README.pdf

Elisca Mastenbroek

July 25, 2023

1 Introduction

This GitHub repository or folder contains the R code that is used for the MSc thesis 'Correcting for measurement error in the Employment Register and the Labour Force Survey using three latent variable models' by Elisca Mastenbroek.

In this README.pdf file, instructions are provided to reproduce the results of the two simulation studies (see Chapters 4-5), the analyses of the real data (see Chapter 6), and the initial analyses in Sections 5.1.1 and 5.1.2. Note that the real data from the ER and the LFS is not publicly available.

General remark: The sample size is referred to as n throughout the thesis, but as N throughout the code. Similarly, the covariates *Z1*, *Z2*, and Q in the thesis are referred to as *baanduur*, *SBIgroep*, and q throughout the code.

2 Software requirements

For this project, the following software is required:

Software	**Version**
RStudio	2022.02.01
R	4.1.3
Latent GOLD	6.0

To perform the analyses and plot the results, the following R packages are required:

Package	**Version**
data.table	1.14.8
dplyr	1.1.2
ggplot2	3.4.2
reshape2	1.4.4

Continues on next page.

RColorBrewer	1.1-3
scales	1.2.1
tidyverse	2.0.0
stringr	1.5.0

3 Instructions to reproduce and plot the results

3.1 Simulation study 1 (Chapter 4)

3.1.1 Instructions to reproduce the simulation study:

1. Load the functions in the following files into R:
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Simulations.R,
   - Functions/Methods_Simulation_Studies.R,
   - Functions/Helpfunctions_Simulate_Data_1.R,
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R,
2. Open the file Analyses/Perform_Simulation_Study_1.R.
3. Change the working directory in line 23.
4. Change the argument *folder* in lines 67, 72, and 77 to where you would like to store the model results. Make sure to end the folder name with a/.
5. Execute the rest of the code. Note that this takes approximately 5 days.

3.1.2 Instructions to plot the results (except for the heatmaps):

1. Follow either step 1a **or** 1b:
   a. Run the code in the file Plots/Prepare_Simulation_Results_for_Plotting.R to prepare the data for plotting. Note that this part requires that the simulation in the file Analyses/Perform_Simulation_Study_1.R has been executed.
   b. Load the file RData/Simulation_1_Reduced.RData into R.
2. Run the code in the file Plots/Plot_Simulation_1.R to create the plots.

3.1.3 Instructions to plot the heatmaps:

1. (*If not already loaded*): Load the functions in the following files into R:
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Simulations.R,
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R,
2. (*If the simulation in the file Perform_Simulation_Study_1.R has not been executed*): Load the file Simulation_1_Complete.RData in R.
3. Specify a working directory in line 37 to store the plots in.
4. Run the code in the file Plot_Simulation_1_Heatmaps.R to create the plots.

3.2 Simulation study 2 (Chapter 5)

3.2.1 Instructions to reproduce the simulation study:

1. Load the functions in the following files into R:
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Simulations.R,
   - Functions/Methods_Simulation_Studies.R,
   - Functions/Helpfunctions_Simulate_Data_2.R,
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R,
2. Open the file Analyses/Perform_Simulation_Study_2.R.
3. Change the working directory in line 32.
4. Make sure the files Analyses/exampleDat_1000.dat and Analyses/exampleDat_10000.dat are in the working directory.
5. Change the argument *folder* in lines 83, 87, and 91 to where you would like to store the model results. Make sure to end the folder name with a/.
6. Execute the rest of the code. Note that this takes approximately 5 days.

3.2.2 Instructions to plot the results (except for the heatmaps):

1. Follow either step 1a **or** 1b:
   a. Run the code in the file Plots/Prepare_Simulation_Results_for_Plotting.R to prepare the data for plotting. Note that this part requires that the simulation in the file Analyses/Perform_Simulation_Study_2.R has been executed.
   b. Load the file RData/Simulation_2_Reduced.RData into R.
2. Run the code in the file Plots/Plot_Simulation_2.R to create the plots.

3.2.3 Instructions to plot the heatmaps:

1. (*If not already loaded*): Load the functions in the following files:
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Simulations.R,
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R,
2. (*If the simulation in the file Analyses/ Perform_Simulation_Study_2.R has not been executed*): Load RData/Simulation_2_Complete.RData.
3. Specify a working directory in line 37 to store the plots in.
4. Run the code in the file Plots/Plot_Simulation_2_Heatmaps.R to create the plots.

3.3 Initial analysis using the first approach in Section 5.1.1

3.3.1 Instructions to reproduce the analysis:

1. Load the functions in the following files into R:
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Simulations.R,
   - Functions/Methods_Initial_Analysis_Approach_1.R,
   - Functions/Helpfunctions_Simulate_Data_2.R,
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R,
2. Open the file Analyses/Perform_Initial_Analysis_Approach_1.R.
3. Change the working directory in line 23.
4. Change the argument *folder* in lines 26 and 27 to where you would like to store the model results. Make sure to end the folder name with a/.
5. Execute the rest of the code.

3.4 Initial analysis using second approach in Section 5.1.2

3.4.1 Instructions to reproduce the analysis:

Load the functions in the following files:

- Functions/Helpfunctions_General.R,
- Functions/Helpfunctions_Simulations.R,

- Functions/Methods_Initial_Analysis_Approach_2.R,
- Functions/Helpfunctions_Simulate_Data_2.R,
- Functions/Helpfunctions_Performance_Measures_and_Plots.R,

1. Open the file Analyses/Perform_Simulation_Study_2.R.
2. Change the working directory in line 32.
3. Make sure the files Analyses/exampleDat_1000.dat and Analyses/exampleDat_10000.dat are in the working directory.
4. Change the argument *folder* in lines 83, 87, and 91 to where you would like to store the model results. Make sure to end the folder name with a/.
5. Change lines 40 and 42 to: `N  <- 1000, iteration <- 1:25`
6. Execute the rest of the code. Note that this takes approximately 12 hours.

3.4.2 Instructions to plot the results (except for the heatmaps):

1. Follow either step 1a **or** 1b:
   a. Run the code in the file Plots/Prepare_Simulation_Results_for_Plotting.R to prepare the data for plotting. Note that this part requires that the simulation in the file Analyses/Perform_Simulation_Study_2.R has been executed as described above.
   b. Load the file RData/Initial_Analysis_Approach_2_Reduced.RData.
2. Run the code in the file Plots/Plot_Initial_Analyis_Approach_2.R to create the plots.

3.5 Analyses on real data from the ER and the LFS (Chapter 6)

Note that the data from the ER and the LFS is not publicly available.

3.5.1 Instructions to reproduce the analyses:

1. Load the functions in the files
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Real_Data.R,
   - Functions/Methods_Real_Data.R,
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R into R.
2. Open the file Analyses/Perform_Real_Data_Analyses.R.

3. Change the working directory in line 19.

4. Change the argument *folder* in lines 82-103 to where you would like to store the model results. Make sure to end the folder name with a/.

5. Execute the rest of the code. Note that this may take a few hours.

3.5.2 Instructions to obtain and plot the results:

1. *(If not already loaded)*: Load the functions in the following files:
   - Functions/Helpfunctions_General.R,
   - Functions/Helpfunctions_Real_Data.R, and
   - Functions/Helpfunctions_Performance_Measures_and_Plots.R

2. (*If the models in the file Analyses/Perform_Real_Data_Analyses.R were not created*): Load the content of the file RData/Real_Data_Analyses.RData in R. Note that this file is not publicly available.

3. Run the code in the file Plots/Plot_Real_Data_Analyses.R to obtain the results and to create the plots.