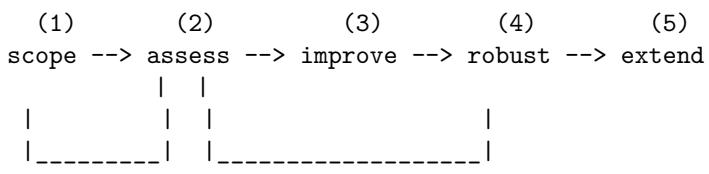
Guidelines for Computational Reproducibility in Economics

ACRE Team

2020-06-07

Contents



(1) Scoping

(2) Assessment

(3) Improvement

(4) Robustness

(5) Extensions

Outcome-level

Paper-level

Select paper

Describe inputs

+ Raw data

+ Version control

Analytical choices

New method

Check ACRE

Reproduction diagrams

+ Analysis data

+ Documentation

Type of choice

New data

Check Rep. pkg exists

Reproduction score

Introduction

In 2019, the American Economic Association updated its Data and Code Availability Policy, which now requires that the AEA Data Editor verify the reproducibility of all papers before they are accepted by an AEA journal. In addition to the requirements laid out in the policy, several specific recommendations were produced to facilitate compliance. This change in policy is expected to improve the computational reproducibility of all published research going forward, after several studies showed that rates of *computational reproducibility* in economics at large range from somewhat low to alarmingly low (**???**).

Replication, or the process by which a study's hypotheses and findings are re-examined using different data or different methods (or both) (**?**) is an essential part of the scientific process that allows science to be "self-correcting." *Computational reproducibility*, or the ability to reproduce the results, tables, and other figures using the available data, code, and materials, is a precondition for replication. Computational reproducibility is assessed through the process of *reproduction*. At the center of this process is the *reproducer* (you!), a party not involved in the production of the original paper. Reproductions sometimes involve the *original author* (whom we refer to as "the author") in cases where additional guidance and materials are needed to execute the process.

This exercise is designed for reproductions performed in economics graduate courses or undergraduate theses, with the goal of providing a common approach, terminology, and standards for conducting reproductions. The goal of reproduction, in general, is to assess and improve the computational reproducibility of published research in a way that facilitates further robustness checks, extensions, collaborations, and replication.

This exercise is part of the Accelerating Computational Reproducibility in Economics (ACRE) project led by the Berkeley Initiative for Transparency in the Social Sciences (BITSS) and Prof. Lars Vilhuber, Data Editor for the journals of the American Economic Association (AEA). ACRE looks to assess, enable, and improve the computational reproducibility of published economics research.

Beyond binary judgments

Assessments of reproducibility can easily gravitate towards binary assessments that declare an entire paper "reproducible" or "non-reproducible." These guidelines suggest a more nuanced approach by highlighting two reasons that make binary judgments less relevant.

First, a paper may contain several scientific claims (or major hypothesis) that may vary in computational reproducibility. Each claim is tested using different methodologies, where results are presented in one or more display items (outputs like table and figures). Each display item will itself contain several specifications. Figure **??** illustrates this idea.

