README and Guidance

Overview

This is the ReadMe file for the replication of "Second-best fairness: The trade-off between false positives and false negatives" by Alexander W. Cappelen, Cornelius Cappelen, and Bertil Tungodden. The replication package uses survey data collected by the authors jointly with NORSTAT. The code in this replication package runs on Stata. One file runs all the 6 dofiles to clean and generate the data for 6 figures and 36 tables, including the online appendices. The replicator should expect the code to run in 2-8 hours.

Data Availability and Provenance Statements

Statement about Rights

I certify that the authors of the manuscript have documented permission to redistribute/publish the data contained within this replication package. The data has been collected by the authors. Appropriate permissions are documented in the Data/LICENSE.txt file.

License for Data

The data are licensed under a Creative Commons/cc-BY 4.0 license. See Data/LICENSE.txt for details.

Summary of availability

All data are publicly available as part of this replication package.

Details on each Data Source:

Raw_Data/surveyusa.dta and Raw_Data/surveynorway.dta data were collected by NORSTAT in the United States and Norway respectively, under contract with the authors. Identifying data were removed and no identifying information was used to estimate the results of the paper. This is the raw data for the experiments in the main paper.

Raw_Data/Norway_2019_raw.dta and Raw_Data/usa_2019_raw.dta were collected by NORSTAT in the United States and Norway respectively, under contract with the authors. Identifying data were removed and no identifying information was used to estimate the results of the paper. This is the raw data for the experiment in Study 1 – Online Appendix B.

Raw_Data/norway2015_raw.dta and Raw_Data/usa2015_raw.dta were collected by NORSTAT in the United States and Norway respectively, under contract with the authors. Identifying data were removed and no identifying information was used to estimate the results of the paper. This is the raw data for the experiment in Study 2 – Online Appendix B.

Processed_Data/studydata.dta appends the two raw datasets (Raw_Data/surveyusa.dta and Raw_Data/surveynorway.dta) and drops non-relevant variables for the analysis.

Processed_Data/analyticaldata.dta constructs the variables for the analysis (treatments, dependent variables, control variables).

Processed_Data/study20152019.dta cleans each dataset and constructs the main dependent variable, independent variables and background dummies. Then, it appends the fours datasets (Raw_Data/norway2019_raw.dta, Raw_Data/usa2019_raw.dta, Raw_Data/norway2015_raw.dta, Raw_Data/usa2015_raw.dta).

Processed_Data/analysis_20152019.dta constructs variables for further analysis.

Other documentation:

Survey/survey_instructions_english.pdf and Survey/survey_instructions_norway.pdf are the questionnaires used to collect Raw_Data/surveyusa.dta and Raw_Data/surveynorway.dta respectively. The files contain the experimental instructions for all 21 treatments in the paper and the additional questions (statements about unemployment benefits, income inequality, generosity, religion, females\males in leadership positions) and demographics for the entire sample.

Survey/Type1Type2_Compensation.qsf, Survey/Type1Type2_Compensation_High.qsf and Survey/Type1Type2_Earningsd.qsf are the files used to develop the survey instrument.

Documentation_type1type2 contains the study documentation of the paper:

- an overview of the study in the main text and the two studies in Appendix B, the countries covered, a list of the primary investigators, the sampling and weighting procedure (for the study in the main text), and the mode of data collections.
- a list of the 60 variables in Raw_Data/studydata.dta with the variable's name, type, format, non-missing observations (valid), missing observations (invalid), and the survey question associated with the variable.
- a list of the 31 variables in Raw_Data/study20152019.dta with the variable's name, type, format, non-missing observations (valid), missing observations (invalid), and the survey question associated with the variable.
- A description of each variable with information (type, format, Range, and missing), statistics (valid, invalid), definitions, literal questions from the survey and notes.

Computational requirements

Software Requirements

- Stata (code was last run with version 17)
  - `Estout`
  - `Coefplot`
  - `grc1leg2`
  - `rwolf2`
  - `wyoung`

Controlled Randomness

Random seed is set at line 22 of program ../Code/master_type1type2

Memory and Runtime Requirements

Summary

Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine:

- 2 - 8 hours

Details

The code was last run on a 12th Gen Intel(R) Core (TM) i5-1240P 1.70 GHz laptop with Windows 10 Enterprise version 21H2.

