations, this seems to not have been the case. As Dewald et al (19) note: "Many authors cited only general sources such as Survey of Current Business, Federal Reserve Bulletin, or International Financial Statistics, but did not identify the specific issues, tables, and pages from which the data had been extracted."

3.2. A History of Sharing Pre-Prints and Code

Economics has a history of sharing "grey literature" – documents such as technical reports, working papers, etc. that are typically not subject to peer-review (20), but are of

[5] Similar later efforts to consolidate and standardize dispersed electronic records, such as the U.S. censuses, the Current Population Survey, and others, led to the creation of the Integrated public use microdata series (IPUMS) in 1995 (18)

sufficient quality that they are worthwhile preserving (21), and in particular, worthwhile citing. Most scientists, when they think of pre-prints, think of arXiv (22, 23), founded in 1991. However, the first National Bureau of Economic Research (NBER) working paper, one of the most prestigious working paper series in economics, was published (in paper form) in 1973 (24). By the early 1990s, there was a wide variety of such working paper series, typically provided by academic departments and research institutions. Since grey literature at the time was not cataloged or indexed by most bibliographic indexes, a distinct effort to identify both working papers and the novel electronic versions grew from modest beginnings in 1992 at Université de Montréal[6] and elsewhere into what is today known as the Research Papers in Economics (RePEc) network, a "collaborative effort by hundreds of volunteers in 99 countries" (25–27). The initial index was split into electronic (WoPEc) (28) and printed working papers (BibEc) (26, 29), testimony to the prevalence of the exchange of scientific research in semi-organized ways. Economists had, in fact, access to a central repository for submitting working papers, based on the arXiv system, but it seems to not have been very popular, in contrast to the decentralized working paper archives (28). In 1997, BibEc counted 34,000 working papers from 368 working paper series (30). RePEc today has data from around 4,600 working paper series and claims about 2.5 million full-text (free) research items, provided in a decentralized fashion by about 2,000 archives (31). These items not only include traditional research papers, but also, since 1994, computer code (32–34). Although still cataloging mostly grey literature, RePEc bibliographic metadata is, in fact, indexed by all major bibliographic indexes.

[6] This author benefited greatly from those early cataloging efforts at Université de Montréal, where he commenced graduate studies in 1992, and often spent many hours in the physical collection of working papers collected by Féthy Mili, the Economics Department librarian and creator of BibEc.

3.3. The Increasing Importance of Non-Public Data

Economists have been using non-public data that they have not themselves collected at least as far back as Adam Smith's pin factory.[7] Economists were requesting access for research purposes to government microdata through various committees at least as far back as 1959 (14). Whether using private sector data, school district data, or government administrative records, from the U.S. and other countries, the use of these data for innovative research has been increasing in recent years. In 1960, 76% of empirical AER articles used public-use data.[8] By 2010, 60% used administrative data, presumably none of which is public-use (see Figure 1, reproduced from (36)). We will return to the effects of this phenomenon on reproducibility later, when discussing the effect of "data policies".

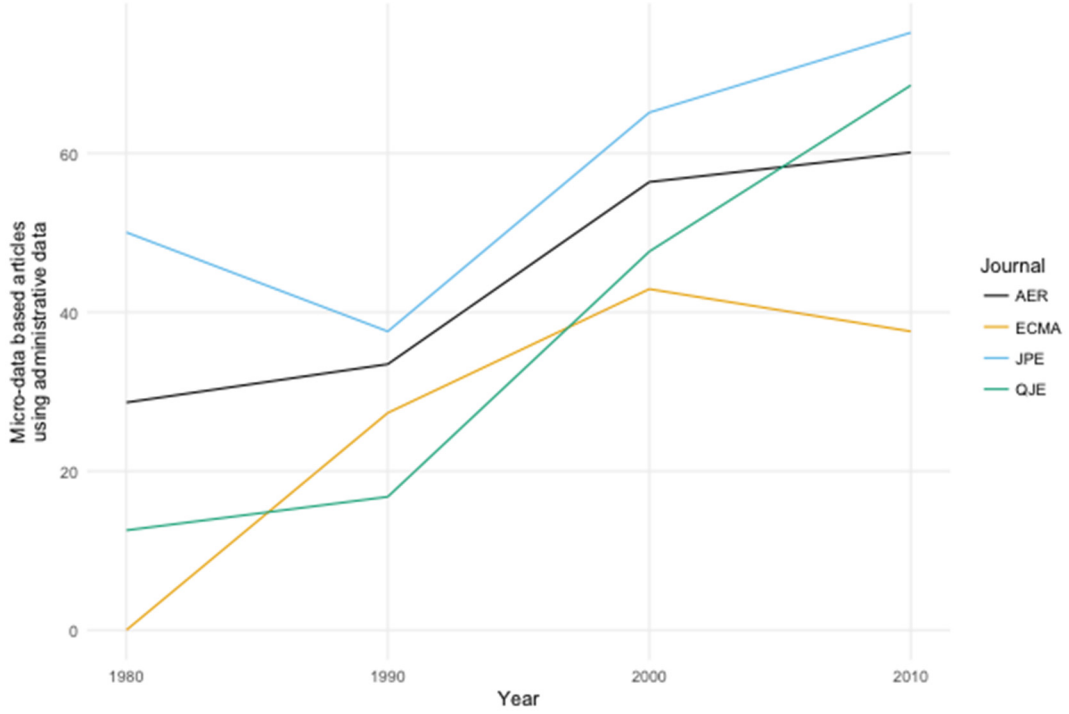

Figure 1: Use of Administrative Data in Publications in Leading Journals, 1980-2010

Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records).

[7] "I have seen a small manufactory of this kind where ten men only were employed […] make among them about twelve pounds of pins in a day." (35)

[8] See Appendix 1 for methodology and data for the AER, JPE, and ReStat.

Sample excludes studies whose primary data source is from developing countries. Figure reproduced from (36).

3.4. Proprietary Software

Software is considered an important component of the reproducibility "package". Many economists have long been willing to (informally) share their custom code,[9] even if others are hesitant to do so. However, underlying this is a large dispersion in software tools, extending from Fortran 77 code to software instructions for popular (and typically proprietary) statistical software such as Minitab, SAS, SPSS, and Stata (released in 1985 for PCs). In particular Stata is very popular among economists (nearly all articles in the *AEJ: Applied Economics* use Stata,[10] though this is likely to overstate the prevalence of Stata in economics). Stata very soon had many of the trappings of today's open-access toolkit. The Stata Journal, where peer-reviewed add-ons for Stata are published, has a paywall, but the underlying programs can be installed for free and in source-code form by any user of Stata.[11] Additional open software archives are widely used and referenced (postings to Statalist since 1994, Statistical Software Components (SSC) archive[12] since 1997). Historically, software such as R (37), Python, Julia (38) have not been widely used by economists, although each has had an active economics community (see f.i. , 39).

