Code and pseudo-data for Kisseleva, Mjøs and Robinson: "Evaluation Selection Bias in Early-Stage Investment Returns", Journal of Financial and Quantitative Analysis, 2025

This study uses a fully anonymized pseudo-data set derived from the original sample of startup firms. The pseudo-data are created by a random draw of 379 firms from the combined population of startup firms plus 500 other firms started during 2004–2018. After removing duplicates, the resulting pseudo-sample comprises 876 firms. Pseudonymous IDs replace all actual identifiers for firms, parent companies, and all investors to protect confidentiality. Identity disclosure is prohibited when using research-data provided by the Norwegian Tax Authority.

All data preparation, analysis, and production of outputs for the pseudo-data are conducted using the original Stata do-files used for the analysis in the paper and in the supplementary material. The do-files generate separate log-files and export outputs (Excel files and PNG figures) as required for the manuscript. Reported results from the pseudo-data analysis differ from those in the published paper due to the sampling design, and some sub-sample statistics are omitted because not all categories of firms and investors are represented in the pseudo-sample.

Folder structure (to be defined at the start of each do-file as required):

- pseudo_data – includes data files in .dta-format and the subfolders which include:
    - do-files (complete)
    - figures (empty)
    - logs (empty)
    - tables (empty)

Despite anonymization, the input data used in the preparatory stage cannot be shared due to confidentiality and re-identification risk. The pseudo-data, associated do-files, and generated outputs are provided to enable replication of tables and figures, but underlying real-world data remain unavailable under applicable data-use restrictions (e.g., Norwegian Tax Authority requirements).

Data preparation and inputs (do-files only):

- 0_sample_data_prep_pseudo.do — Preparing the data for analysis (first-stage data preparation)
- 1_investortypes_pseudo.do — Defining investor types
- 2_valuations_pseudo.do — Firm valuations
- 3_founding_dates.do — Firm founding dates

The following do-files execute the tables and figures and the necessary data-files are included in the folder:

B: Creating the specific outputs used in the paper – do-files and outputs:

- Tables 1-3:
    - Tables1_3 _pseudo.do
        - Tables 1 and 3 require manual transcription/formatting from the corresponding log-file outputs.

        - Table 2 is produced and exported in Excel format.

- Table 4:
    - Table4_pseudo.do
        - Exported to Excel; labels and formatting should be added manually.
- Table 5:
    - Table5_pseudo.do
        - Exported to Excel; labels and formatting to be completed manually.
- Table 6:
    - Table6_pseudo.do
        - Exports to separate Excel files for equal-weighted and capital-weighted returns, following the structure of Table 6 in the paper, with unwinsorized means appended at the end. Labels and formatting require manual completion.
- Table 7:
    - Table7_pseudo.do
        - Exported to Excel; labels and formatting to be completed manually.
- Figure 3:
    - Figure3_pseudo.do
        - Statistics used to construct the figure are exported to the folder "figures." The figure is saved as Fig3.png for the manuscript; alternative versions are prepared as Figure3_99win.png (99% winsorization) and Figure3_nowins.png (no winsorization).
- Figure 4 & 5:
    - Figures4_5_A2_BA_pseudo.do
        - Statistics exported to the folder "figures."
        - Figure4_mv_rounds.png and Figure5_mv_rounds.png are saved accordingly.

C. Figures used in the supplementary material – online appendix:

- Figure A1:
    - Figure3_pseudo.do
        - Saved as FigureB1.png.
- Figure A2:
    - Figures4_5_A2_BA_pseudo.do
        - Saved as Figure4_mv_all.png.
- Figure A3:
    - Figures4_5_A2_BA_pseudo.do
        - Saved as Figure5_mv_all.png.