Description of programs/code

The dofile Code/master_type1type2.do installs all dependencies and runs the other dofiles in the correct sequence.

The dofile Code/cleaning_type1type2.do appends the two raw datasets, cleans the merged dataset and defines the variables for analysis.

The dofile Code/figures_type1type2_2022 generates all figures in the main body of the article.

The dofile Code/tables_type1type2_2022 generates the regression and estimated shares tables in the main text and Appendix A. At the end of the file, the code provides instructions on how to calculate the numbers in the running text.

The dofile Code/mht_type1type2 generates the multiple hypothesis test tables using the Holm-Bonferroni and Romano-Wolf corrections.

The dofile Code/ AppendixBcleaning2_type1type2.do cleans the four raw datasets for study 1 and study 2 in the Online Appendix B. It creates the main dependent variables and independent variables for the analysis and appends the four datasets together.

The Dofile/AppendixBanalysis_type1type2.do generates all figures and tables in the Online Appendix B. At the end of the file, the code provides instructions on how to calculate the numbers in the running text.

License for Code

Instructions to Replicators

To proceed with replication, first save the folder on your computer. Then open Stata ad select the folder as your working directory.

Open the master_type1type2.do and run the entire code. The first part of the code consists of the installation of the packages needed for analysis; the second part of the code runs the rest of the dofiles.

In Code/ tables_type1type2.do please note the following:

- Table 2 – household income reports the median household income. Household income is a categorical variable. The table output reports the median category for this variable. Median household income equals category 5 ($50,000 to $59,999) for the US and 9 (800.001-900.000 NOK) for Norway. The manuscript reports the rounded mean value of each category, thus a value of 55.000 (USD) for the US and 850.000 (NOK) for Norway.
- The household income categories can be found in Survey/survey_instructions_english.pdf and Survey/survey_instructions_norway.pdf for the American and Norwegian surveys respectively and are as follows:

US:
Under $20,000 - Category 1
$20,000 to $29,999 - Category 2
$30,000 to $39,999 - Category 3
$40,000 to $49,999 - Category 4
$50,000 to $59,999 - Category 5
$60,000 to $69,999 - Category 6
$70,000 to $79,999 - Category 7
$80,000 to $89,999 - Category 8
$90,000 to $99,999 - Category 9
$100,000 to $119,999 - Category 10
$120,000 to $149,999 - Category 11
$150,000 to $199,999 - Category 12
Over $200,000 - Category 13
Would rather not say - Category 14

Norway:
100.001-200.000 NOK - Category 2
200.001-300.000 NOK - Category 3
300.001-400.000 NOK - Category 4
400.001-500.000 NOK - Category 5
500.001-600.000 NOK - Category 6
600.001-700.000 NOK - Category 7
700.001-800.000 NOK - Category 8
800.001-900.000 NOK - Category 9
900.001-1.000.000 NOK - Category 10
1.000.001-1.100.000 NOK - Category 11
1.100.001-1.200.000 NOK - Category 12
1.200.001-1.300.000 NOK - Category 13
1.300.001-1.400.000 NOK - Category 14
1.400.001-1.500.000 NOK - Category 15
1.500.001 NOK eller mer - Category 16

Vil ikke svare - Category 17
Vet ikke - Category 18

In Code/mht_type1type2:

- the multiple hypothesis p-values tables are constructed manually as the commands `rwolf2` and `wyoung` do not allow output to be captured in tables or matrices.

In Code/AppendixBanalysis_type1type2.do:

- Table 2 – household income reports the median household income. Household income is a categorical variable. The table output reports the median category for this variable. Median household income equals category 5 ($50,000 to $59,999) for the US and 8 (700.001-800.000 NOK) for Norway in 2019 (study 1), and category 7 ($70,000 to $79,999) for the US and 7 (600.001-700.000 NOK) for Norway in 1025 (study 2). The manuscript reports the rounded mean value of each category, thus a value of 55.000 (USD) for the US in 2019, 750.000 (NOK) for Norway in 2019, 75.000 (USD) for the US in 2015, and 650.000 (NOK) for Norway in 2015.
- The income categories are as follows:

