

README for "Communicating Scientific Uncertainty via Approximate Posteriors" Andrews and Shapiro.

Overview

- This replication package contains the data and programs used in the paper referenced above, excluding data posted on trusted online repositories.
- The replication package can produce all figures and tables in the main text and Supplemental Appendix.
- Replication of calculations underlying Supplemental Appendix Figure 2 is time-intensive and is disabled by default. Instead, we have provided the generated data needed to reproduce the plot.
- Programs in this replication package include: Python, Stata, and Matlab. The replication package uses a Python package called `scons` to build all outputs based on Python configuration files.
- Expect total run time for the package is approximately 5 hours.

Data Availability and Provenance Statements

Summary of Availability

- [X] All data **are** publicly available.
- [] Some data **cannot be made** publicly available.
- [] No data **can be made** publicly available.

Details on each Data Source

The data sources are:

- Published data from [Nickerson, Fredrichs and King \(2006\)](#).
- A published [Bootstrap Census](#)
- Published replication packages for four articles in the *American Economic Review*.

[Nickerson, Fredrichs and King \(2006\)](#)

Hand-entered data from [Nickerson, Fredrichs and King \(2006, NFK\)](#) are used for the running example. These data are:

- findings from a get-out-the-vote experiment as reported in [Nickerson, Fredrichs and King \(2006, Table 3\)](#);
- the wage rate as reported in [Nickerson, Fredrichs and King \(2006, page 96\)](#); and
- the contacts per hour as reported in [Nickerson, Fredrichs and King \(2006, Table 6, row "Volunteer Phone Call," column "Contacts Per Hour."\)](#).

[Bootstrap Census](#)

The published [Bootstrap Census](#) contains published estimates and bootstrap replicates from a census of papers using the bootstrap published in the 2021 *American Economic Review*. The paper and appendix contain additional details about the papers and objects of interest included in the census.

Published Replication Packages

Most of our applications use only the published estimates and bootstrap replicates from the original articles which can be obtained from the Bootstrap Census. Supplemental Appendix Figure 2 requires additional calculations that use the published replication code and data for four articles. The corresponding replication packages are published on openICPSR. The articles are:

- Abebe, Caria, and Ortiz-Ospina (2021): [paper](#), [replication package](#)
- Braguinsky, Ohyama, Okazaki, and Syverson (2021): [paper](#), [replication package](#)
 - For this article, replication requires a custom Stata command which can be obtained from the authors' Github repository. We have included a static copy in the replication archive under `/datastore/raw/bootstrap_census/Braguinsky_Ohyama_Okazaki_Syverson_2021/orig/boos2-main.zip`.
- Dinerstein and Smith (2021): [paper](#), [replication package](#)
- Mueller, Spinnewijn, and Topa (2021): [paper](#), [replication package](#)

Details on Data Location within the Replication Archive

- Data from [Nickerson, Fredrichs and King \(2006\)](#) are located in `source/raw/example/data`.
- Data from [Bootstrap Census](#) are located in `source/raw/bootstrap_census`.
- Data from published replication archives can be obtained from the corresponding trusted published repositories.
 - To obtain these, for each paper in `datastore/raw/bootstrap_census`, follow the instructions in `/datastore/raw/bootstrap_census/AUTHORNAMES/readme.md` to download replication files as `/orig/AUTHORINITIALS2021.zip`, where `AUTHORINITIALS` are the first initials of the authors' last names in their order of authorship.
 - To verify `AUTHORINITIALS`, see the filename listed under `/orig` in the *Description* section of `/datastore/raw/bootstrap_census/AUTHORNAMES/readme.md`.
 - To verify that files are in the correct locations, files should be placed only in subfolders that have a file called `placeholder.txt`.

Computational Requirements

Software Requirements

Main analysis (excluding Supplemental Appendix Figure 2)

- [X] The replication package contains one or more programs to install all dependencies and set up the necessary directory structure.
- If typed at a system (e.g. bash or command) prompt, each command in `source/lib/`

JMSLab/builders/executables.yml (except *tablefill*) should start the corresponding software. If not, please ensure that (i) the path to the given command is part of the environment's path and (ii) the command name in `executables.yml` matches the command on your system.

- For example, typing `python` in a prompt (as below) should start Python. If it does not, first confirm that Python is on the path in your environment, then confirm that `executables.yml` gives the correct command for Python. If the latter fails, rename the command in `executables.yml`. For example, if your system calls Python with `python3`, change the corresponding line in `executables.yml` from `python: python` to `python: python3`.

ExamplePrompt> `python`

- Python (code was last run on version 3.10.9)
 - The file `source/lib/requirements.txt` lists all dependencies, please set up a virtual Python environment (e.g. via Anaconda) and run `pip install -r source/lib/requirements.txt` to install dependencies.
 - The Python package `BootstrapReport` is installed from github by default; a static snapshot is available in a [trusted repository](#).
 - The Python package `PyMC` requires `PyTensor` which in turn requires a C compiler; see [here](#) for an example of an error that can occur when the C compiler is not configured correctly.
- The code structure follows this public template: [JMSLab/Template](#). The documentation there provides more background on setup steps, execution, and requirements. The replication package is designed to be self-contained and does not require the public template to execute.

