Replication Instructions

Task Mismatch and Salary Penalties: Evidence from the Biomedical PhD Labor Market

Holden A. Diethorn & Gerald R. Marschke

NOTE: If downloading from Dataverse, the file DM_TaskMismatch.zipfile must be renamed to DM_TaskMismatch.zip before extracting.

Data Availability

Confidential Microdata

The core microdata used in this paper is confidential and requires a license with the National Center for Science and Engineering Statistics (NCSES). Users with an approved data license can access the requested data via a virtual data enclave. The following datasets are required to replicate the results in the present paper:

- **Survey of Doctorate Recipients (SDR)** – All survey waves between 1993 and 2017, including the international SDR files in 2010 and 2013.
- **1991 SDR Longitudinal File** – Data from SDR waves prior to 1993.
- **Doctorate Records File (DRF)** – Data from the annual Survey of Earned Doctorates (SED).

Aggregated Statistics Generated from Confidential Microdata

Aggregated statistics generated from the core microdata underlying the figures appearing in the paper are available in folder `6_data/external`:

- `predicted_salary_profiles.xlsx` — data underlying the predicted experience-earnings profiles in Figures 1, 2, & A.4. Code `predicted_salary_profiles.do` generates all predicted salary plots from the paper using this data.
- `tasks.xlsx` — task data underlying Figures 3 & 4. Code `Fig3.R` & `Fig4.R` generates figures.
- `FigA2.xlsx` — data underlying Figure A.2. Code `FigA2.do` generates figure.
- `FigA3.xlsx` — data underlying Figure A.3. Code `FigA3.do` generates figure.

Publicly-Available Data

Publicly-available data used in paper:

- `SED_bio_phd_count.xlsx` — data derived from public-use SED data downloaded via https://ncsesdata.nsf.gov/builder/sed?type=chart and used to construct Figure A.1. Available in folder `6_external` along with code `FigA1.do`.
- CPI-U data downloaded from FRED (https://fred.stlouisfed.org/graph/?g=1Ivda) was used to convert the nominal annual salary of SDR respondents to real values (in 2018 dollars).

Replication Instructions after Data Access

Before replication, user must edit the working directory in programs `master.do` and `preamble.do`. Running the file `master.do` will run all Stata programs used to construct data and produce the results in the paper's figures and tables.[1] Folders `1_data/raw`, `1_data/intermediate`, `3_figures`, `4_tables`, and `5_logs` are intentionally empty; they serve as locations where data and results produced from the programs in `2_programs` will be stored after running `master.do`. All programs are located in folder `2_programs` unless explicitly stated otherwise (i.e., Figure 3 and Figure 4).

Replication Programs

Data Construction

To construct the datasets needed as input to programs that generate the tables and figures in the paper, the user must run the following programs in the order that they appear:

- `00_sdr15.sas`
- `00_sdr17.sas`
- `00_sdrdrf17.sas`
- `00_compress.do`
- `01_sdr_panel.do`
- `02_identify_postdocs.do`
- `03_var_construction.do`
- `04_baseline_sample.do`
- `05_task_tenure_vars.do`
- `06_task_mismatch_vars.do`
- `07_task_summary_vars.do`

Tables and Figures in Main Text

The following programs construct the tables in the main body of the paper:

- **Table 1:** `Table_1.do`
- **Table 2:** `Table_2.do`
- **Table 3:** `Table_3.do`
- **Table 4:** `Table_4.do`
- **Table 5:** `Table_5.do`
- **Table 6:** `Table_6.do`
- **Table 7:** `Table_7.do`
- **Table 8:** `Table_8.do`

The following programs construct the figures in the main body of the paper:

- **Figures 1 & 2:** `Fig_1A_2A.do` & `Fig_1B_2B.do`
- **Figure 3:** `6_external/Fig_3.R`
- **Figure 4:** `6_external/Fig_4.R`

[1] Programs ending in `.sas` or `.R` will need to be run separately in SAS and R, respectively.

Tables and Figures in Appendices

The following programs construct the tables in the appendices:

- **Table A1:** `Table_A1.do`
- **Table A2:** `Table_A2.do`
- **Table A3:** `Table_A3.do`
- **Table A4:** `Table_A4.do`
- **Table A5:** `Table_A5.do`
- **Table A6:** `Table_A6.do`
- **Table C1:** `Table_C1.do`
- **Table C2:** `Table_C2.do`
- **Table D1:** `Table_D1.do`
- **Table D2:** `Table_D2.do`
- **Table D3:** `Table_D3.do`
- **Table F1:** `Table_F1.do`
- **Table G1:** `Table_G1.do`
- **Table G2:** `Table_G2.do`

The following programs construct the figures in the appendices:

- **Figure A1:** `Fig_A1.do`
- **Figure A2:** `Fig_A2.do`
- **Figure A3:** `Fig_A3.do`
- **Figure A4:** `Fig_A4.do`