DATA AND CODE README

Forecasting the Results of Experiments:

Piloting an Elicitation Strategy

Stefano DellaVigna
UC Berkeley and NBER

Nicholas Otis
UC Berkeley

Eva Vivalt
Australian National University

This document:

1. **Data and Code:** A list of files for replication.
2. **Instructions:** Instructions to replicate tables and figure.
3. **List of Tables and Programs:** A mapping from code to tables and figure.
4. **Codebook:** An overview of variables included in the datasets.
5. **System information:** System specifications used when executing code.
6. **Acknowledgements:** Code acknowledgements.

1 Data and Code

1. Data.rda: This r data file contains data to replicate all tables and figures in the main text and appendix.
2. Analysis.r: This r script generates all tables and figures in the main text and appendix.
3. Data.csv: For non r users, we also provide the data in a .csv format.

2 Instructions

To generate the tables and figure open the file Analysis.r. Change the directory (line 29) to the folder containing the data. Execute the script to produce the paper- and appendix-tables and figure. The table below maps scripts and data sources to specific outputs.

3 List of Tables and Programs

Figure/Table #	Program	Line #
Table 1	Analysis.r	32-116
Table 2	Analysis.r	118-207
Table A1	Analysis.r	209-294
Table A2	Analysis.r	296-443
Figure 1	Analysis.r	445-461

4 Codebook

This section describes variables contained in the dataset Data.rda. Data are in "long" format: multiple forecasts from the same individual are stacked.

List of variables

1. `ResponseId`
   - Description: This variable uniquely identifies individual respondents.
2. `Background`
   - Description: This variable captures whether respondents identify as faculty, PhD students, researchers or practitioners.
3. `Group`
   - Description: This variable captures the pool respondents are drawn from (List serv=PP_ListServ ; Twitter= PP_Twitter ; All other groups = PP_xOtr)
4. `Pred`
   - Description: Predictions of treatment effects in standard deviations.
5. `AnySlider`
   - Description: Dichotomous variable taking a value of 1 for slider responses, and 0 for text entry responses.
6. `AnySD`
   - Description: Dichotomous variable taking a value of 1 for responses in SD, and 0 for responses in raw units.
7. `AnyWide`
   - Description: Dichotomous variable taking a value of 1 for responses using a wide slider scale and 0 for responses using a narrow slider scale (note: non-slider scales are also coded as 0, and must be omitted if you are interested in the effects of the slider scale size).
8. `TreatLargeRef`
   - Description: Dichotomous variable taking a value of 1 for responses following a large reference example, and 0 for responses following a small reference example.
9. `Cond`
   - Description: This variable specifies what outcome / treatment is being predicted.
10. `Treat`
    - Description: This variable specifies what outcome / treatment is being predicted *and* how predictions are elicited. For example "C_Sli_SD_Large_Food_T1" indicates that the response is for: food consumption from treatment 1 (Food_T1) from the Cash study (C_), elicited in SD using a large slider (Sli_SD_Large).
11. `C_Food_T1`
    - Description: Dichotomous variable taking a value of 1 for predictions of food consumption from the cash treatment, and 0 otherwise.
12. `C_Food_T2`
    - Description: Dichotomous variable taking a value of 1 for predictions of food consumption from the cash and asset treatment, and 0 otherwise.
13. `C_Food_T3`

- Description: Dichotomous variable taking a value of 1 for predictions of food consumption from the cash, asset, and nutrition treatment, and 0 otherwise.