Additional requirements for Supplemental Appendix Figure 2

- Stata (code was last run with version 17)
 - Run `source/lib/requirements.do` to install dependencies.
- Matlab (code was run with Matlab R2021a)
 - The file `source/lib/requirements.m` lists all dependencies. Installation must be done manually through the MathWorks installer. When the script is run, it will halt and list any missing toolboxes.

Memory, Runtime, Storage Requirements

Summary

Approximate time needed for the main computations (excluding Supplemental Appendix Figure 2) on a standard (2025) desktop machine:

- [] <10 minutes
- [] 10-60 minutes
- [] 1-2 hours
- [X] 2-8 hours
- [] 8-24 hours

- ☐ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days

Approximate storage space needed:

- ☐ < 25 MBytes
- ☐ 25 MB - 250 MB
- ☐ 250 MB - 2 GB
- ☒ 2 GB - 25 GB
- ☐ 25 GB - 250 GB
- ☐ > 250 GB
- ☐ Not feasible to run on a desktop machine, as described below.

The computations to reproduce all output not included in the archive were last run on:

- **A MacBook Air (M1, 2020) running macOS Monterey version 12.3 and 9.35 GB of free space and 8GB of RAM.** The runtime was five hours.
- **A 60-core Slurm-managed HPC system with Rocky Linux version 8.10 with 53GB of free space and 15GB of memory per core.** The runtime was less than one hour.

The computations underlying Supplemental Appendix Figure 2, for which we have included output in the archive, require repeatedly rerunning the authors' original procedure and take several weeks on a high-performance computing cluster.

Description of programs/code

- `source/lib` contains utilities used throughout the pipeline
- `source/derived` cleans and organizes data for analysis and sends output to `/output/derived`
 - `source/derived/example`: bayes bootstrap replicates used in running example
 - `source/derived/bayes_bootstrap`: bayes bootstrap calculation for subset of Bootstrap Census papers
- `source/analysis` performs analysis and sends output to `/output/analysis`
 - `source/analysis/bayes_bootstrap`: compares SK distance across bootstrap methods
 - `source/analysis/bootstrap_census`: p-p plots and SK distance calculations for Bootstrap Census papers
 - `source/analysis/example`: plots for running example
 - `source/analysis/pp_plots`: p-p plots for Bootstrap Census papers
- `source/{tables, figures}` contain code and templates to build tables and figures in `/output/{tables, figures}`

Instructions to Replicators

These instructions presume that all software described in section "Software Requirements" has been installed

- Follow the instructions from "Details on Data Location within the Replication Archive"

to download data to `datastore/raw/bootstrap_census`.

- At a command prompt, type `scons`, as in the example below. This command uses Python configuration scripts to execute the package.

```
ExamplePrompt> scons
```

- (Optional) To reproduce the intermediate files needed for Supplemental Appendix Figure 2, uncomment the line `#SConscript('bayes_bootstrap/SConscript')` in `source/derived/SConscript` before running `scons`. This step takes several weeks on a high-performance computing cluster and has been excluded from the journal's replication process.

List of exhibits, in-text numbers, and programs

The tables below list, respectively, each exhibit and each in-text number in the paper and the supplemental appendix, with the corresponding location and main source script.

For more details on dependencies, type `scons -n --tree=prune PATH/TO/TARGET`. For example, `scons -n --tree=prune source/figures/example_posterior.lyx` prints the full dependency graph for Figure 2.

Exhibits

Figure/Table #	Program	Output file
Figure 1	None	<code>source/figures/example_prior.lyx</code>
Figure 2	<code>source/analysis/example/example_prior_dist.py</code> <code>source/analysis/example/example_posterior_effect.py</code> <code>source/analysis/example/example_posterior_risk.py</code>	<code>source/figures/example_posterior.lyx</code>
Figure 3	<code>source/analysis/example/example_sk.py</code>	<code>source/figures/example_sk.lyx</code>
Figure 4	<code>source/analysis/example/example_bb.py</code>	<code>source/figures/example_bb.lyx</code>
Figure 5	<code>source/figures/compile_pp_plots.py</code> <code>source/figures/LyxBuilder.py</code>	<code>output/figures/pp_plot_some.lyx</code>
Supplemental Appendix Figure 1	<code>source/analysis/example/example_bvm.py</code>	<code>source/figures/example_bvm.lyx</code>
Supplemental Appendix Figure 2	<code>source/analysis/bayes_bootstrap/bayes_scatter.py</code>	<code>source/figures/bayes_scatter.lyx</code>
Supplemental Appendix Figure 3	<code>source/analysis/bootstrap_census/summary_plots.py</code>	<code>source/figures/distance_sk.lyx</code>