4. Reproducibility and Replicability in Modern Economics

It is generally argued that the ability to replicate and validate scientific findings is an important, even critical part of the scientific method. When reproducing results, researchers can check for inadvertent errors, and code and data archives provide a basis for subsequent

[9] The author worked with algorithms and code shared informally among labor economists as a research assistant in the early 1990s, and informally, this seems to have been standard practice.
[10] For details, see Appendix TBD and unpublished manuscript.
[11] Prior to 2001, the Stata Technical Bulletin (STB) fulfilled the role of distributing software components.
[12] https://ideas.repec.org/s/boc/bocode.html, maintained by Christopher Baum, an economist at Boston College Department of Economics. Note that bibliographic information on SSC items is disseminated via RePEc; thus, it is one of the earlier examples of citable software components.

replications and extensions by others (40). In economics, complaints about the inability to properly conduct reproducibility studies, or about the absence of any attempt to do so by editors, referees and authors, can be traced back to comments and replies in the 1970s (see (19) for examples). Calls for better journal policies to support replicability were made (41). While the *Journal of Political Economy (JPE)* added a section to the journal for "verifications and contradictions" of papers published in the *JPE* between 1976-1987, this seems to not have been effective (19): only 36 notes were published, of which 5 were actually reproductions (7). The best-cited example was the imposition of a "data availability policy" by the *Journal of Money, Credit, and Banking (JMCB).* The subsequent analysis thereof (19) considered all papers published, accepted, or under review by the *JMCB* between 1980 and 1984, of which some had been published before the announcement of the new data policy in 1982. The results suggested several problem areas. Authors, even among those whose article was still under review, had lost the data, or did not respond to the request for data and code. The non-submission rate was 65% for articles published before the announcement of the data policy, 26% after the announcement. Resource constraints (and the complexity of undertaking some of the replications) led to only 8 replication attempts being made, of which 5 were successful.[13] Only few such systematic replication or reproducibility attempts were made in subsequent years. It was concluded that "there is no tradition of replication in economics" (7).

4.1. Journal Policies Supporting Reproducibility

In the early 2000s, as in other sciences (42), journals started to implement "data" or "data availability" policies. Typically, they required that data and code be submitted to the journal, for publication as "supplementary materials." The *JMCB* had re-implemented their policy in 1996, after a brief hiatus, and the *Journal of Applied Econometrics* has a data archive going

[13] There is no summary table of replications in (19). I classified their verbose descriptions into "success" or "failure" based on the extent of the replication success. Only 2 studies were perfectly replicated, others had some (minor) deviations, and 3 failed to replicate.

back to 1988. The American Economic Association announced its "data availability policy" in 2003, implemented it in 2004, and extended it to the new domain-specific journals in 2009-2012. The first data supplements appear in *Econometrica* in 2004. The *JPE* announced its policy in 2004 and implemented it in 2005 (see Table 1 for details and links). Depending on how the sample of journals is selected, between 8.1% and 29.5% of economics journals (43, 44) have a "data availability policy."[14]

Table 1- Journal Policies

Table 1 ("Table 1 – Policies.xlsx") about here

The policies of the top journals listed in Table 1 generally reflect the lessons learned from earlier experiences. Although typically called "data availability" policies, they are more accurately described as "data and code deposit" policies. Most attach an archive of both data and programs as "supplementary data" on the journal website, with only the two Harvard-based journals (QJE, ReStat) depositing materials on the (Harvard-based) Dataverse (45). A consequence of this treatment of data supplements as secondary digital objects is that they do not generally obtain their own digital identifiers, the exception being data supplements stored on Dataverse. Few allow the data to be independently explored or discovered. Few if any articles during this time period cite the data, even when the data and code objects have citable identifiers.

Journals in economics that have introduced data deposit policies tend to be higher-ranked even before introducing the more stringent policy (46), possibly biasing analyses that focus on high-ranked journals (47). None of the journals in Table 1 request that the data be provided before or during the refereeing process,[15] nor does a review of the data or code enter

[14] The two studies cited differ in their time frames (2013 and 2011, respectively), how economics journals are identified (denominator) and how data availability policies are counted (numerator).

[15] Personal correspondence in 2018 with several editors and co-editors of the American Economic Association's journals suggests that a very small number may, of their own accord, request and

the editorial decision, in contrast to other domains (48). All make provision of data and code a condition of publication, unless an exemption for data provision is requested.

Other journals have taken a more low-key approach, only requesting that authors provide data and code upon request post-publication. Studies old and new have found that the probability of obtaining sufficient data and code to actually attempt a reproduction is lower when no formal data or code deposit policy is in place (19, 49, 50). In a simple experiment we conducted in 2015, we emailed all 117 authors that had published in a lower-ranked economics journal between 2011 and 2013. The journal has no data deposit policy, and only requires that authors promise to collaborate. We sent a single request for data and code. Only 48 (41%) responded, in line with other studies of the kind (49), and of those, only 12 (10% of total requests) provided materials upon first request.[16] Others report 35.5% (19) and 42% (50), with different request protocols and article selection criteria in each case.

4.2. Reproducibility Studies

If the announcement and implementation of data deposit policies improves the availability of researchers' code and data (19, 51), what has the impact been on overall reproducibility? A journal's data deposit policy needs to be enforced and verified – absence thereof can still lead to low data and code availability: the non-submission rate amongst the 193 *JCMB* articles studied by (7) was 64%, even after the policy was theoretically in place (see Table 2, Panel A).

Table 2, Panel B shows the reproduction rates both conditional on data availability as well as unconditionally, for a number of reproducibility studies.[17] In our own analysis, as well as in (7), a census of all articles over a certain time period was undertaken, whereas (50) and (55)

successfully obtain data and code as part of the refereeing process. When doing so, they report excellent compliance.

[16] Refusals included "too complicated for you", "I no longer have access to the data at my prior employer", and sadly "the co-author with the data has passed away". The data and report are not yet available, to allow for a follow-up contact.

[17] For additional reproducibility studies, see (52–54)

selected specific articles under certain search criteria. The studies undertaken in the past three years find a higher conditional reproduction rate than (7) (between 49% and 61%).

