Replication Code and Data for
"A Check for Rational Inattention"
by Greg Howard
Journal of Political Economy Microeconomics

The data source for this project is the lichess.org open database
which can be downloaded from this url: https://database.lichess.org/
I use the April 2017 data, which I downloaded on 9/10/2020.
You can find that copy of the data in the rawdata folder.

In this folder, you will also find the code to replicate the data.

The first step after downloading the files is to unzip the file
in the rawdata folder which is split into two files due to
size limits on the JPE Micro Repository. This can be done
using 7-Zip or similar. Alternatively, the database
can be downloaded directly from lichess and put in that folder.
After that has been unzipped, you can continue with the following steps:
If you open the master.do and set the directory to the current folder,
the code will then run, including the cleaning and the analysis code.
The entire code runs in about 2 weeks on a PC with 16GB of RAM.
It was run using Stata 17.0 and Python 3.9.

If you wish to skip the data cleaning aspects of the code,
comment out everything between the "Data Cleaning" and the "Data
Analysis" lines in the master.do file. This should run in about 1.5
hours on a similar machine, and does not require python.

Program files are stored in dofiles, which contain Stata files, and
pythonfiles, which contain python files. The data cleaning files,
described below, save data in the intermediatedata folder, and the
cleandata folder. The analysis files output pdf's and LaTeX tables (.tex)
into the exhibits folder.

Currently, the intermediatedata folder contains only two files.
These are used by motivationtable.do. Cleaning the data
will create other files for this folder, too, which I have
not uploaded since many of them are too large for the
repository.

Here is a list of programs that master.do calls, and what they do.

pythonfiles
-import_lichess
        This reads in the rawdata, which is in a pgn format (which is
        chess-specific) and saves the relevant information into
        lichess.csv in intermediatedata

dofiles

-clean_lichess.do
	This reads lichess.csv and saves it to lichess.dta, as well as subfiles lichess15.dta to lichess600.dta, where the number corresponds to the number of seconds in the time control.

-comparison_groups_and_win_prob_main.do (calls comparison_groups_and_win_prob.do)
	This takes the intermediatedata, and estimates a value function in each time control as described in the paper. It then matches observations to the next faster and slower time controls, and saves them as cleandata/winprob_matched_datasetT1_T2.dta
	where T1 is the time control and T2 is the matched time control.

-comparison_groups_and_win_prob_main_100bins.do (calls comparison_groups_and_win_prob_100bins.do) This does the same thing, but uses a slightly different algorithm that uses 100 bins. The output names are appended with "_100bins"

-comparison_groups_and_win_prob_main_lassosplines.do (calls comparison_groups_and_win_prob_lassosplines.do)
	This does the same thing, but uses a slightly different algorithm that uses LASSO splines. The output names are appended with "_lassospline"

-summarystats.do
	Calculates summary statistics and creates Table 1 (summarystats.tex).

-summarystats_matchedsample.do
	Calculates summary statistics only for the matched observations. Creates Tables C.1 and C.2 (summarystats_lowermatch.tex and summarystats_uppermatch.tex)

-motivation_table.do
	Creates Table A.1 (motivation.tex)

-test_fit.do
	Creates Table B.1 (value_function_table.tex) and Figures B.1 and B.2 (winprob_fit.pdf, eval_fit.pdf, clock_fit.pdf, opp_clock_fit.pdf, winprob_hist.pdf, and marginalvalueoftime_hist.pdf)

-comparing_marginal_benefit_to_marginal_cost.do
	Creates Figure 1 (marginal_benefit_cost_upperbound.pdf and marginal_benefit_and_cost_lowerbound.pdf) and Tables C.3, C.4, C.5, and C.6

	(firststage_lower.tex, lowerbound.tex, firststage_upper.tex, upperbound.tex)

-comparison_by_elo.do
	Creates Figure 2 (elo_comparison.pdf) and elo_pvalues.csv, which is used for p-values in the text.

-comparison_by_startclock.do

Creates Figure 3 (startclock_comparison.pdf) and start_clock_pvalues.csv, which is used for p-values in the text.

-comparison_by_elo_robust.do
Creates Figure C.7 (elo_comparison_robust.pdf)

-comparison_by_startclock_robust.do
Creates Figure C.8 (start_clock_comparison_robust.pdf)

-comparison_by_elo_100bins.do
Creates Figure C.3 (elo_comparison_100bins.pdf)

-comparison_by_startclock_100bins.do
Creates Figure C.4 (start_clock_comparison_100bins.pdf)

-comparison_by_elo_lassosplines.do
Creates Figure C.5 (elo_comparison_lassosplines.pdf)

-comparison_by_startclock_lassosplines.do
Creates Figure C.6 (start_clock_lassosplines.pdf)

-beep_analysis.do
Creates Figure 4a, 4b, 4c (hist_beep60.pdf, time_beep60.pdf, and eval_beep60.pdf) and Figure C.1a, C.1b, C.1c (hist_beep30.pdf, time_beep30.pdf, and eval_beep30.pdf)

-beep_analysis_by_elo.do
Creates Figure 4d and 4e (beep_analysis_by_elo751_1357.pdf and beep_analysis_by_elo1823_2890.pdf)
and Figure C.2 (beep_analysis_by_elo.pdf)

-beep_analysis_by_elo30seconds.do
Creates Figure C.1d and C.1e (beep_analysis_by_elo918_1613_30sec.pdf and beep_analyis_by_elo2018_2930_30sec.pdf)

-difficulty_balance.do
Creates Table C.7 (balance_captures.tex, balance_checks.tex, and balance_pieces.tex)