Readme File: Replication Package for “Optimal Procurement With Quality Concerns”

Giuseppe Lopomo Nicola Persico Alessandro T. Villa

November 2022

1 Overview

The codes in this replication package were used to generate the numerical results in the paper and in the Online Appendix using Matlab. Figures are generated by a specific Matlab file with the exceptions of Figure 1, which is descriptive. The replicator should expect the code to run for approximately one hour.

2 Data Availability Statement

Regarding data availability, we are not permitted to post the original data. The editor asked us to create and post a "replication dataset" that could be run through the replication code and yields similar results as the original data. Accordingly, we created a new "replication" dataset by adding positive-mean random noise to the original observations described below.[1] The "replication" dataset is available in the replication package uploaded on AEA Data and Code Repository (Project number: openicpsr-182801) as well as on one of the author's website: https://www.alessandrotenzinvilla.com/research.html. Once the "replication" dataset is run through the kernel density estimation, which is also provided in the replication package, the resulting distributions are visually similar, but not identical, to Figure 5. Inputing these distributions into the computational procedures provided in the replication package yields results and figures similar to the ones in Section 6 and Appendix D.

2.0.1 Statement about Rights

We, the authors, certify that we have legitimate access to and permission to use the original data provided by Francesco De Carolis. Francesco De Carolis gave us permission to use the original data which we analyze in the paper, but not to distribute it. We do however have his permission to distribute a "replication dataset" that contains noise.

[1]For each of the variables (cost, delay ratio, and overrun ratio) the "fake" observations have mean that is larger than the original by 20-25 percent, and the differences between original and fake observations have standard deviations equal to 12-15 percent of the original standard deviations. This means that our procedure added noise in the amount of about 13 percent of the standard deviations calculated on the original dataset.

2.0.2 License for Data

We, the authors, certify that we have legitimate access to and permission to use the original data provided by Francesco De Carolis. Francesco De Carolis gave us permission to use the original data which we analyze in the paper, but not to distribute it. We do however have his permission to distribute a “replication dataset” that contains noise.

2.0.3 Summary of Availability

Only the “Replication” dataset can be made publicly available. The “Replication” dataset is available in the replication package uploaded on AEA Data and Code Repository (Project number: openicpsr-182801) as well as on one of the author’s website:
https://www.alessandrotenzinvilla.com/research.html.

2.0.4 Details on each Data Source

The folder “Section 6 Illustrative Application” contains two data sources:
1. Costs_Sample_Noise.csv and 2. TurinRenegotiationsNoise.csv.
The “Noise” part of the filenames indicates that these files contain “fake data” as described above.

- Both files are in `.csv` open format
- File Costs_Sample_Noise.csv contains two columns:
  1. progressive identification number
  2. bidder cost
- File TurinRenegotiationsNoise.csv contains three columns:
  1. progressive identification number
  2. delay ratio
  3. overrun ratio

The Matlab script “PreliminariesAndFigure5Noise.m” is the only one that takes in inputs these data to create the Kernel distributions in Figure 5.

2.0.5 Example for public use data collected by the authors

We do not have data collected by the authors.

2.0.6 Example for public use data sourced from elsewhere and provided

We do not have other data other the ones provided to us by Francesco De Carolis.

2.0.7 Example for public use data with required registration and provided extract

We do not have other data other the ones provided to us by Francesco De Carolis.

2.0.8 Example for free use data with required registration, extract not provided

We do not have other data other the ones provided to us by Francesco De Carolis.

2.0.9 Example for confidential data

The data for this project are confidential, and provided by Francesco De Carolis. Prof. De Carolis is a Professor in Bocconi. His webpage is: https://faculty.unibocconi.eu/francescodecarolis/.

2.0.10 Example for confidential Census Bureau data

We do not have any Census Bureau data.

2.0.11 Example for preliminary code during the editorial process

We do not have any preliminary code to run. The data are processed directly by this Matlab script "PreliminariesAndFigure5Noise.m".

2.1 Dataset list

The folder "Section 6 Illustrative Application" contains two data sources:
1. Costs_Sample_Noise.csv and 2. TurinRenegotiationsNoise.csv.
The "Noise" part of the filenames indicates that these files contain "fake data" as described above.

- Both files are in `.csv` open format
- File Costs_Sample_Noise.csv contains two columns:
  1. progressive identification number

2. bidder cost

- File TurinRenegotiationsNoise.csv contains three columns:

  1. progressive identification number
  2. delay ratio
  3. overrun ratio

The Matlab script "PreliminariesAndFigure5Noise.m" is the only one that takes in inputs these data to create the Kernel distributions in Figure 5.

3 Computational Requirements

3.1 Software Requirements

The code was run with Matlab Release 2021a (including symbolic matlab toolbox), AMPL, and IBM Ilog CPLEX 12.9 64 bit.

3.2 Controlled Randomness

There is no random seed to set.

3.3 Memory and Runtime Requirements

The approximate time to reproduce the analyses on a standard desktop machine is approximately one hours. The code was last run on a 6-core Intel-based desktop computer with Windows 11 with 16GB RAM.

4 Description of Programs

The replication package is organized in six main directories:

1. "Section 2 Illustrative Example" contains the programs to reproduce Figure 2.
2. "Section 5 Extensions" contains the programs to reproduce Figure 3. Figure 4 is purely descriptive and requires no code.

3. "Section 6 Illustrative Application" contains the programs to reproduce Figure 5, 6 , 7 and 8. Figure 8 is in the Online Appendix D but refer to section 6.
4. "Appendix C Numerical Analysis of Asymmetric Environments" replicates all numerical results and figures contained in the Online Appendix C.
5. "Software 1 Optimal Lola". This software is a visually handy procedure realized in Matlab that does not require IBM ILOG CPLEX and computes the optimal LoLA based on user specified inputs. This software is described in greater details in the Online Appendix E.1.
6. "Software 2 Optimal Mechanism". This software is realized in Matlab and IBM ILOG CPLEX. This application requires the same inputs as Software 1, and it computes the optimal mechanism even when that mechanism is not a LoLA. This software is described in greater details in the Online Appendix E.2.

5 Instructions to Replicators

To reproduce all results, it is sufficient to run the following programs:

- Figure2.m in "Section 2 Illustrative Example"
- Figure3.m in "Section 5 Extensions"
- For all figures in "Section 6 Illustrative Application":
    1. Run PreliminariesAndFigure5Noise.m to reproduce Figure 5
    2. Save output as LoblafromDataT100S7.mat
    3. In Figure6.m, Figure7.m, specify installation paths for AMPL (amplPath) and CPLEX (cplexPath)
    4. Run Figure6.m, Figure7.m, Figure8.m to reproduce the corresponding figures.

6 List of Tables, Figures, and Corresponding Programs

The two main programs main all simple.m in stylized and main all quant.m in quantitative reproduce all tables and figures in the paper and Online Appendix as well as the nu- merical results provided in the text. The following subsections provide further details on the specific programs that produce each output.

6.1 Tables

There are no tables in the paper.

6.2 Figures

All figures can be replicated by following the steps indicated in Section 5.

6.3 Numbers in Text

- The numerical results discussed in Section 6.4 are derived from Figure7, and Figure 7 can be replicated by running Figure7.m
- The numerical value of b_L discussed at the end of Section 6.4 can be derived by expression 53