JPE Data and Code Policy

Table of contents

It is the policy of the Journal of Political Economy (JPE) to publish papers only if the data and computer code used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of conditionally accepted papers, in particular those that contain empirical work, simulations, experimental work, or

numerical computations, must provide, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. They must also provide sufficient information to replicate the process of obtaining the raw data from the original sources and cite all the sources of data appropriately.

The JPE will perform reproducibility checks of empirical, experimental and simulation results of all conditionally accepted papers and their approved online appendices. Eventual publication of the paper is conditional on a positive outcome of those checks. Requests for an exemption from providing the materials described in this policy, or for restricting their usage, should be stated clearly when the paper is first submitted for review. The handling editor will decide whether the paper should be reviewed in this case. Exceptions will not be considered later in the review and publication process.

By submitting to the JPE, authors indicate their acceptance of this Data and Code Policy.

DCAS v1.0 The JPE endorses DCAS, the Data and Code Availability Standard [v1.0], and its data and code availability policy is compatible with DCAS.

The specific terms of the JPE Data and Code Policy are as follows.

1 Data

1.1 Data Availability Statement

A Data Availability Statement (DAS) must be provided with sufficient detail for independent researchers to replicate the necessary steps to access the original data, including information on any limitations and the expected monetary and time costs associated with data access. When applicable, the DAS should also specify the version of the dataset and the original date of access by the authors. Similarly, the DAS should clearly indicate which datasets are included and excluded from the replication package. The DAS should be included as a section of the README file (see Section 3.4 below).

1.2 Raw Data

The raw data utilized in the research, including primary data collected by the author and secondary data, must be included in the replication package. If the exact extract of the raw data used in the analysis is published in a trusted repository that satisfies the FAIR data principles (guidance here), including a permanent identifier (e.g. its DOI) linking to these raw data is considered sufficient to fulfill the obligation of including the raw data in the replication package. In cases where the terms of use of the secondary data or the ethics requirements of the primary data used in the analysis prohibit the authors from publishing them, the authors must request an exemption and provide a justification for not complying with the policy. When

exemptions are granted, authors are required to comply with at least one of the following two procedures:

1. Whenever possible, provide the JPE with temporary access to the data affected by the exemption for the purpose of implementing reproducibility checks. It is the authors' responsibility to obtain permission from the data provider to confidentially share the data with the JPE.
2. Include in the replication package a synthetic or simulated dataset that enables users to execute the code and verify that it generates all outputs presented in the paper and appendices, even if the results differ from those in the paper. While including synthetic/simulated data is not required when temporary access for JPE is provided, it is still recommended, as it allows future users of the package to run the codes, increasing the publication's impact.

In either of these two procedures, the authors are expected to clarify the nature of the exemption in the DAS (see Section 1.1).

1.3 Analysis Data

Analysis Data is provided as part of the replication package unless they can be fully reproduced from accessible data within a reasonable time frame. Exceptions are to be explained in the DAS.

1.4 Format

The data files must be either in plain `ASCII` format, such as comma-separated value (`.csv`), or any other non-proprietary format so that they can be read by any researcher on any machine. Additionally, the authors may choose to submit data in a format that is read by specific programs, such as Matlab (`.mat`) files, Stata (`.dta`), or Excel (`.xlsx`) files, but a copy of these files in a non-proprietary format is required in every case.

1.5 Metadata

A description of the variables included in the data and their allowed values must be made publicly accessible. Such a description could take the form of labels in the dataset, comments in the code, easy-to-identify variable names, codebooks, and indications in the `README` file.

1.6 Citations

All data used in the paper and the approved online appendices must be appropriately cited in both the paper/appendices and in a dedicated references section of the `README` file. As a general guideline, citations of data employed in the paper should be included in the paper's references section, while citations exclusively pertaining to data used in the approved online appendices may be relegated to the appendix. However, in exceptional circumstances, such as when there is a large number of data sources to cite or when recommended by the handling co-editor, citations of data used in the paper may be included in a references section of the approved appendix only.

2 Code

2.1 Data Transformations

All programs used to generate final and analysis data sets from raw data must be included, even if the raw datasets cannot be provided due to approved exemptions to comply with Section 1.2 above. Whenever transformations or simulations include randomly generated numbers, the code must ensure reproducibility with appropriate means (e.g. setting a seed for the random number generator).

2.2 Analysis Code

Programs that produce any kind of computational results (e.g. estimation, simulation, model solution, visualization, etc.) must be included. These programs should produce all such computational exhibits in the paper and approved online appendices (i.e. tables, figures, in-text numbers) with minimal human intervention.

In cases where execution times of the programs are excessively long, authors are encouraged to provide simplified versions of their programs that allow partial reproduction of results in a reasonable timeframe. Similarly, authors of such computationally intensive papers are encouraged to include intermediate results in the replication package, such that reproducibility checks can be performed, taking those results as given. Note that this does not invalidate the requirement to provide code that reproduces all output without relying on intermediate results. In such cases, authors are required to collaborate with the Data Editor to develop a feasible strategy for reproducibility checks. The extent of the eventual partial reproducibility checks must be clearly documented in the package `README`.