Table 2 - Submission and Reproduction Rates

Table 2 about here – Rates.xlsx

4.3. The Importance and Impact of Restricted-Access Data

As the increase of non-public data in Figure 1 suggests, however, even if compliance among users of public-use data increases, it is possible for overall availability, and thus reproducibility, to decline. In the journals of the *AEA*, all authors complied with the policy, as evidenced by the various "Reports by the Editor" published each year by the *AEA* (56, 57), an improvement on earlier years (58, 59). However, as noted earlier, exemptions are given when restricted-access data is used in an article. In our analysis of all 157 articles appearing in *American Economic Journal: Applied Economics (AEJ:AE)* between 2009 and 2013, only 60% of articles have some data available – a lower percentage than in the original *JCMB* study (Table 2, Panel A). Note that exemptions are not clearly published or posted, and because all such papers are still required to provide the code used to process the confidential or proprietary data, most such papers still have a supplementary material ZIP file, but without data.

Data that is not provided due to licensing, privacy, or commercial reasons (often incorrectly collectively referred to as "proprietary" data[18]) can still be useful in attempts at reproduction, as long as others can reasonably expect to access the data. For instance, while confidential data provided by the Health and Retirement Survey (HRS) or through the U.S. Federal Statistical Research Data Center (FSRDC) cannot be posted to journal websites, hundreds if not thousands of researchers have gained secure access to these data over the

[18] The term "proprietary" refers to ownership. Many datasets limit access not because of they are "owned" by somebody, but because balancing concerns of privacy and access yields a non-open access solution in order to provide both. The reason for limiting access is thus not ownership - many data owners and custodians provide open access to their data.

years, and could potentially reproduce or replicate the published research.[19] We thus analyzed each of the papers in the *AEJ:AE* that did not provide data, and classified the data used into five categories. Administrative data could be provided by a "national" provider (a national statistical office or similar), a "regional" entity (a state or province), or a "local" entity (a school district, county, or other governmental institution). Private providers might be commercial (data for which access can be purchased, such as from Dun and Bradstreet, State Street, or Bureau van Dijk), or some other type. Table 3 Panel A tabulates the distribution of characteristics amongst the 2009-2013 *AEJ:AE* articles with non-public data. National data providers dominate in this journal, providing nearly 50% of all non-public data.

Providers will differ in the presence of formal access policies, and this is quite important for reproducibility: only if researchers other than the original author can access the non-public data can an attempt at reproducibility even be made, if it at some cost. We made a best effort to classify the access to the confidential data, and the commitment by the author or third-parties to provide the data if requested. For instance, a data curator with a well-defined, non-preferential data access policy would be classified under 'formal commitment'. The FSRDC or the German Research Data Center of the Institute for Employment Research (IAB) have such policies. If the author personally promises to provide access to the data, we further distinguished 'with commitment', where the author would engage a third party to provide access in a well-defined fashion, from 'no commitment', where the author would simply promise to work with replicator, without being able or willing to guarantee such access. Our ability to make this classification depends critically on information provided by the authors. Table 3 Panel B tabulates the results from that exercise. We could identify a formal commitment or process to access the data only for 35% of all non-public datasets.

[19] For an example of a recent discussion of reproducibility relying on restricted access French data, see (60–63).

Table 3 Characteristics of non-public data

Table 3 – non-public data about here

The results above on type and access mode of non-public data are derived from a single journal's articles, and should be interpreted with caution. A more generalized assessment is difficult to undertake, since no journal in economics provides consistent data or metadata on the mode of access.

It is worth pointing out the increase in the past two decades of formal restricted-access data environments (RADEs), sponsored or funded by national statistical offices and funding agencies. RADE networks, with formal, non-discriminatory, albeit often lengthy access protocols, have been set up in the United States (FSRDC) (64), Canada (65), Germany (66), France (67–69), and many other countries. Often, these networks have been initiated by economists, though widespread use is made by other social scientists and in some cases health researchers. Restricted-use agreements with physical shipment of data are being phased out in favor of remote access arrangements. The use of such arrangements is less common for private sector data, although certain initiatives have made progress (Institute for Research on Innovation and Science (IRIS) (70, 71), Health Care Cost Institute (HCCI) (72, 73), Private Capital Research Institute (PCRI) (74, 75). A novel method for unbiased and rules-based access to social media data has recently been proposed (76).

Some widely used datasets are accessible by any researcher, but the license they are subject to prevents their redistribution and thus their inclusion as part of data deposits. This includes non-confidential datasets from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID) at the University of Michigan and data provided by IPUMS at the Minnesota Population Center. All of these data can be freely downloaded, subject to agreement to a license. IPUMS lists 963 publications for 2015 alone that use one of its data sources. The typical user will create a custom extract of the PSID and IPUMS databases

through a data query system, not download specific datasets. Thus, each extract is essentially unique. Yet that same extract cannot be redistributed, or deposited at journal or any other archive. Within the last year, the PSID, in collaboration with ICPSR, has addressed this issue with the PSID Repository (77), which allows researchers to deposit their custom extracts in full compliance with the PSID Conditions of Use.

Commercial ("proprietary") data is typically subject to licenses which also prohibit redistribution. Larger companies may have data provision as part of their service, but providing it to academic researchers is only a small part of the overall business. Dun and Bradstreet's Compustat, Bureau van Dijk's Orbis, or Twitter's data are all used frequently by economists and other social scientists. But providing robust and curated archives of data as used by clients over 5 or more years is typically not part of their service.[20] Most researchers also do not think to include or request redistribution rights in the acquisition contract, or at a minimum, the right to provide some level of access for the purpose of reproducibility. Nevertheless, such agreements exist, but are often hard to find due to the opaque nature of the "supplemental data" package on journal websites.[21]

4.4. The Importance of Transparent Public Data

While I point out the potential impact on reproducibility that restricted-access data may have, it is worthwhile to point out that even when data is shareable, there are issues related to reproducibility and replicability. While reproductions that identify errors in programs *used* by the researchers (e.g. 78–80) are testimony to the power of reproducible research, studies that have

[20] In personal conversation, Twitter's data team was receptive to improving transparency and reproducibility of research using Tweets. A conversation with Bureau Van Dijk's sales personnel was less conclusive.

[21] For instance, Molinari and co-authors specifically foresaw the interest of potential replicators to access the confidential insurance company data they used for their 2013 article (69). The supplemental data package, provided on the AER's website as a ZIP file, contains a user agreement, allowing interested users to request access to the data in a secure environment at the Cornell (University) Institute for Social and Economic Research (CISER). This fact is not discoverable by searching on either the AER's or on CISER's website.

focused on errors in the production and appropriate use of public-use data are just as important. Widely used datasets, such as the Current Population Survey (CPS) and the American Community Survey (ACS), have a long history of use by academics, and have a vast amount of accompanying documentation. Thanks to methodology documentation, it has been possible to show that incorrect use of data can lead to misleading conclusions.[22] In some cases, previously undocumented errors in the data publication itself were discovered (82). These are examples of replication studies, but also of the need for adequate documentation of the data collection, cleaning, and dissemination. Many other public-use datasets, as well as most researcher-collected datasets lack the amount of documentation to support such transparency, critical for the downstream use of these data in research. For official statistics, the National Academies' workshop on "Transparency and Reproducibility in Federal Statistics" will publish a report later in 2018.