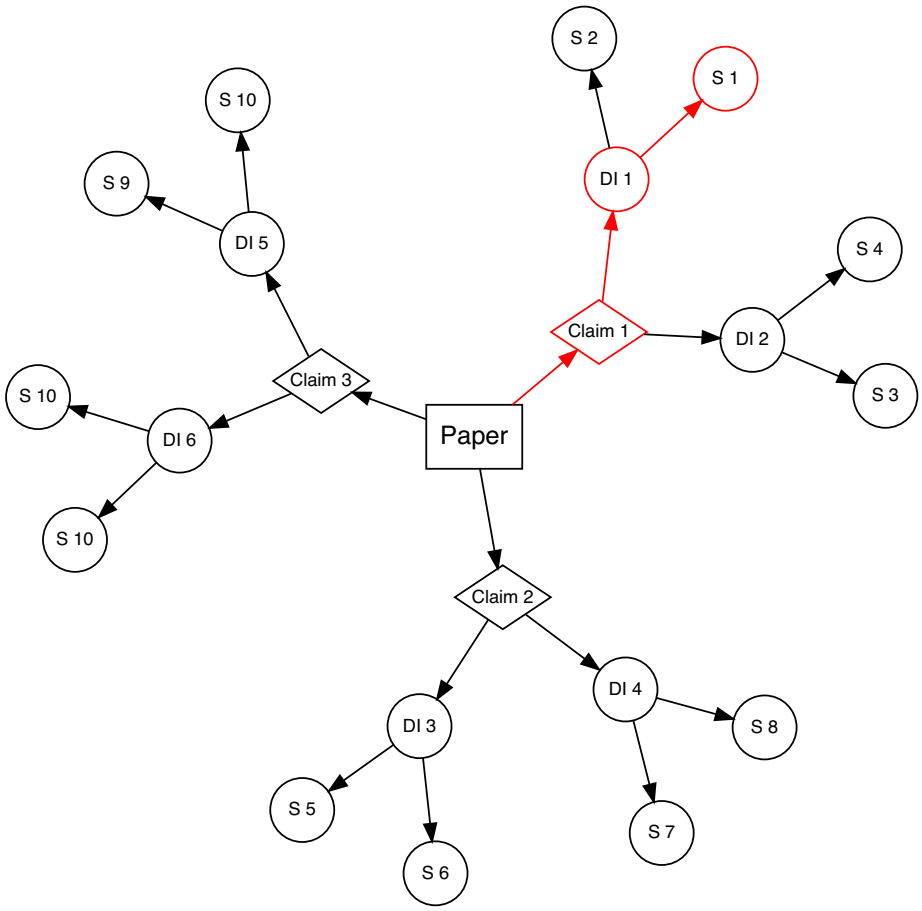


Figure 1: One paper has multiple components to reproduce. DI: Display Item, S: Specification

Second, for a given specification there are several levels of reproducibility,

ranging from the absence of any materials to complete reproducibility starting from raw data. And even for a specific claim-specification, distinguishing the appropriate level can be far more constructive than simply labeling it as (ir)reproducible.

Note that the highest level of reproducibility, which requires complete reproducibility starting from raw data, is very demanding to achieve and should not be expected of all published research – especially before 2019. Instead, this level can serve as an aspiration as the field of economics at large seeks to improve the reproducibility of research and facilitate the transmission of knowledge throughout the scientific community.

Stages of the exercise

This reproduction exercise is divided into four stages, corresponding to the first four chapters of these guidelines, with a fifth optional stage:

1. **Scoping**, where you (the reproducer) will define the scope of the exercise by declaring a paper and the specific output(s) on which you will focus in the remainder of the exercise;

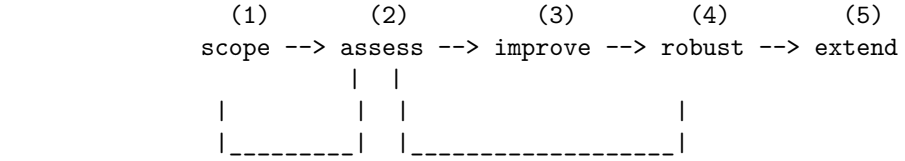
2. **Assessment**, where you will review and describe in detail the available reproduction package, and assess the current level of computational reproducibility of the selected outputs;

3. **Improvement**, where you will modify the content and/or the organization of the reproduction package to improve its reproducibility;

4. **Robustness checks**, where you will assess the quality of selected analytical choices; and

5. **Extension** (if applicable), where you may extend the current paper by including new methodologies or data. This step brings the reproduction exercise a step closer to *replication*.

Figure 2: Steps for reproduction



Suggested level of effort:					
- Graduate research:	5%	10%	5%	10%	70%
- Graduate course:	10%	25%	20%	40%	5%
- Undergrad. thesis:	10%	30%	40%	20%	0%

Figure 2 depicts suggested levels of effort for each stage of the exercise depending on the context in which you are performing a reproduction. This process need not be chronologically linear. For example, you may realize that the scope of a reproduction is too ambitious and switch to a less intensive one. Later in the exercise, you can also begin testing different specifications for robustness while also assessing a paper's level of reproducibility.

