Readme: High-Pressure, High-Paying Jobs?

Markus Nagler Johannes Rincke Erwin Winkler

July 10, 2023

This ReadMe contains a description of the data sets and programs used for our analysis. We used Stata16 for our analysis, based on Microsoft Windows 11.

1 Data required for replication of our results

The results of the first part of our analysis are based on data from the Federal Institute for Vocational Education and Training (BIBB) and their BIBB/BAuA Employment Surveys. We are not allowed to upload these data to the replication archive. However, it can be easily obtained for a small fee through the BiBB and the GESIS Leibniz Institute for the Social Sciences, an infrastructure institute for the social sciences that stores the data.

The data for the survey waves from 1979 to 2018 can be obtained via GESIS, accessible either directly (`https://www.gesis.org/en/services/finding-and-accessing-data`, searching for "BIBB/BAuA-Employment Survey", last accessed 3rd July 2023) or via these links: `https://www.bibb.de/en/15182.php` and `https://www.bibb.de/en/14781.php` (last accessed 3rd July 2023) . These links also contain codebooks that define all variables. All changes to these variables are documented in the do-files listed below.
The dataset "kldb88_3_10_5.dta" included in this replication kit contains a crosswalk between two occupation classifications (kldb88 and kldb10) that we use during the data preparation process.

The second part of the analysis was conducted with self-collected survey data including a stated choice experiment. These data are part of this replication kit. We performed the main analysis on the subset of respondents who passed all attention checks. The

corresponding dataset is named "HighPressure_attent.dta" in the replication kit. We also provide the data on all participants in our experiment (dataset "HighPressure_all.dta") These two datasets allow for a full replication of the second part of our results using the do-file described in the next section.

2 Do-files in the replication kit

This replication kit contains a master do-file, named "0_Master.do". In this do-file, the central path to folders can be changed. In addition, the master do-file calls all other do-files necessary for replication of our results. The do-file "0_setup.do" creates the general setup and installs all necessary programs. To execute the master do-file, one needs to create three folders: one folder called "do" containing all do-files in the kit, one folder called "data", and one folder called "orig" containing all original datasets.

The do-file "1_exp_analyze.do" generates all results which are based on the self-collected experimental data.

All do-files that contain "obs" in their name are for the analysis of the BIBB data, which we are not allowed to share but which can be easily obtained as explained above. These are commented out in the master do-file at the moment. The do-file "2_obs_prepare.do" prepares the 2018 wave of the BIBB data. It calls two do-files which are used to translate variable and value labels to English, named "2b_obs_translation_ETB2018_valuelabels_EN.do" and "2c_obs_translation_ETB2018_variablelabels_EN.do". The do-file "3_obs_analyze.do" replicates all tables and figures based on the 2018 wave of the BIBB data. The do-file "4_obs_timetrend.do" prepares the data and produces all results which are based on multiple waves of the BIBB data. The do-file "5_obs_rforest_health.do" uses a random forest to predict the health index and, analogously, "6_obs_rforest_pressure.do" predicts the pressure index. At the top of every do-file, we mention which tables and figures are replicated in this do-file.

3 Data dictionary

The self-collected survey data that we provide as part of this replication kit contain too many variables to list below. In the dataset, however, each variable contains a label in the

data set that explains the underlying definition of the variables. In these labels, the prefix "survey:" indicates survey information on respondents while the other variables relate to choices about hypothetical jobs these respondents make. Note that differences between job attributes are defined such that positive values indicate better amenities in Job A. This is why some labels are shown as "Difference A-B" while others (commuting, hours, where less is supposedly better) are shown as "Difference B-A".

For the analyses using the BiBB data, we include two do-files in the replication kit that label variables and values, as described above. In addition, for each BiBB dataset, there is a separate codebook defining all variables (see links above).