4.5. Reproducible Research in Academic Education

One of the more difficult topics to empirically assess is the extent to which reproducibility is taught in economics, and to what extent in turn economic education is helped by reproducible data analyses. The extent of the use of replication exercises in economics classes is anecdotally high, but I am not aware of any study or survey demonstrating this. Most empirical economists teaching graduate economics classes will ask students to reproduce or replicate one or more relevant articles, though few of these replications are ever systematically made public if replications are successful,[23] though many failed reproductions and replications may have triggered articles and entire theses, attesting to the publication bias in replications. The

[22] (81) show that the imputation methods of the CPS bias any conclusion with regard to earnings or wage gaps when the relevant variable (e.g. union status) is not part of the earnings imputation model, and point to prior well-cited studies that ignored that point (e.g., when measuring the union wage gap).

[23] Some recent examples are cited in (83). The seminal study of Dewald et al (19) makes note of one example.

most famous example in economics is, of course, the exchange between Reinhart and Rogoff, and graduate student Thomas Herndon, together with professors Pollin and Ash (84, 85).

It is worthwhile pointing out that the Canadian Research Data Center Network expedites access requests to confidential data for students, greatly facilitating work on masters and doctoral theses, and potentially opening the door to easier reproductions and replications using confidential data.

More recently, explicit training in reproducible methods (86, 87), and participation of economists in data science programs with reproducible methods has increased substantially, but again, no formal and systematic survey has been conducted.

5. Looking ahead

Many of the issues facing reproducibility and replicability in economics are not unique to economics, and affect many other of the empirical social and clinical sciences. I touch here on a few of these topics.

5.1. Citing Data

A contributor to transparent and reproducible use of data is the ability to cite data, and to do so with precision. While data citation standards are well-established (88–90), only recently have style guides at major economics journals provided a suggested data citation format. The Chicago or Harvard citation styles are generally followed, but as of the 15th edition, the *Chicago Manual of Style* does not provide strong guidance or examples on data citations (91). The American Economic Review now requires datasets to be cited, and provides a suggested data citation (92, 93) to supplement the Chicago style. Other journals, even when having an explicit reproducibility policy (94), do not provide guidance on how to cite datasets. The absence of a data citation policy also affects incentives to create reproducible research (54).

5.2. Big Data, Changing Data

Difficulties when citing data are compounded when the data is either changing, or is a potentially ill-defined subset of a larger static or dynamic databases. “Big data” have always posed challenges – see the earlier discussion of the 1950s-1960s demand for access to government databases. By nature, they most often fall into the “proprietary” and “commercial” category, with the problems that entails for reproducibility. However, beyond the (solvable) problem of providing replicators with authorized access and enough computing resources to replicate original research, even defining or acquiring the original data inputs is hard. When the original inputs cannot be identified, reproducibility is impossible, though replicability and generalizability exercises can still be undertaken. This is not just an issue with what is commonly referred to as “big data”, but also for more traditional very large datasets. IPUMS, as pointed out earlier, can only be accessed via a query interface, but without a method to clearly define the precise revision of the underlying database, the version of the query system used, and the exact query used, and without a mechanism to redistribute the resulting data extract, it is not feasible to reliably create reproducible analyses using IPUMS, though the analysis could still be replicable. The same generic problem affects many other systems, though some, like the PSID (95), allow for storage and later retrieval of the query parameters, and others, such as the Census Bureau’s OnTheMap (96), provide mechanisms for users themselves to store reusable query parameters.

The above examples involve large, but slowly evolving databases. However, in the presence of “big data”, including data from social media (Facebook, Twitter), data and possibly data schemas evolve quite rapidly, and the simple mechanisms that the PSID, IPUMS, and the Census Bureau use, fail. Data citation of such data sources remains an active research topic for institutions like the Research Data Alliance (97, 98), with no robust solution yet adopted.

While in theory, researchers are able to at least informally describe the data extraction and cleaning processes when run on third-party controlled-systems that are typical of big data,

in practice, this does not happen. An informal analysis of various Twitter-related economics articles shows very little or no description of the data extraction and cleaning process. The problem, however, is not unique to big-data articles – most articles provide little if any input data cleaning code in reproducibility archives, in large part because provision of the code that manipulates the input data is only suggested, but not required by most data deposit policies.

5.3. Registration of Trials, Analysis Plans, and Reports

Related to concerns about replicability, but primarily aiming to address issues of publication bias and selective reporting of outcomes, the pre-registration of research hypotheses, analysis plans, and trials has made inroads in economics. Formal trial registries are inspired by similar efforts in the medical sciences. An early implementation in economics was the J-PAL Hypothesis Registry (99). In 2012, the *American Economic Association* instantiated the AEA Randomized Control Trial (RCT) Registry (100), "as a source of results for meta-analysis; as a one-stop resource to find out about available survey instruments and data."[24] The AEA RCT Registry keeps track of IRB protocol and approval numbers. Since 2017, registrations are reviewed for compliance with minimal criteria. As of May 2018, nearly 1800 studies have been registered. Reference to the AEA RCT registry in published articles (of the AEA or elsewhere) has not been studied systematically.

Pre-analysis plans (PAP) offer similar benefits, without a particular focus on trials. Registries that allow to register both trials and PAPs include the Registry for International Development Impact Evaluations Registration (RIDIE, 101, 102), Evidence in Governance and Politics (EGAP, 103), and AsPredicted (104). The Open Science Framework (OSF) provides the ability to record snapshots of projects, providing a similar proof as formal registries (105). Registration in general is voluntary (in contrast to clinical trials), but is strongly encouraged by

[24] The AEA RCT Registry is managed by J-PAL and supersedes the J-PAL Hypothesis Registry.

some journals (such as the AER). Even without formal registries, several prominent articles have used PAPs to effectively frame their results, see (106) for some examples.

I note that several potential forms of PAPs are hiding in plain sight. For instance, by making time-stamped research grant proposals or research data access requests (for RADEs) public, researchers could use such routinely submitted and, in the case of RADEs, compulsory documents as a form of PAPs (107). Funders and data custodians could support such efforts, by implementing such functionality within their systems, or explicitly encouraging researchers to routinely submit proposal documents at the relevant registries. Given the prevalence of restricted-access datasets, such a mechanism would have a potentially large and positive impact. To the best of my knowledge, this is not widely used at present.

