Replication Files for "Overinference from Weak Signals and Underinference from Strong Signals"

Ned Augenblick, Eben Lazarus, and Michael Thaler

Overview

This replication package contains the data and code needed to reproduce the experimental, empirical, and simulation analyses in the *Quarterly Journal of Economics* paper "Overinference from Weak Signals and Underinference from Strong Signals," by Augenblick, Lazarus, and Thaler.

Data

All data needed to produce the paper's tables and figures have been included in the replication package. For the experiments, this includes the anonymized raw experimental data. For the empirical analysis, this includes beliefs data that have been initially processed from the underlying sports-betting and options data following the description in the paper. The following tables describe the data structure and locations.

Experiments: Raw Data Files

Description	File Name	Location	Provided
Study 1a Data	experiment1a_data.csv	Experiment/data/	Yes
Study 1b Data	experiment1b_raw.csv	Experiment/data/	Yes
Study 2 Data	experiment2_raw.csv	Experiment/data/	Yes

Experiments: Intermediate Input Files

Description	File Name	Location	Provided
Study 1a Cleaned File	expt1a_cleaned.dta	Experiment/data/	Yes
Study 1b Cleaned File	experiment1b_data.dta	Experiment/data/	Yes
Study 2 Initial Input	experiment2_data.dta	Experiment/data/	Yes
Study 2 Cleaned File	expt2_cleaned.dta	Experiment/data/	Yes
File for Lit. Comparison	benjamin2019_data.dta	Experiment/data/	Yes (via Benjamin 2019)

Empirical Analysis: Data Files

Description	File Name	Location	Provided
Sports-Betting Data	ARData.dta	Empirical/BetFair/	Yes (via Augenblick and Rabin 2021)
Options-Based Data	ALData.dta	Empirical/Finance/	Yes (via Augenblick and Lazarus 2023, processed to obtain risk-neutral histograms)

Computational Requirements

Software Requirements and Settings

All code in the repository runs on Stata. The code was last run on Stata/MP 18.5 in October 2024.

The empirical code conducts a number of simulation and resampling analyses, for the model simulations reported in Section IV.A and the bootstrap-based inference in the subsequent empirical analysis. These three sets of simulations are independent from each other, so we set the random seed for Stata's pseudorandom number generator to 1 ("`set seed 1`") at the top of each of the three relevant files (`00simulatingData.do` in the `simulatedData` folder, `02regressionsBootstrap.do` in the `BetFair` folder, and `r06regressionsBootstrap.do` in the `Finance` folder).

The replication package also contains the necessary code to install all dependencies and set up the necessary subdirectory structures:

- For the experimental analyses, the program `config_stata.do` in the `Experiment` folder will install all dependencies locally. These include `cdfplot`, `coefplot`, `egenmore`, `erepost`, `estout`, `reghdfe`, and `addplot`.
- For the empirical analyses, the program `00MASTER.do` in the `Empirical` folder will install all dependencies locally. These include `unique` and `estout`. This program also sets up the needed subdirectories for later analysis.

Memory, Runtime, and Storage Requirements

The code has been tested on both a 2023 MacBook Pro (with Apple M2 Pro chip, 12 cores, and 32 GB of RAM), and on a 2022 Intel PC (with 4 cores and 16 GB of RAM). Approximately 3 GB of storage space are needed for the data provided in the replication package, and the provided code then saves files that require roughly an additional 5 GB of storage space.

The following estimated run times are based on code run on the 2023 MacBook Pro:

- Experimental analyses: Around 1 hour. The longest run times are for the cleaning and analysis for Study 1a.
- Empirical analyses: Around 2 days. The vast majority of this run time is for the bootstrap draws in `02regressionsBootstrap.do` in the `BetFair` folder and `r06regressionsBootstrap.do` in the `Finance` folder. To obtain point estimates with a much shorter run time, you can change the number of bootstrap draws to a small number at the top of each of these files (e.g., "`local numBootStraps=10`").

