Replication Files for "German Long-Term Health Insurance: Theory Meets Evidence"

Juan Pablo Atal, Hanming Fang, Martin Karlsson and Nicolas R. Ziebarth

This readme file describes the data and replication codes in the main body of the paper and in the Online Appendix. The attached table "Instructions_Addendum_AFKZ.xlsx" provides a detailed list with all codes needed to produce each table and figure in the main text and Online Appendix, as well as the additional results described in the text that do not belong to tables or figures. Questions can be directed to ataljp@econ.upenn.edu

1. Input (raw) data

In total, the paper uses four datasets: two publicly available datasets (SOEP, PSID) and two proprietary claims datasets from a single private insurer and a single public insurer (PKV and GKV). We describe the main use of each data below and instructions on how to access it.

- **SOEP data**
  - **data/raw/SOEP/pl.dta**
  - **data/raw/SOEPpgen.dta**
  - **data/raw/SOEP/pequiv.dta**

  These data are publicly available, but researchers need to apply for access and sign a Data Use Agreement. Instructions on how to obtain data access are provided in Section 3.

- **PSID data**
  - **data/raw/PSID/pequiv_long.dta**

  These data are publicly available, but researchers need to apply for access and sign a Data Use Agreement. Instructions on how to obtain data access are provided in Section 3.

- **PKV**
  - **raw/PKV/MainPKVSample.dta**

  Claims panel data from a big PKV insurer. This data is proprietary. Access requires submitting an application to the insurer. Instructions on how to obtain data access are provided in Section 3.

- **GKV**
  - **raw/GKV/ ACG2010.csv**
  - **raw/GKV/ ACG2011.csv**

  Processed claims panel data from a big GKV insurer are proprietary, containing ACG scores. This data is proprietary. Access requires submitting an application to the insurer. Instructions on how to obtain data access are provided in Section 3.

Deleted: dta

Deleted: dta

2. Codes

Income paths

Codes to construct the life-cycle income path for both educational groups in Germany and the US.

- **1. SOEP1984_2016_prepare.do**
- **2. income-process_SOEP.do**
- **3. PSID1984_2015_prepare.do**
- **4. income-process_PSID.do**

Health process

- **5. HealthRiskPKV.do**

Input dataset: "MainPKVSample.dta"

This code and dataset carries out the risk classification described in Section 5 and then produces the main inputs for the analysis from the health process.

Main results (Section 6.1-6.5)

The following codes in the "**calibrations"** folder produce the main results of the paper (Tables and Figures in Sections 6.1-6.5):
- **a_code_suite.m**: produces results
- **b_TabsandFigs.m**: exports results to .xlsx files
- **6_makefig_main.do**: produces figures.

Robustness analysis (Section 6.4.1).

The following codes in the "**calibrations"** folder produce the robustness results discussed in Section 6.4.1. They are listed in order in which the corresponding results appear in the paper, along with the title of the paragraph in the paper where the results are presented. These codes can be run independently, after running a_code_suite.m

- **c_robustness_gamma.m**: "Risk aversion"
- **d_robustness_crra.m**: "CRRA preferences"
- **e_robustness_EZ.m**: "Epstein-Zin Recursive Preferences"
- **f_robustness_gkv.m**: "Initial Health Status Representative of Germany"
- **g_robustness_sampler_probs.m**: "Different Starting States"
- **h_robustness_savings**: "Savings"
- **i_robustness_PSID.m**: "Income Profiles"
- **j_robustness_savings_PSID**: "Income Profiles" (results allowing for savings).

Section 6.5

- **k_robustness_state7.m**

Other Appendix Results:

Please see InstructionsAddendum.xlsx for the details on what is produced with each code

- **l_ arrow_security_suite.m.** Produces the results for the Arrow Securities in Appendix A.7
- **7_HealthCatDesc.do**: Summary statistics for health risk for PKV population.
- **8_SOEP_analsysis_replication_final**: Analysis of switching between PKV and GKV, and summary statistics for SOEP data and comparison with PKV population.
- **9_Descriptives_PKV.do**: Summary statistics for PKV population and health risks
- **10_makefig_appendix.do**: Processes results to produce appendix figures (lapsation, robustness to starting states, and calibrated premiums).
- **11_FinalGKV.do**: Summary statistics for GKV population and health risks, and comparisons with PKV population.
- **12_PremiumComparison.do**: Figure of calibrated v.s. observed starting premiums

3. Data Access

3.1 SOEP

For researchers residing outside of Germany, an international 95% version of the SOEP can be ordered. The instructions for data access are here:
https://www.diw.de/en/diw_01.c.601584.en/data_access.html

We use the following version as cited in the paper:

> "SOEP (2018). Data for years 1984-2016. version 33, SOEPlong, 2018 doi:10.5684/soep.v33"
> https://www.diw.de/sixcms/detail.php?id=diw_01.c.743253.en

3.2 PSID

As, for the main analysis, we rely on several SOEP income measures that were cleaned, checked and generated by the SOEP Group in Berlin, we use the equivalent income measures for the PSID. This is feasible thanks to the international data harmonization project CNEF based at Ohio State University:
https://www.cnefdata.org/

And as described here:

> Frick, J. R., S. P. Jenkins, D. R. Lillard, O. Lipps, and M. Wooden (2007). The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Journal of Applied Social Science Studies (Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften)* 127(4), 626–654.

The data can be ordered here:
https://www.cnefdata.org/data/data-application

The main PSID files can be downloaded for free directly here:
https://simba.isr.umich.edu/data/data.aspx

3.3 PKV and GKV

The paper uses two proprietary claims datasets: one from a private PKV insurer and one from a public GKV insurer.

Access to these data is granted by the respective insurers, who have both requested that their identity is not made public. Researchers interested in using the data may get in touch with the administrator at the Chair of Health Economics, University of Duisburg-Essen (sekretariat.goek@ibes.uni-due.de), who will assist the application for access.

4. Software

For both claim datasets, we have used the German version of the ACG software, which may be acquired from Johns Hopkins University. Instructions may be found here:

https://www.hopkinsacg.org/