14. `C_Hea_T1`
    - Description: Dichotomous variable taking a value of 1 for predictions of health consumption from the cash treatment, and 0 otherwise.
15. `C_Hea_T2`
    - Description: Dichotomous variable taking a value of 1 for predictions of health consumption from the cash and asset treatment, and 0 otherwise.
16. `C_Hea_T3`
    - Description: Dichotomous variable taking a value of 1 for predictions of health consumption from the cash, asset, and nutrition treatment, and 0 otherwise.
17. `H_HIV`
    - Description: Dichotomous variable taking a value of 1 for predictions of the effects of the health and education intervention on self-reported HIV testing, and 0 otherwise.
18. `E_Exam`
    - Description: Dichotomous variable taking a value of 1 for predictions of the effects of teacher training on exam scores, and 0 otherwise.
19. `E_Drop`
    - Description: Dichotomous variable taking a value of 1 for predictions of teacher training on student drop out (reverse coded), and 0 otherwise.
20. `E_Bus`
    - Description: Dichotomous variable taking a value of 1 for predictions of business participation (self-reports of earning money from a business in the last month), and 0 otherwise.

5 System Information:

The output below depicts version and session information used when generating the manuscript and appendix tables and figure.

```
               -
platform       x86_64-apple-darwin15.6.0
arch           x86_64
os             darwin15.6.0
system         x86_64, darwin15.6.0
status
major          3
minor          5.2
year           2018
month          12
day            20
svn rev        75870
language       R
```

```
version.string R version 3.5.2 (2018-12-20)
nickname       Eggshell Igloo

R version 3.5.2 (2018-12-20)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS  10.15.1

Matrix products: default
BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/
Versions/A/libBLAS.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] texPreview_1.4.3     knitr_1.26           plotrix_3.7-7        kableExtra_1.1.0     lmtest_0.9-37
 [6] zoo_1.8-4            multiwayvcov_1.2.3   data.table_1.12.8    expss_0.10.1         readxl_1.3.1
[11] DeclareDesign_0.20.0 estimatr_0.20.0      fabricatr_0.10.0     randomizr_0.20.0     forcats_0.4.0
[16] stringr_1.4.0        dplyr_0.8.3          purrr_0.3.0          readr_1.3.1          tidyr_1.0.0
[21] tibble_2.1.3         ggplot2_3.2.1        tidyverse_1.2.1

loaded via a namespace (and not attached):
 [1] httr_1.4.1         svgPanZoom_0.3.3   jsonlite_1.6       viridisLite_0.3.0  modelr_0.1.5
 [6] Formula_1.2-3      assertthat_0.2.1   cellranger_1.1.0   pillar_1.4.3       backports_1.1.5
[11] lattice_0.20-38    glue_1.3.1         digest_0.6.23      checkmate_1.9.1    rvest_0.3.5
[16] colorspace_1.4-1   sandwich_2.5-0     htmltools_0.4.0    pkgconfig_2.0.3    broom_0.5.2
[21] haven_2.1.1        magick_2.0         scales_1.1.0       webshot_0.5.2      whisker_0.3-2
[26] htmlTable_1.13.1   farver_2.0.1       generics_0.0.2     withr_2.1.2        lazyeval_0.2.1
[31] cli_2.0.0          magrittr_1.5       crayon_1.3.4       evaluate_0.14      fs_1.2.6
[36] fansi_0.4.0        nlme_3.1-137       xml2_1.2.2         foreign_0.8-71     tools_3.5.2
[41] hms_0.5.2          lifecycle_0.1.0    matrixStats_0.55.0 munsell_0.5.0      compiler_3.5.2
[46] rlang_0.4.2        grid_3.5.2         rstudioapi_0.10    htmlwidgets_1.3    labeling_0.3
[51] base64enc_0.1-3    rmarkdown_2.0      boot_1.3-20        gtable_0.2.0       rematch2_2.1.0
[56] R6_2.4.1           lubridate_1.7.4    zeallot_0.1.0      stringi_1.4.3      parallel_3.5.2
[61] Rcpp_1.0.3         vctrs_0.2.1        tidyselect_0.2.5   xfun_0.11
```

6 Acknowledgements

The code under the header “Install and load required packages” in the file Analysis.r is from Lars Vilhuber:

`https://gist.github.com/larsvilhuber/85026976027b58714c00420d75f04281`