Recording the results of the exercise

You will be asked to record the results of their reproduction progress through each stage.

In *Stage 1: Scoping*, complete **Survey 1**, where you will declare your paper of choice and the specific display item(s) and specifications on which you will focus for the remainder of the exercise. This step may also involve writing a brief 1-2 page summary of the paper (confirm this with your instructor).

In *Stage 2: Assessment*, you will inspect the paper's reproduction package (raw data, analysis data, and code), connect the display item to be reproduced with its inputs, and assign a reproducibility score to each output.

In *Stage 3: Improvement*, you will try to improve the reproducibility of the selected outputs by adding missing files, documentation, and report any potential changes in the level of reproducibility. Use **Survey 2** to record your work at Stages 2 and 3 (you will receive access instructions for Survey 2 when you submit Survey 1).

In *Stage 4: Robustness Checks*, you will assess different analytical choices and test possible variations. Use **Survey 3** to record your work at this stage.

Chapter 1

Scoping

In this stage, you will choose or be assigned a candidate paper to reproduce. You will verify whether or not a reproduction package for your candidate paper exists. We define a **reproduction package** (sometimes referred to as a "replication package") as the collection of all materials that make it possible for a reproducer to reproduce the paper. This package may contain data, code, and/or documentation.

If your candidate paper has a reproduction package, only then it will become your declared paper. If your candidate paper does not have a reproduction package you will be asked to leave a brief record in the ACRE database (under current development), and choose another candidate paper. If you still want to explore the reproducibility of a paper with no reproduciton package, these guidelines will provide instructions on how contact the authors with a specific request for materials to create a public reproduction package, or if this route proves unsuccesful, on how to build your reproduction package from scratch.

In this stage, *you are not expected to review the reproduction materials in detail*, as you will dedicate most of your time to this in later stages of the exercise. If materials are available, you will read the paper and declare the scope of the reproduction exercise. You can expect to spend between 1-3 days in the Scoping stage.

Use Survey 1 to record your work for this stage.

1.1 From candidate to declared paper

1.1.1 Select a candidate paper or be assigned one (several weeks before starting the excercise).

The first step is to identify a candidate paper. You might be able to choose among a pool of paper assinged by your instructor, or might be assigned a specific paper. In any case, it is important ot highlight that this might not be your final paper. It can be the case that you will not find any reproduction package and switch to another paper. The distinction between candidate and declared paper is meant to track precicely this circunstances.

Verifiy the existence of a reproduction package by going through these steps (stop whenever you find the reproduction package):

1. Verify the existence of a previous entry on ACRE (under construction) on your candidate paper.

2. Check the paper's webpage in the official journal or publishers website. Look for links similar to "Data and Materials", "Supplemnetal Materials", "Reproduction/Replication Package/Materials".

3. Look for links in the paper (review footnotes and appendices).

4. Review the personal websites of all the authors in the paper.

Once you have cover steps 1-4 and have not find any reproduction materials, then we suggest to select another candiate paper or to contact the authors.

We strongly encourage you to contact authors several weeks before officially starting the main parts of Scoping stage of this exercise. Instructors should also plan to use the ACRE exercise accordingly (eg. if the reproducction is expected to take place in the middle of the semester, students should plan on contacting the authors of a missing reproduction package on the first few weeks of the semester.

1.1.2 Verify existence of reproduction package for candidate paper

At this point you are *only validating the existence* of (at least) one reproduction package and not assessing the quality of its content.