An under-appreciated tool that has most of the characteristics of PAPs is the use of validation and verification servers in combination with synthetic data (108–111). When using synthetic data, researchers build sophisticated models using data that is not guaranteed to provide the correct inferences (112). By submitting their code for validation, researchers are in effect submitting a PAP. Various U.S. statistical agencies as well as those in other countries (113, 114) have been experimenting with these methods, but have not been viewed as part of the toolkit addressing reproducibility.

Registered Reports carry the idea of pre-registration further, and condition the publication of an article only on the pre-specified analysis. Not only do the authors have no (significant) leeway in analyzing the data, but the editors and reviewers also cannot select publications based on the statistical results (115–117). Registered reports are intended to counter the publication bias in favor of "significant" results, and encourage replications regardless of outcomes. As of 2018, registered reports are uncommon in economics.[25]

[25] According to one source (116), *Work, Aging and Retirement* is the only economics-related journal using registered reports.

5.4. Published Replications

Registered reports are seen as a potential solution to obtain more published reproducibility studies (54). Because most reproducibility studies of individual articles “only” confirm existing results, they fail the “novelty test” that most editors apply to submitted articles (54). In one particular case, all papers in Volume 100 of the *AER* were analyzed in how many were referenced as part of replication or follow-on work (52). While partially confirming earlier findings that strongly cited articles will also be replicated (3), the authors found that 60% of the original articles were referenced in some sort of replication or extension work, but only 20% appeared in explicit replications. Of the roughly 1500 papers that cite the papers in the volume, only about 50 (3.5%) are replications, and of those, only 8 (0.5%) focused explicitly on replicating one paper. Out of roughly 2600 articles in the *AER* between 2004 and 2016, the ReplicationWiki (83) identifies 44 “Comments” as “replications” of some sort. A few journals have introduced specific sections for reproducibility studies, following the longtime lead of the *Journal of Applied Econometrics*. Some journals have had calls for special issues dedicated to specific replication studies (118).

Even rarer are studies that conduct replications prior to their publication, of their own volition. (119) predict the unemployment rate from Twitter data. After having written the paper, they continued to update the statistics on their website (120), thus effectively replicating their paper’s results on an ongoing basis. Shortly after release of the working paper, the model started to fail. The authors posted a warning on their website in 2015, but continued to publish new data and predictions until 2017, in effect demonstrating themselves that the originally published model did not generalize. Similarly, (121) present their original experiment, and their own failure to replicate the original results when conducting the experiment a second time.

5.5. Elevating the Importance of Data and Code Availability

Enabling easier publication of replication studies and reproductions is one approach that will likely enhance overall reproducibility of economic research. A complementary approach is to make the analysis of code and data associated with the research a part of the peer review process. The AEA recently appointed a data editor (122) with the task to review not just the data availability policy, but also the methods and procedures supporting the implementation of the policy.[26] One of the outcomes is likely to be an emphasis on pre-publication verification of data and code packages (123), and the timely availability of results of such tests to referees and editors, prior to final decisions on acceptance. Similar considerations are under way at the *Review of Economic Studies*.

At statistical agencies and RADEs, reproducibility and its interaction with access restrictions and the protection of confidentiality is being discussed. At the U.S. Census Bureau and the Canadian Research Data Centers, working groups are looking into how the visibility of reproducibility within secure research environments can be increased.[27] Most of the processes in place to ensure confidentiality actually imply reproducibility, since research results and methods are vetted by reviewers before being released to the public. It should thus be relatively straightforward to demonstrate reproducibility of studies that are conducted in these environments (107).

Multiple research institutions are not waiting for journals to implement more stringent criteria. For instance, projects at J-PAL that collect data with funding from their research initiatives are subject to a data availability policy (124), and all J-PAL affiliated researchers are encouraged to publish their datasets in a Dataverse. Some research institutions offer a "code

[26] Disclosure: I am that data editor. As July 2018, no changes had yet been announced regarding the policy or procedures, but are expected shortly.

[27] Disclosure: I am involved in both of those working groups, either as a member, or as an outside expert. Reproducibility has also been discussed at the Scientific Advisory Committee of the French *Centre d'accès sécurisé aux données* (CASD), of which I am the current chair.

check" service to researchers prior to submission to journal (125, 126). The Federal Reserve Bank of Kansas City is implementing properly curated data supplements for its working papers, prior to any journal publication (127).

6. Conclusion

Reproducibility has certainly gained more visibility and traction since Dewald *et al*'s wake up call. Twenty years after Dewald *et al* saw the emergence of data archives and data availability policies at top economics journals. Thirty years after Dewald *et al*, the largest association of economists has designated a data editor for its journals, More general projects that provide training on reproducibility (TIER (87), BITSS (86)) and infrastructure for curated reproducibility (RunMyCode (128), Dataverse (45), OpenICPSR (129), Zenodo (130), CodeOcean (131), Whole Tale (132)) are gaining traction in economics, and many other initiatives that are likely to yield improved reproducibility are in their early stages.

Still, after 30 years, the results of reproducibility studies consistently show problems with about a third of reproduction attempts, and the increasing share of restricted-access data in economic research requires new tools, procedures, and methods to enable greater view onto the reproducibility of such studies. Incorporating consistent training in reproducibility into graduate curricula remains one of the challenges for the (near) future.

7. References

1. Bollen K, Cacioppo JT, Kaplan RM, Krosnick JA, Olds JL (2015) *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science* (National Science Foundation) Available at: https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf [Accessed May 20, 2018].

2. Hamermesh DS (2017) Replication in Labor Economics: Evidence from Data, and What It Suggests. *American Economic Review* 107(5):37–40.

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

4. Clemens MA (2017) The Meaning of Failed Replications: A Review and Proposal. *Journal of Economic Surveys* 31(1):326–342.

5. Journal of Applied Econometrics (2014) EXTENSION OF THE REPLICATION SECTION'S COVERAGE. Available at: https://onlinelibrary.wiley.com/page/journal/10991255/homepage/News.html#replication [Accessed July 17, 2018].

6. Pesaran H (2003) Introducing a replication section. *Journal of Applied Econometrics* 18(1):111–111.

7. McCullough BD, McGeary KA, Harrison TD (2006) Lessons from the JMCB Archive. *Journal of Money, Credit, and Banking* 38(4):1093–1107.

8. Stigler GJ, Stigler SM, Friedland C (1995) The Journals of Economics. *Journal of Political Economy* 103(2):331–359.

9. Margo RA (2011) The Economic History of the American Economic Review: A Century's Explosion of Economics Research. *American Economic Review* 101(1):9–35.

10. Frisch R (1933) Editor's Note. *Econometrica* 1(1):1–4.

11. Anderson MJ (2015) *The American census: a social history* (Yale University Press, New Haven). Second Edition.

12. Bullock CJ (1919) Prefatory Statement. *The Review of Economics and Statistics* 1(1). Available at: http://www.jstor.org/stable/1928753 [Accessed July 18, 2018].