  US:
  Under $20,000 - Category 1
  $20,000 to $29,999 - Category 2
  $30,000 to $39,999 - Category 3
  $40,000 to $49,999 - Category 4
  $50,000 to $59,999 - Category 5
  $60,000 to $69,999 - Category 6
  $70,000 to $79,999 - Category 7
  $80,000 to $89,999 - Category 8
  $90,000 to $99,999 - Category 9
  $100,000 to $119,999 - Category 10
  $120,000 to $149,999 - Category 11
  $150,000 to $199,999 - Category 12
  Over $200,000 - Category 13
  Would rather not say - Category 14

  Norway:
  100.001-200.000 NOK - Category 2
  200.001-300.000 NOK - Category 3
  300.001-400.000 NOK - Category 4
  400.001-500.000 NOK - Category 5
  500.001-600.000 NOK - Category 6
  600.001-700.000 NOK - Category 7
  700.001-800.000 NOK - Category 8
  800.001-900.000 NOK - Category 9
  900.001-1.000.000 NOK - Category 10
  1.000.001-1.100.000 NOK - Category 11
  1.100.001-1.200.000 NOK - Category 12
  1.200.001-1.300.000 NOK - Category 13

1.300.001-1.400.000 NOK - Category 14
1.400.001-1.500.000 NOK - Category 15
1.500.001 NOK eller mer - Category 16
Vil ikke svare - Category 17
Vet ikke - Category 18

List of tables and programs

The provided codes reproduce selected tables and figures in the paper, as explained and justified below.

Figure/Table #	Program	Line Number	Output file	Note
Figure1_a	Code/figures_type1type2	30	../Figures/figure1_a.pdf	Upper left panel of Figure 1
Figure1_b	Code/figures_type1type2	47	../Figures/figure1_b.pdf	Upper right panel of Figure 1
Figure1_c	Code/figures_type1type2	65	../Figures/figure1_c.pdf	Middle left panel of Figure 1
Figure1_d	Code/figures_type1type2	83	../Figures/figure1_d.pdf	Middle right panel of Figure 1
Figure1_e	Code/figures_type1type2	101	../Figures/figure1_e.pdf	Lower left panel of Figure 1
Figure1_f	Code/figures_type1type2	119	../Figures/figure1_f.pdf	Lower right panel of Figure 1
Figure 1	Code/figures_type1type2	126	../Figures/Figure1.pdf	
Figure 2_a	Code/figures_type1type2	182	../Figures/figure2_a.pdf	Upper left panel of Figure 2

Figure2_b	Code/figures_type1type2	235	../Figures/figure2_b.pdf	Upper right panel of Figure 2
Figure2_c	Code/figures_type1type2	288	../Figures/figure2_c.pdf	Middle left panel of Figure 2
Figure2_d	Code/figures_type1type2	341	../Figures/figure2_d.pdf	Middle right panel of Figure 2
Figure2_e	Code/figures_type1type2	394	../Figures/figure2_e.pdf	Lower left panel of Figure 2
Figure2_f	Code/figures_type1type2	447	../Figures/figure2_f.pdf	Lower right panel of Figure 2
Figure 2	Code/figures_type1type2	452	../Figures/Figure2.pdf	
Figure3_a	Code/figures_type1type2	488	../Figures/figure3_a.pdf	Panel A of Figure 3
Figure3_b	Code/figures_type1type2	518	../Figures/figure3_b.pdf	Panel B of Figure 3
Figure 3	Code/figures_type1type2	522	../Figures/Figure3.pdf	
Figure 4_a	Code/figures_type1type2	541	../Figures/figure4_a.pdf	Upper left panel of Figure 2
Figure4_b	Code/figures_type1type2	640	../Figures/figure4_b.pdf	Upper right panel of Figure 2
Figure4_c	Code/figures_type1type2	560	../Figures/figure4_c.eps	Middle left panel of Figure 2
Figure4_d	Code/figures_type1type2	693	../Figures/figure4_d.pdf	Middle right panel of Figure 2

