Data and coding instructions for "Automation Enables Specialization: Field Evidence"

The paper uses two datasets.

1. Taxi drivers. This is a primary survey dataset collected by the authors.
    a. Dataset: taxi.dta
    b. Stata do file: taxi_code.do. This file produces the summary statistics of the taxi driver survey and the main results on map app usage. The produced results are shown in Online Appendix Table A1 and A2.
    c. Original survey: taxi-survey.pdf

2. Supermarket cashiers. This is a proprietary dataset and governed by a non-disclosure agreement with the collaborating supermarket group. We disguised the data to protect sensitive information (such as worker identifiers, transaction value, purchased quantity, etc). We provide the disguised dataset and do file that replicate the main results. Due to the manipulation, some of the coefficients can be replicated to one decimal place. The main results and conclusion remain unchanged.
    a. Dataset:
        i. cashier.dta: for Tables 1 to 4, Figures 2 and 3, and Online Appendix Table B2-B5, B7 (columns (d) and (e)).
        ii. cashier-wt0.dta: for Online Appendix Table B7, columns (a)-(c).
    b. Stata do file: cashier_code.do. This file produces the summary statistics and main regression analyses of the cashier study. The produced results are shown in Tables 1 to 4, Figures 2 and 3, and Online Appendix Table B2-B5, B7.[1]
    c. Description of disguised data: cashier-disguise.pdf

[1] In order to protect sensitive information such as transaction value, quantity and average price, we manipulated the relevant variables. As a result, the disguised dataset does not allow for replication of Online Appendix Tables B6.