2.3 Format of Code

Codes must be provided in source format which can be directly interpreted or compiled by appropriate software. Single-driver scripts that run all of the code from raw data to final results are strongly encouraged, and must be provided at the Data Editor's specific request (e.g. to limit the number of human intervention steps). The code must save all exhibits (e.g. tables and figures) appearing in the paper and appendices in a specified directory within the replication package. When the codes are written in compiled languages, precise instructions for all steps and compiling options must be included in the documentation. A `make` file (or similar build tool) that reproduces compilation steps is strongly encouraged. In general, all necessary steps to recreate a stable computational environment should be taken (such as, for example, precise specification of software and library versions, virtual environments, and containerization). Software that does not allow generating output using scripts (e.g., ArcGIS) is discouraged. When this type of software is used, very precise step-by-step instructions allowing users to exactly reproduce the generated outputs independently of the authors must be included in the `README` file.

3 Supporting Materials

3.1 Survey Instruments and Experimental Instructions

3.1.1 Material Expected in Replication Package

In case the raw data are collected or generated via surveys or experiments, authors are required to include survey instruments or experimental instructions and subject selection criteria in the replication package. Specifically, it is required that the entirety of the following information be included in the package `README`, regardless of whether some of it appears in the paper or the appendix already:

1. The subject pool and recruiting procedures.
2. The experimental technology – when and where the experiments were conducted; by computer or manually; online, and so forth.
3. Any procedures to test for comprehension before running the experiment, including the use of practice trials and quizzes.
4. Matching procedures, especially for game theory experiments.
5. Subject payments, including whether artificial currency was used, the exchange rate, show-up fees, average earnings, lotteries, and/or grades.
6. The number of subjects used in each session and, where relevant, their experience.
7. Timing, such as how long a typical session lasted, and how much of that time was instructional.
8. Any use of deception and/or any instructional inaccuracies.

9. Detailed statement of protocols.
10. Samples of permission forms and record sheets.
11. Copies of instructions and slides/transparencies used to present instructions.
12. Source code for computer programs used to conduct the experiment and to analyze the data. This does not include compilers (such as zTree) that are publicly available.
13. Screen shots showing how the programs are used.

The documentation provided in the replication package must be self-contained, regardless of the content included in the paper and approved online appendices, as the replication package is a different citable object than the paper. As a general rule, the replication package must be at least as exhaustive as the paper and approved appendices, since there are no space constraints in the documentation provided in the replication package.

3.1.2 Initial Submission of Experimental Papers

Similarly to requests about exemptions from the Data and Code Policy, authors should include as much information about their experimental procedure as possible in their initial submission. All the materials listed above are desirable and typically expected, but further details about what is needed in each case can be obtained from the co-editor handling the paper. If, during the review process, the editor or referees feel additional information is needed, requests for that material will be made and may naturally cause a delay in processing. Hence, we encourage as complete a submission as feasible.

3.2 Ethics

Authors who collect primary data (e.g. via experiment or survey) are required to include the IRB approval documentation (or similar) from their institution.

3.3 Pre-registration

If applicable, pre-registration of the research project must be documented in the `README`.

3.4 Documentation in `README` file

A `README` document in either `pdf` or `markdown` format has to be included in the replication package. The file `README.pdf` or `README.md` must be located at the root of the replication package, it must contain all information required to reproduce the results (there should not be multiple locations for documentation). As a minimal requirement, the `README` should contain at least the following items:

1. A DAS as explained in Section 1.1

2. A description of the content of the replication package.
3. Precise instructions on all the steps required to run the code to reproduce all exhibits included in the paper and appendices.
4. Precise instructions on where in the replication package the produced outputs will be saved, and how each of them maps to the exhibits included in the paper and appendices.
5. Precise specifications of software and hardware used by the authors when preparing the package, including expected running time and specific requirements needed to successfully reproduce the results (e.g., software versions, libraries to be installed, etc.). When the requirements and execution time are heterogeneous across significant portions of the package, specific requirements and running times for each of the different parts must be indicated.
6. Data citations, according to Section 1.6.

It is strongly recommended to use the `README` template (available here) of the Social Sciences Data Editors.

4 Sharing

4.1 Location

By default, conditionally accepted papers should deposit their replication package in the JPE dataverse repository, where it will be reviewed by the Data Editor. After the successful conclusion of reproducibility checks, the package will be published in this dataverse space.

Other repositories and archives may be acceptable for all or part of the replication package, as long as they are considered to be trusted archives or repositories, as alluded to in Section 1.2 above. The Data Editor will assess the suitability of any such repositories and archives.

In cases where data cannot be published in an openly accessible trusted data repository, authors who have requested an exemption to publish them at the time of first submission must commit to preserving the data and code for a period of no less than five years following the publication of the paper. They should also provide reasonable assistance to requests for clarification and replication.

4.2 License

The replication package must be deposited with a license that specifies the terms of use of the code and data in the replication package. The license must, at the very least, allow for unrestricted access to all files included in the deposit and permit the usage of the package for replication purposes by researchers unconnected to the original parties.

4.3 Omissions

The README must clearly indicate any omission of the required parts of the package as a result of a granted exemption. The README must also indicate the reasons for such omission, such as legal requirements, limitations, or other approved agreements. In cases where the extent or complexity of the reproducibility checks impedes the exact reproduction of all the results in the paper and approved appendices, such as when synthetic datasets are provided (see Section 1.2) or partial checks have been implemented (see Section 2.2), the README must clarify which results have not been checked for reproducibility and why.