Supplemental Appendix Table 1	source/tables/build_census.py	output/tables/census.lyx
-------------------------------	-------------------------------	--------------------------

In-text numbers

In-text number(s)	Program	Output file
Regret values from section 3.3	source/analysis/example/example_posterior_risk.py	output/autofill/example_risk_regret.tex
Signed variation bound in section 4.2	source/analysis/example/density_ratio_bounds.py	output/autofill/density_ratio_bounds.tex
SK metric values from section 4.3	source/analysis/example/example_sk.py	output/autofill/example_sk.tex
SK metric values from section 5.3	source/analysis/example/example_bb.py	output/autofill/example_bb.tex
Bootstrap Census statistics in section 6	source/analysis/bootstrap_census/make_autofills.py	output/autofill/bootstrap_census.tex

Sequential list of execution steps

For more details on dependencies, type `scons -n -Q` in a command prompt.

Below is a simplified list of the steps (excluding Supplemental Appendix Figure 2).

This list of steps assumes that:

- the command `python` executes Python from the system prompt
- the environment variable `PYTHONPATH` is set to `.`
 - on Linux and similar systems, this can be achieved by executing the command `export PYTHONPATH=.` at the system prompt
 - on Windows Powershell, this can be achieved by executing the command `$env:PYTHONPATH = "."` at the system prompt

```
python source/derived/example/make_data.py
python source/derived/example/estimate_mle.py
python source/derived/example/draw_bootstrap.py
python source/derived/example/draw_posterior.py
python source/analysis/bootstrap_census/calculate_statistics.py
python source/analysis/bootstrap_census/make_autofills.py
python source/analysis/bayes_bootstrap/bayes_scatter.py
python source/analysis/example/example_prior_dist.py
python source/analysis/example/example_posterior_costpervote.py
```

```
python source/analysis/example/example_posterior_effect.py
python source/analysis/example/example_bvm.py
python source/analysis/example/example_bb.py
python source/analysis/example/example_sk.py
python source/analysis/example/example_lb_regret.py
python source/analysis/example/example_posterior_risk.py
python source/analysis/bootstrap_census/summary_plots.py
python source/analysis/pp_plots/pp_plots.py
python source/figures/compile_pp_plots.py
python source/tables/build_census.py
python source/analysis/example/density_ratio_bounds.py
```

References

Abebe, Girm, A. Stefano Caria, and Esteban Ortiz-Ospina. 2021. "The Selection of Talent: Experimental and Structural Evidence from Ethiopia." *American Economic Review* 111 (6): 1757–1806. <https://doi.org/10.1257/aer.20201656>

Abebe, Girm, Caria, A. Stefano, and Ortiz-Ospina, Esteban. Data and Code for: The Selection of Talent: Experimental and Structural Evidence from Ethiopia. Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-05-26. <https://doi.org/10.3886/E127401V1> (Version 1).

Andrews, I., & Shapiro, J. M. (2025). Data and code for: Bootstrap Census (Version 1) [Data set]. Zenodo [publisher]. <https://doi.org/10.5281/zenodo.17085552>

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. 2021. "Product Innovation, Product Diversification, and Firm Growth: Evidence from Japan's Early Industrialization." *American Economic Review* 111 (12): 3795–3826. <https://doi.org/10.1257/aer.20201656>

Braguinsky, Serguey, Ohyama, Atsushi, Okazaki, Tetsuji, and Syverson, Chad. Data and Code for: Product Innovation, Product Diversification, and Firm Growth: Evidence from Japan's Early Industrialization. Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-11-19. <https://doi.org/10.3886/E145421V1> (Version 1).

Dinerstein, M., & Smith, T. D. (2021). Quantifying the supply response of private schools to public policies. *American Economic Review*, 111(11), 3583–3616. <https://doi.org/10.1257/aer.20151723>

Dinerstein, Michael, and Smith, Troy D. Data and Code for: Quantifying the Supply Response of Private Schools to Public Policies. Nashville, TN: American Economic Association [publisher], 2021. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-09-28. <https://doi.org/10.3886/E141201V1> (Version 1).

Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa. 2021. "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias." *American Economic Review* 111 (1): 324–63. <https://doi.org/10.1257/aer.20190808>

Mueller, Andreas I., Spinnewijn, Johannes, and Topa, Giorgio. Data and Code for: Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence and Bias. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-

university Consortium for Political and Social Research [distributor], 2020-12-17. <https://doi.org/10.3886/E120501V1> (Version 1).

Nickerson, D. W., Friedrichs, R. D., & King, D. C. (2006). Partisan mobilization campaigns in the field: Results from a statewide turnout experiment in Michigan. *Political Research Quarterly*, 59(1), 85–97. <https://doi.org/10.1177/106591290605900108>