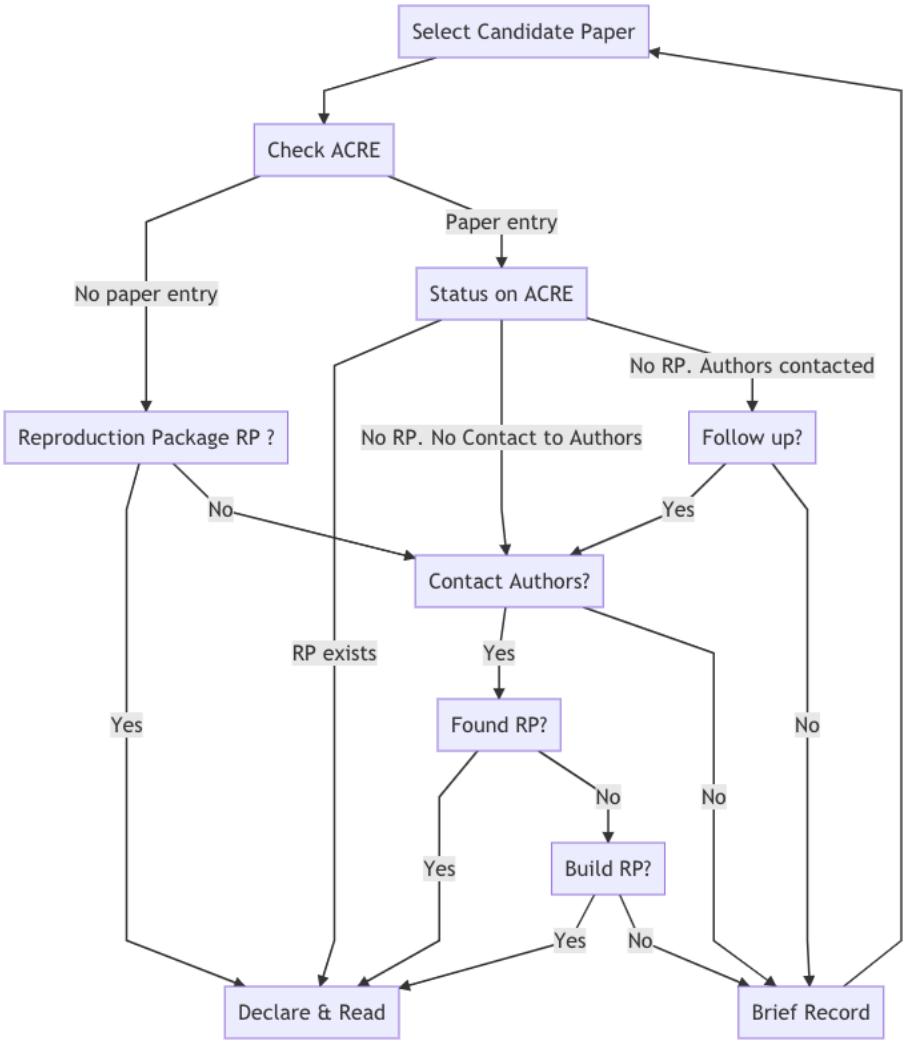
Check the ACRE database (under development) for previous assessments of your candidate paper. If there are previous entries, you will see a brief report card with the following information:

Box 1: Summary Report Card for ACRE Paper Entry
Title: Sample Title
Authors: Jane Doe & John Doe
Original Reproduction Package Available: URL/No
[If "No"] **Contacted Authors?:** Yes/No
[If "Yes(contacted)"] **Type of Response:** Categories (6).
Additional Reproduction Packages: Number (eg., 2)
Authors Available for Further Questions for ACRE Reproductions: Yes/No/Unknown
Open for reproductions: Yes/No

1.1.3 Decision Tree to Move From Candidate to Declared Paper

To move from candidate to declared paper, you will basically need to asssess if a reproduction package exists. Following the steps describe above, and leaving a record for each reproduction package that did not had a reproduction package will lead to a a decision tree (collection of sequencial decisions) like the one outlined next.

View Decision Tree To Select Paper (Emma: add title)



When you follow step 1, and check ACRE for a previuous entry on the paper you will either find and entry, or that no entry exists yet. Then:

- If there is no paper entry, create one. For this you will need to verify if a reproduction package exist in any of the steps outlined above (2-4). Here there will be two possible scenarios:

  - A reproduction package exists. In this case, the candidate paper becomes the declare paper and you can move onto the next section (read the paper).