Figure4_e	Code/figures_type1type2	579	../Figures/figure4_e.eps	Lower left panel of Figure 2
Figure4_f	Code/figures_type1type2	746	../Figures/figure4_f.pdf	Lower right panel of Figure 2
Figure4_a_c_e	Code/figures_type1type2	586	../Figures/figure4_a_c_e.pdf	Left panels of Figure 4
Figure 4_b_d_f	Code/figures_type1type2	750	../Figures/figure4_b_d_f.pdf	Right panels of Figure 4
Figure 4	Code/figures_type1type2	756	Figure 4: Unemployment-experiment	
Table2a	Code/tables_type1type2	117	../Tables/Table2a.tex ../Tables/Table2b.tex ../Tables/Table2c.tex	Table 2a – descriptive statistics for income (expressed as categories in the code, and as mean value of category in the manuscript) Table 2b – descriptive statistics for education, age, gender, and political ideology Table 3a – lists the number of observations
Table2b	Code/tables_type1type2	145		
Table2c	Code/tables_type1type2	162		
Table 2	Code/tables_type1type2			Constructed by manually appending Tables 2a, 2b and 2c

Table 3	Code/tables_type1type2	182	../Tables/Table3.tex	
Table 4	Code/tables_type1type2	287	../Tables/Table4.tex	
Table5a	Code/tables_type1type2	304	../Tables/Table5a.tex ../Tables/Table5b.tex	Panel A Table 5
Table5b	Code/tables_type1type2	324		Panel B Table 5
Table 5	Code/tables_type1type2			Constructed by manually appending Table 5a and 5b
Table6a	Code/tables_type1type2	376	../Tables/Table6a.tex ../Tables/Table6b.tex	Political difference shares in Table 6
Table6b	Code/tables_type1type2	402		Country difference shares in Table 6
Table 6	Code/tables_type1type2			Constructed by manually appending Table 6a and 6b
Table 7	Code/tables_type1type2	419	../Tables/Table7.tex	
TableA2a	Code/tables_type1type2	447	../Tables/Tablea2a.tex using Tablea2b.tex	Regression results in Table A2
TableA2b	Code/tables_type1type2	458		F-test in Table A2
Table A2	Code/tables_type1type2			Constructed by manually appending Table A2a and A2b

TableA3a	Code/tables_type1type2	475	../Tables/Tablea3a.tex ../Tables/Tablea3b.tex	Panel A Table A3
TableA3b	Code/tables_type1type2	496		Panel B Table A3
Table A3	Code/tables_type1type2			Constructed by manually appending Table A3a and A3b
Table A5	Code/tables_type1type2	522	../Tables/Tablea5.tex	
Table A7	Code/tables_type1type2	543	../Tables/Tablea7.tex	
Table A9	Code/tables_type1type2	570	../Tables/Tablea9.tex	
Table A11	Code/tables_type1type2	640	../Tables/Tablea11.tex	
Table A12	Code/tables_type1type2	656	../Tables/Tablea12.tex	
Table A13	Code/tables_type1type2	675	../Tables/Tablea13.tex	
TableA14a	Code/tables_type1type2	737	../Tables/Tablea14a.tex ../Tables/Tablea14b.tex	Political difference shares in Table A14
TableA14b	Code/tables_type1type2	773		Country difference shares in Table A14
Table A14	Code/tables_type1type2			Constructed by manually appending Table A14a and A14b
TableA15a	Code/tables_type1type2	788	../Tables/Tablea15a.tex. ./Tables/Tablea15b.tex	Panel A of Table A15
TableA15b	Code/tables_type1type2	793		Panel B of Table A15

Table A15	Code/tables_type1type2			Constructed by manually appending Tables A15a and A15b
Table A16	Code/tables_type1type2	813	../Tables/Tablea16.tex	
Table A17	Code/tables_type1type2	837	../Tables/Tablea17.tex	
Table A18	Code/tables_type1type2	861	../Tables/Tablea18.tex	
Table A19	Code/tables_type1type2	885	../Tables/Tablea19.tex	
Table A20	Code/tables_type1type2	909	../Tables/Tablea20.tex	
Table A21	Code/tables_type1type2	933	../Tables/Tablea21.tex	
Table A22	Code/tables_type1type2	965	../Tables/Tablea22.tex	
Table A23	Code/tables_type1type2	988	../Tables/Tablea23.tex	
Table A1	Code/mht_type1type2	18	Output table produced manually because `rwolf2` and `wyoung` do not allow output to be captured in tables or matrices.	
Table A4	Code/mht_type1type2	40	Output table produced manually because `rwolf2` and `wyoung` do not allow output to be captured in tables or matrices.	
Table A6	Code/mht_type1type2	59	Output table produced manually because `rwolf2` and `wyoung` do not allow output to be captured in tables or matrices.	
TableA8a	Code/mht_type1type2	78	Output table produced manually because	Panel A of Table A8