13. Standard Charts and Tables: Original Data (1919) *The Review of Economics and Statistics* 1(1):64–103.

14. Kraus R (2013) Statistical déjà vu: The National Data Center Proposal of 1965 and its descendants. *Journal of Privacy and Confidentiality*. doi:10.29012/jpc.v5i1.624.

15. Recommendations on Availability of Federal Statistical Materials to Nongovernmental Research Workers (1959) *The American Statistician* 13(4):15–37.

16. ICPSR: The Founding and Early Years Available at: https://www.icpsr.umich.edu/icpsrweb/content/membership/history/early-years.html [Accessed July 18, 2018].

17. Miller WE (2016) The Inter-University Consortium for Political Research: *American Behavioral Scientist*. doi:10.1177/000276426300700304.

18. Sobek M, Ruggles S (1999) The IPUMS Project: An Update. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 32(3):102–110.

19. Dewald WG, Thursby JG, Anderson RG (1986) Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *The American Economic Review* 76(4):587–603.

20. Rousseau R, Egghe L, Guns R (2018) *Becoming Metric-Wise* (Elsevier) doi:10.1016/C2017-0-01828-1.

21. Schöpfel J (2011) Towards a Prague Definition of Grey Literature. *The Grey Journal (TGJ): An international journal on grey literature* 7(1). Available at: http://hdl.handle.net/10068/700015.

22. Ginsparg P (1997) Winners and Losers in the Global Research Village. *The Serials Librarian* 30(3–4):83–95.

23. Halpern JY (1998) A Computing Research Repository. *D-Lib Magazine* 4(11). doi:10.1045/november98-halpern.

24. Welch F (1973) *Education, Information, and Efficiency* (National Bureau of Economic Research) Available at: http://www.nber.org/papers/w1 [Accessed July 16, 2018].

25. RePEc: Research Papers in Economics Available at: http://repec.org/ [Accessed July 19, 2018].

26. Krichel T, Zimmermann C (2009) The Economics of Open Bibliographic Data Provision. *Economic Analysis and Policy* 39(1):143–152.

27. Bátiz-Lazo B, Krichel T (2012) A brief business history of an on-line distribution system for academic research called NEP, 1998-2010. *Journal of Management History* 18(4):445–468.

28. Krichel T (1997) WoPEc: Electronic Working Papers in Economics Services. *Ariadne* (8). Available at: http://www.ariadne.ac.uk/issue8/wopec [Accessed July 19, 2018].

29. Cruz JMB, Krichel T (2000) Cataloging Economics Preprints. *Journal of Internet Cataloging* 3(2–3):227–241.

30. BibEc main page (1997) Available at: http://web.archive.org/web/19971211044921/http://netec.mcc.ac.uk:80/BibEc.html [Accessed July 19, 2018].

31. IDEAS/RePEc Available at: https://ideas.repec.org/ [Accessed July 19, 2018].

32. Economics Software | IDEAS/RePEc Available at: https://ideas.repec.org/i/c.html [Accessed July 19, 2018].

33. Eddelbüttel D (1997) A Code Archive for Economics and Econometrics. *Computational Economics* 10(4). Available at: http://web.archive.org/web/19980515055648/http://netec.mcc.ac.uk:80/~adnetec/CodEc/ce97.pdf [Accessed July 19, 2018].

34. CodEc - Programs for Economics and Econometrics (1998) Available at: http://web.archive.org/web/19980121224535/http://netec.mcc.ac.uk:80/CodEc.html [Accessed July 19, 2018].

35. Smith A (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations* (MetaLibri Digital Library). 2007 edition, S. M. Soares.

36. Raj Chetty (2012) Time Trends in the Use of Administrative Data for Empirical Research. Available at: http://www.rajchetty.com/chettyfiles/admin_data_trends.pdf [Accessed July 19, 2018].

37. R Core Team (2000) *R: A Language and Environment for Statistical Computing (Version 1.0)* (R Foundation for Statistical Computing, Vienna, Austria) Available at: https://www.R-project.org.

38. Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A Fresh Approach to Numerical Computing. *SIAM Review* 59(1):65–98.

39. Sargent TJ, Stachurski J (2017) Lectures in Quantitative Economics. *Quantitative Economics*. Available at: https://lectures.quantecon.org/about_lectures.html [Accessed July 16, 2018].

40. King G (1995) Replication, Replication. *PS, political science & politics* 28(3):443–499.

41. Feige E (1975) The Consequences of Journal Editorial Policies and a Suggestion for Revision. *Journal of Political Economy* 83:1291–95.

42. Council NR (2003) *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences* doi:10.17226/10613.

43. Duvendack M, Palmer-Jones RW, Reed W (2015) Replications in Economics: A Progress Report. *Econ journal watch* 12(2):164–191.

44. Vlaeminck S, Herrmann L-K (2015) Data Policies and Data Archives: A New Paradigm for Academic Publishing in Economic Sciences? *Proceedings of the 19th International Conference on Electronic Publishing*, eds Schmidt B, Dobreva M, pp 145–155.

45. The Dataverse Project - Dataverse.org Available at: https://dataverse.org/home [Accessed July 20, 2018].

46. Höffler JH (2017) Replication and Economics Journal Policies. *American Economic Review* 107(5):52–55.

47. Crosas M, et al. (2018) Data policies of highly-ranked social science journals. doi:10.17605/osf.io/9h7ay.

48. Stodden V, Guo P, Ma Z (2013) Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. *PLoS ONE* 8(6):e67111.

49. Stodden V, Seiler J, Ma Z (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. *PNAS*:201708290.

50. Chang AC, Li P (2017) A Preanalysis Plan to Replicate Sixty Economics Research Papers That Worked Half of the Time. *American Economic Review* 107(5):60–64.

51. Anderson RG, Dewald WG (1994) Replication and Scientific Standards in Applied Economics A Decade After the Journal of Money, Credit and Banking Project. *Federal Reserve Bank of St Louis Review* 76(6). doi:10.20955/r.76.79-83.

52. Berry J, Coffman LC, Hanley D, Gihleb R, Wilson AJ (2017) Assessing the Rate of Replication in Economics. *American Economic Review* 107(5):27–31.

53. Duvendack M, Palmer-Jones R, Reed WR (2017) What Is Meant by "Replication" and Why Does It Encounter Resistance in Economics? *American Economic Review* 107(5):46–51.

54. Galiani S, Gertler P, Romero M (2017) *Incentives for Replication in Economics* (National Bureau of Economic Research) doi:10.3386/w23576.