Code Description and Instructions

- Programs in `Experiment` load and clean the experimental data and conduct analyses to produce all tables and figures in Sections II–III of the paper and in Online Appendix B.1. All analyses can be run with the master file `Experiment/00masterExperiments.do`.
  - **Instructions:** Unzip the `Experiment` folder and open `Experiment/00masterExperiments.do`. Change the path (`cd` to the `Experiment` folder of whatever subfolder you save this package in) as directed at the top of the file. That master file can then be run to reproduce all the experimental analyses in the paper, the results of which will be output to `tables` and `figures`.
  - For further details, or to modify any particular aspect of the analysis for any of the studies, see the corresponding `do` file as listed in the `00masterExperiments.do` code. The master file contains an overview and description of each of the remaining `do` files, and each underlying `do` file is commented to describe the steps being taken and the table or figure it produces.
  - The order in which the code is run is not important for reproducing the experimental results shown in the paper, since all the intermediate data files are already included in the repository (in the `data` subfolder, and also see the `tables` and `figures` subfolders for output corresponding to each of the experiment-related tables and figures in the paper). If you change any aspect of the code for a given study, all code files for that study should be run in numerical order.
  - The `screenshots`, `irb`, and `PAPs` subfolders of `Experiment` contain additional details for the experiment. The `screenshots` folder contains all the experimental study materials shown in Figure IV and Online Appendix C. The remaining subfolders contain additional detail on the IRB approvals and pre-analysis plans.
- Programs in `Empirical` load and clean the empirical data, conduct model simulations, and conduct analyses to produce the tables and figures in Section IV of the paper and in Online Appendix B.2. All analyses can be run with the master file `Empirical/00MASTER.do`.
  - **Instructions:** Unzip the `Empirical` folder and open `Empirical/00MASTER.do`. Change the path (`cd` to the `Empirical` folder of whatever subfolder you save this package in) as directed at the

top of the file. That master file can then be run to reproduce all the empirical analyses in the paper, the results of which will be output to subfolders created in `simulatedData`, `BetFair`, and `Finance`.

- For further details, or to modify any particular aspect of the analysis for any of the studies, see the corresponding `do` file as listed in the `00MASTER.do` code. The master file contains a detailed overview and description of each of the remaining `do` files, and each underlying `do` file is commented to describe the steps being taken and the table or figure it produces.
- For each of the three sets of analysis in the `Empirical` folder (`simulatedData`, `BetFair`, and `Finance`), the code needs to be run in numerical order according to the `do` file name prefix. (For the `Finance` folder, start with the numbered programs, e.g. `01cleanData.do`, and then move to the programs prefixed with "`r`", e.g. `r01cleanDataRisk.do`.) This is because the intermediate files need to be constructed for each step of later analysis.

List of Tables and Figures and Corresponding Code

The table below provides more detail on the program responsible for producing the output shown in each table and figure in the paper. The output file names in the listed locations are straightforward (e.g., for column 1 of Table I, the output file is `table1_study1a.tex`), and the listed code files are commented so as to be able to find the code block that generates each corresponding table or figure. The code can be edited to undertake additional analyses as desired.

Figure/Table (and Col.)	Program	Location	Output Dir.
Fig. I–II, A1–A4	02study1aAnalysis.do	Experiment/	figures
Tab. I (1), II (1)–(3)	02study1aAnalysis.do	Experiment/	tables
Fig. III	03study1bAnalysis.do	Experiment/	figures
Tab. I (2)–(3), II (4)–(6), A1	03study1bAnalysis.do	Experiment/	tables
Fig. V	05study2Analysis.do	Experiment/	figures
Tab. I (4)–(5), A2	05study2Analysis.do	Experiment/	tables
Fig. VI	01generatePlots.do	Empirical/simulatedData/	Output
Fig. VII, A5, A7	01overTimeMR.do	Empirical/BetFair/	Output
Tab. III (1)–(5)	03regressionsBootstrapResults.do	Empirical/BetFair/	Output
Tab. A3–A4 (1)–(5)	03regressionsBootstrapResults.do	Empirical/BetFair/	Output
Fig. VIII (a), A6 (a), A8 (a)	03plotMain.do	Empirical/Finance/	Graphs
Fig. VIII (b), A6 (b), A8 (b)	r05plotMainRisk.do	Empirical/Finance/	Output
Tab. III (6)–(7)	r07regressionResults.do	Empirical/Finance/	Output
Tab. A3–A4 (6)–(7)	r07regressionResults.do	Empirical/Finance/	Output
Fig. A9	r08smoothedRiskAdjustedPlot.do	Empirical/Finance/	Graphs

References

Augenblick, Ned and Eben Lazarus (2023), "A New Test of Excess Movement in Asset Prices," Working Paper.

Augenblick, Ned and Matthew Rabin (2021), "Belief Movement, Uncertainty Reduction, and Rational Updating," *Quarterly Journal of Economics* 136, 933–985.

Benjamin, Daniel J. (2019), "Errors in Probabilistic Reasoning and Judgment Biases," in *Handbook of Behavioral Economics: Applications and Foundations*, ed. by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, vol. 2, Amsterdam: Elsevier, 69–186.