TableA8b	Code/mht_type1type2	99	`rwolf2` and `wyoung` do not allow output to be captured in tables or matrices.	Panel B of Table A8
Table A8	Code/mht_type1type2			Constructed by manually appending Table A8a and A8b
Table A10	Code/mht_type1type2	118	Output table produced manually because `rwolf2` and `wyoung` do not allow output to be captured in tables or matrices.	
TableB1a	Code/AppendixBanalysis_type1type2	107	../Tables/TableB1a.tex ../Tables/TableB1b.tex ../Tables/TableB1c.tex	Table B1a – descriptive statistics for income (expressed as categories in the code, and as mean value of category in the manuscript) Table B1b – descriptive statistics for education, age, gender, and political ideology Table B1a – lists the number of observations
TableB1b	Code/AppendixBanalysis_type1type2	144		
TableB1c	Code/AppendixBanalysis_type1type2	169		
Table B1	Code/AppendixBanalysis_type1type2			Constructed by manually appending Table B1a, B1b, B1c

Table B3	Code/AppendixBanalysis_type 1type2	189	../Tables/TableB3.tex	
Table B4	Code/AppendixBanalysis_type 1type2	443	../Tables/TableB4.tex	
Table B5	Code/AppendixBanalysis_type 1type2	464	../Tables/TableB5.tex	
Table B6	Code/AppendixBanalysis_type 1type2	521	../Tables/TableB6.tex	
Table B7	Code/AppendixBanalysis_type 1type2	541	../Tables/TableB7.tex	
barmeansstudy1	Code/AppendixBanalysis_type 1type2	567	../Figures/barmeansstudy1.pdf	Upper left panel of Figure B1
barmeansstudy2	Code/AppendixBanalysis_type 1type2	637	../Figures/barmeansstudy2. pdf	Upper right panel of Figure B1
barmeansUSstudy1	Code/AppendixBanalysis_type 1type2	592	../Figures/barmeansUSstudy1. pdf	Middle left panel of Figure B1
barmeansUSstudy2	Code/AppendixBanalysis_type 1type2	660	../Figures/barmeansUSstudy2. pdf	Middle right panel of Figure B1
barmeansNorwaystudy1	Code/AppendixBanalysis_type 1type2	617	../Figures/barmeansNorwaystudy1. pdf	Lower left panel of Figure B1
BarmeansNorwaystudy2	Code/AppendixBanalysis_type 1type2	687	../Figures/barmeansNorwaystudy2. pdf	Lower right panel of Figure B1
FigureB1	Code/AppendixBanalysis_type 1type2	691	../Figures/FigureB1. pdf	
classification_1	Code/AppendixBanalysis_type 1type2	737	../Figures/classification_1. pdf	Upper left panel of Figure B2

classification_2	Code/AppendixBanalysis_type 1type2	775	../Figures/classification _2.pdf	Upper right panel of Figure B2
classification_1US	Code/AppendixBanalysis_type 1type2	814	../Figures/ classification_1US.pdf	Middle left panel of Figure B2
classification_1N	Code/AppendixBanalysis_type 1type2	853	../Figures/classification _1N.pdf	Middle right panel of Figure B2
classification_2US	Code/AppendixBanalysis_type 1type2	892	../Figures/classification _2US.pdf	Lower left panel of Figure B2
classification_2N	Code/AppendixBanalysis_type 1type2	929	../Figures/classification _2N.pdf	Lower right panel of Figure B2
FigureB2	Code/AppendixBanalysis_type 1type2	934	../Figures/FigureB2.pdf	

References

Cappelen, Alexander W., Cornelius Cappelen, and Bertil Tungodden." Second-best fairness: The trade-off between false positives and false negatives." American Economic Review (forthcoming)

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." Scandinavian Journal of Statistics, 6(2): 65–70

Romano, Joseph P., and Michael Wolf. 2010. "Balanced control of generalized error rates." The Annals of Statistics, 38(1): 598–633.