55. Camerer CF, et al. (2016) Evaluating replicability of laboratory experiments in economics. *Science*:aaf0918.

56. Duflo E (2018) Report of the Editor: American Economic Review. *AEA Papers and Proceedings* 108:636–651.

57. Goldberg PK (2017) Report of the Editor: American Economic Review. *American Economic Review* 107(5):699–712.

58. Moffitt RA (2011) Report of the Editor: American Economic Review (with Appendix by Philip J. Glandon). *American Economic Review* 101(3):684–93.

59. Glandon P (2011) Report on the American Economic Review Data Availability Compliance Project. *Appendix to American Economic Review Editors Report*. Available at: http://www.aeaweb.org/aer/2011_Data_Compliance_Report.pdf.

60. Chemin M, Wasmer E (2017) Erratum. *Journal of Labor Economics* 35(4):1149–1152.

61. Chemin M, Wasmer E (2009) Using Alsace-Moselle Local Laws to Build a Difference-in-Differences Estimation Strategy of the Employment Effects of the 35-Hour Workweek Regulation in France. *Journal of Labor Economics* 27(4):487–524.

62. Godechot O (2016) *L'Alsace-Moselle peut-elle décider des 35 heures?* (SciencesPo - Observatoire Sociologique du Changement) Available at:

http://www.sciencespo.fr/osc/sites/sciencespo.fr.osc/files/ND_2016-04.pdf [Accessed July 20, 2018].

63. Godechot O (2016) *Can We Use Alsace-Moselle for Estimating the Employment Effects of the 35-Hour Workweek Regulation in France?* (SciencesPo - Observatoire Sociologique du Changement) Available at: http://olivier.godechot.free.fr/hopfichiers/fichierspub/Comment_on_Chemin_Wasmer_2009_Jole.pdf [Accessed July 20, 2018].

64. Weinberg DH, Abowd JM, Steel PM, Zayatz L, Rowland SK (2007) *Access Methods for United States Microdata* (Center for Economic Studies, U.S. Census Bureau) doi:10.2139/ssrn.1015374.

65. Currie R, Fortin S (2015) *Social statistics matter: History of the Canadian Research Data Center Network* (Canadian Research Data Centre Network) Available at: http://rdc-cdr.ca/sites/default/files/social-statistics-matter-crdcn-history.pdf [Accessed December 15, 2017].

66. Bender S, Heining J (2011) The Research-Data-Centre in Research-Data-Centre Approach: A First Step Towards Decentralised International Data Sharing. *IASSIST quarterly / International Association for Social Science Information Service and Technology* 35(3). Available at: http://www.iassistdata.org/iq/issue/35/3.

67. Une bulle pour protéger les fichiers (2014) *Le Monde*:4–5.

68. Bozio A, Geoffard P-Y (2017) *L'accès des chercheurs aux données administratives* (Conseil national de l'information statistique).

69. Gadouche K, Picard N (2017) *L'accès aux données très détaillées pour la recherche scientifique* (Université Cergy-Pontoise).

70. About IRIS *IRIS*. Available at: http://iris.isr.umich.edu/about/ [Accessed July 20, 2018].

71. Weinberg BA, et al. (2014) Science Funding and Short-Term Economic Activity. *Science* 344(6179):41–43.

72. About the Health Care Cost Institute | HCCI *Health Care Cost Institute*. Available at: http://www.healthcostinstitute.org/about-hcci/ [Accessed July 20, 2018].

73. Newman D, Herrera C-N, Parente ST (2014) Overcoming Barriers to a Research-Ready National Commercial Claims Database. *American Journal of Managed Care* 11(17):eSP25–eSP30.

74. PCRI - The Private Capital Research Institute Available at: http://www.privatecapitalresearchinstitute.org/about.php [Accessed July 20, 2018].

75. Jeng L, Lerner J (2016) Making Private Data Accessible in an Opaque Industry: The Experience of the Private Capital Research Institute. *American Economic Review* 106(5):157–160.

76. King G, Persily N (2018) *A New Model for Industry-Academic Partnerships* (Harvard University) Available at: http://j.mp/2q1IQpH.

77. About the PSID Repository Available at: https://www.openicpsr.org/openicpsr/psid [Accessed July 21, 2018].

78. Welch F (1974) MINIMUM WAGE LEGISLATION IN THE UNITED STATES. *Economic Inquiry* 12(3):285–318.

79. Welch F (1977) MINIMUM WAGE LEGISLATION IN THE UNITED STATES: REPLY. *Economic Inquiry* 15(1):139–142.

80. Siskind FB (1977) MINIMUM WAGE LEGISLATION IN THE UNITED STATES: COMMENT. *Economic Inquiry* 15(1):135–138.

81. Hirsch BT, Schumacher EJ (2004) Match Bias in Wage Gap Estimates Due to Earnings Imputation. *Journal of Labor Economics* 22(3):689–722.

82. Alexander JT, Davern M, Stevenson B (2010) Inaccurate Age and Sex Data in the Census Pums Files: Evidence and Implications. *Public Opinion Quarterly* 74(3):551–569.

83. Höffler JH (2017) ReplicationWiki: Improving Transparency in Social Sciences Research. *D-Lib Magazine* 23(3/4). doi:10.1045/march2017-hoeffler.

84. Herndon T, Ash M, Pollin R (2014) Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics* 38(2):257–279.

85. Reinhart CM, Rogoff KS (2010) Growth in a Time of Debt. *American Economic Review* 100(2):573–578.

86. Berkeley Initiative for Transparency in the Social Sciences (2015) *Berkeley Initiative for Transparency in the Social Sciences*. Available at: https://www.bitss.org/about/ [Accessed July 22, 2018].

87. Ball R, Medeiros N (2012) Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis. *The Journal of Economic Education* 43(2):182–189.

88. Data Citation Synthesis Group, Martone M (2014) *Joint Declaration of Data Citation Principles* (Force11) doi:10.25490/a97f-egyk.

89. ICPSR - Data Citations (2018) Available at: https://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp [Accessed July 22, 2018].

90. Starr J, et al. (2015) Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science* 1:e1.

91. Author-Date: Sample Citations (2018) *The Chicago Manual of Style Online*. Available at: http://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-2.html [Accessed July 22, 2018].

92. Sample References - Styles of the AEA (2018) Available at: https://www.aeaweb.org/journals/policies/sample-references [Accessed July 22, 2018].

93. AER Style Guide for Accepted Articles (2018) Available at: https://www.aeaweb.org/journals/aer/submissions/accepted-articles/styleguide [Accessed July 22, 2018].

94. Submissions - Royal Economic Society (2018) Available at: http://www.res.org.uk/view/submissionsEconometrics.html [Accessed July 22, 2018].

95. PSID - Data Center - Previous carts (2018) Available at: https://simba.isr.umich.edu/VS/c.aspx [Accessed July 22, 2018].

