Indications for the replication of the results in the paper **Personal Bests and Gender**

Julio González-Díaz[1], Ignacio Palacios-Huerta[2], and José M. Abuín[3]

[1]Universidad de Santiago de Compostela. Email: julio.gonzalez@usc.es.

[2]London School of Economics and Ikerbasque UPV/EHU. Email: i.palacioshuerta@gmail.com.

[3]Universidad de Santiago de Compostela. Email: josemanuel.abuin@usc.es

November 2021

1 Testing environment

All the analyses reported in the paper have been performed using language Language R (v. 4.1.1) through RStudio (v. 1.4.1717) environment. The pdfs in the online appendices have been automatically generated with knitr and the attached scripts and data files are meant to be self contained for replication purposes.

2 Data files

The compressed file with the supplementary data and scripts contains three main data folders, which contain the corresponding RData files with the relevant data:

- **RawData.** This folder contains the raw data as imported from FIDE website into R dataframes. The full dataset corresponds with files of the form:

  "`Data-CDB-289238-freq_1_batch_n-60.RData`".

  Additionally, some smaller test set is contained in the files of the form:

  "`Data-999-freq_1_batch_n-60.RData`".

- **ProcessedData.** This folder contains processed data, obtained from the raw data in order to run the analyses reported in the paper. These files are not necessary to replicate the results, since the attached code generates them automatically if needed.

- **MatchingEstimatorsData.** This folder contains additional processed data, specific for the matching estimator's analysis reported in Appendix B. Again, this files are not necessary to replicate the results, since the attached code generates them automatically if needed.

3 Master files for replication purposes

The supplementary data and scripts uploaded to the online repository of the Review of Economics and Statistics have been prepared to allow for a direct replication of the results presented in the paper and in the online appendices. Moreover, an easy inspection of the master files allows to change different parameters in order to perform additional tests. It is worth noting that the generation of the appendices is quite demanding computationally and make take several hours (some intermediate processed data files have been also uploaded to save some computing time when the very same configurations used in online Appendices A and B are run).

There are three main files, which are described below:

- `++GRD_Appendix_A.Rnw`: This is the master file to generate online Appendix A.
- `++GRD_Appendix_B.Rnw`: This is the master file to generate online Appendix B.
- `++GRD_Main_Results_Reported_in_Paper.Rnw`: This is the master file to generate a pdf containing only the main regressions reported in the paper (corresponds to Chapter 2 in online Appendix A).