96. OnTheMap (2018) Available at: https://onthemap.ces.census.gov/ [Accessed July 22, 2018].

97. Data Versioning WG (2016) *RDA*. Available at: https://www.rd-alliance.org/groups/data-versioning-wg [Accessed July 22, 2018].

98. Rauber A, Asmi A (2016) Identification of Reproducible Subsets for Data Citation, Sharing and Re-Use. *Bulletin of IEEE Technical Committee on Digital Libraries* 12(1):10.

99. The Abdul Latif Jameel Poverty Action Lab - Hypothesis Registry (2009) Available at: https://www.povertyactionlab.org/Hypothesis-Registry [Accessed July 22, 2018].

100. Katz L, Duflo E, Goldberg P, Thomas D (2013) Email: AEA Registry for Controlled Trials. Available at: https://web.archive.org/web/20131128040053/http://www.aeaweb.org:80/announcements/20131118_rct_email.php [Accessed July 17, 2018].

101. RIDIE - Registry for International Development Impact Evaluations Available at: http://ridie.org [Accessed July 22, 2018].

102. Dahl Rasmussen O, Malchow-Møller N, Barnebeck Andersen T (2011) Walking the talk: the need for a trial registry for development interventions. *Journal of Development Effectiveness* 3(4):502–519.

103. EGAP Rules and Procedures (2009) Available at: http://egap.org/sites/default/files/pdfs/20110608_EGAP_structure.pdf [Accessed July 22, 2018].

104. AsPredicted: About (2018) Available at: https://aspredicted.org/messages/about.php [Accessed July 22, 2018].

105. OSF Guides - Registrations (2018) Available at: http://help.osf.io/m/registrations [Accessed July 22, 2018].

106. Christensen G, Miguel E Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*. doi:10.1257/jel.20171350.

107. Lagoze C, Vilhuber L (2017) Making Confidential Data Part of Reproducible Research. *Chance*. Available at: http://chance.amstat.org/2017/09/reproducible-research/.

108. Reiter JP (2003) Model diagnostics for remote-access regression servers. *Statistics and computing* 13:371–380.

109. Reiter JP, Oganian A, Karr AF (2009) Verification servers: Enabling analysts to assess the quality of inferences from public use data. *Computational statistics & data analysis* 53(4):1475–1482.

110. Kinney SK, et al. (2011) Towards Unrestricted Public Use Business Microdata: The Synthetic Longitudinal Business Database. *International statistical review* 79(3):362–384.

111. Vilhuber L, Abowd JM, Reiter and JP (2016) Synthetic Establishment Microdata Around the World. *Statistical Journal of the International Association of Official Statistics* 32(1):65–68.

112. Vilhuber L, Abowd J (2016) Usage and outcomes of the Synthetic Data Server. Available at: http://hdl.handle.net/1813/43883 [Accessed July 22, 2018].

113. Nowok B, Raab G, Dibben C (2016) synthpop: Bespoke Creation of Synthetic Data in R. *Journal of Statistical Software, Articles* 74(11):1–26.

114. Drechsler J (2012) New data dissemination approaches in old Europe – synthetic datasets for a German establishment survey. *Journal of Applied Statistics* 39(2):243–265.

115. Nosek BA, Lakens D (2014) Registered Reports: A Method to Increase the Credibility of Published Results. *Social Psychology* 45(3):137–141.

116. OSF - Registered Reports (2018) Available at: https://cos.io/rr/ [Accessed July 22, 2018].

117. Chris Chambers (2014) Registered Reports: A step change in scientific publishing. *Reviewers' Update*. Available at: https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/registered-reports-a-step-change-in-scientific-publishing [Accessed July 22, 2018].

118. Burman LE, Reed WR, Alm J (2010) A Call for Replication Studies. *Public Finance Review* 38(6):787–793.

119. Antenucci D, Cafarella M, Levenstein M, Ré C, Shapiro and MD (2014) *Using Social Media to Measure Labor Market Flows*.

120. Prediction of Initial Claims for Unemployment Insurance (2017) Available at: http://econprediction.eecs.umich.edu/ [Accessed July 22, 2018].

121. Bowers J, Higgins N, Karlan D, Tulman S, Zinman J (2017) Challenges to Replication and Iteration in Field Experiments: Evidence from Two Direct Mail Shots. *American Economic Review* 107(5):462–465.

122. Duflo E, Hoynes H (2018) Report of the Search Committee to Appoint a Data Editor for the AEA. *AEA Papers and Proceedings* 108:745.

123. Jacoby WG, Lafferty-Hess S, Christian T-M (2017) Should Journals Be Responsible for Reproducibility? *Inside Higher Ed*. Available at: https://www.insidehighered.com/blogs/rethinking-research/should-journals-be-responsible-reproducibility [Accessed July 22, 2018].

124. The Abdul Latif Jameel Poverty Action Lab - Transparency & Reproducibility (2015) Available at: https://www.povertyactionlab.org/research-resources/transparency-and-reproducibility [Accessed May 22, 2018].

125. Results Reproduction (R-squared) – CISER (2018) Available at: https://ciser.cornell.edu/research/results-reproduction-r-squared-service/ [Accessed July 22, 2018].

126. Arguillas F, Christian T-M, Peer L (2018) Education for (a) CURE: Developing a prescription for training in data curation for reproducibility. Available at: https://www.openconf.org/IASSIST2018/modules/request.php?module=oc_program&action=summary.php&id=160 [Accessed July 22, 2018].

127. Butler C, Kulp C (2018) The role of data supplements in reproducibility: Curation challenges. Available at: https://www.openconf.org/IASSIST2018/modules/request.php?module=oc_program&action=summary.php&id=41 [Accessed July 22, 2018].

128. Stodden V, Hurlin C, Perignon C (2012) RunMyCode.Org: A Novel Dissemination and Collaboration Platform for Executing Published Computational Results. *SSRN Electronic Journal*. doi:10.2139/ssrn.2147710.

129. openICPSR: Share your behavioral health and social science research data Available at: https://www.openicpsr.org/openicpsr/ [Accessed July 22, 2018].

130. Zenodo - Research. Shared. Available at: http://about.zenodo.org/ [Accessed July 22, 2018].

131. About | Code Ocean Available at: https://codeocean.com/about [Accessed July 22, 2018].

132. Brinckman A, et al. (2018) Computing environments for reproducibility: Capturing the “Whole Tale.” *Future Generation Computer Systems*. doi:10.1016/j.future.2017.12.029.

8. Figure Legends

Figure 1: Use of Administrative Data in Publications in Leading Journals, 1980-2010

Note: “Administrative” datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries. Figure reproduced from (35).