An Experimental Analysis of the Prize–Probability Tradeoff in Stopping Problems

Replication Guide

Yair Antler and Ayala Arad (Tel Aviv University)

1. Introduction

This guide describes how to replicate the maximum likelihood estimation (MLE) and leave-one-out prediction exercise (LOO) of alternate theories, section 4.1 of the paper:

Antler Y., Arad A. (2023): "An Experimental Analysis of the Prize–Probability Tradeoff in Stopping Problems", *The Review of Economics and Statistics*, forthcoming.

If you use any of the data or the code described in this paper in your own work, even in altered form, please do not forget to properly cite the above paper. Thank you.

Please note that the results of the maximum likelihood estimation may be sensitive to the chosen starting parameters. Hence, minor differences between the replicated results and the reported results may occur. Importantly however, the number of types reported in the paper is not sensitive to the chosen starting parameters.

2. Estimation Methodology

In order to estimate the utility functions of individuals we utilize a random utility model (RUM) and use the maximum likelihood estimation. RUM assumes that expected utility is a function of utility from the selected choice and a random error component.

An individual's utility can be expressed as:

$$U_j = \beta' X_j + \varepsilon_{j,}$$

Where X_j is a column vector of observed attributes of alternative j and the individual, β is a conformable vector of constant parameters, and $\varepsilon_{j,}$ is a random variable that accounts for the effects on preferences of unobserved attributes of the alternative and individual.

The probability that an individual chooses alternative *i* is:

$$P(i|X) = P(\beta' X_i + \varepsilon_i > \beta' X_j + \varepsilon_j) \quad for\ all\ j = 1, \ldots, J;\ j \neq i$$

Where X = (X_1',……,X_j')'

Assuming the error follows i.i.d and type 1 extreme value distribution the analytical solution to *P(i|X)* is given by the multinomial logit (MNL) model:

$$P(i|X) = \frac{exp(\beta' X_i)}{\sum_{j=1}^{J} exp(\beta' X_j)}$$

This model is estimated using the maximum likelihood estimation.

3. Requirements

To run the replication, you need:

- An installation of the R Software Environment which can be obtained for free from the website https://www.r-project.org/ for various platforms (including Windows, Mac OS, and Linux).

  The replication was tested with version 4.2.2 of R on Windows 11 (OS build 22621.1992). However, it should also work with any newer version of R and on a different platform.

- A text editor to view and edit the code, we recommend RStudio which can be obtained for free from the website https://posit.co/download/rstudio-desktop/ for various platforms (including Windows, Mac OS, and Linux).

  The replication was tested with version 2023.06.1+524 of RStudio on Windows 11 (OS build 22621.1992). However, it should also work with any newer version of R and on a different platform.

- Additional RStudio packages which can be installed and loaded using the code provided within the scripts. The version numbers of the packages used are:
  - tidyverse: 2.0.0
  - lubridate: 1.9.2
  - rio: 0.5.29
  - bbmle: 1.0.25

4. Description of Files

The '*stopping_rules_replication_folder'* will serve as the working directory and has 3 folders:

1. *Data* (file format – comma separated value file or .csv):
   - *stopping_rules_data*: Main dataset used for all estimations.
   - *regret_numbers*: Secondary data set used only for the estimations of regret aversion.
   - *all_data:* In this file one can find the raw data from Parts A, B, C, and D of the experiment.
2. *Replication_scripts* (file format – R script or .R):

- *stopping_rules_leave_one_out_script*: Main replication script. Make sure to set the replication folder as your working directory. The script estimates the leave-one-out prediction exercise for each utility function.
- *stopping_rules_LOO_random_data_script*: This script runs the "placebo test" on page 27 (footnote 15). Runs the LOO prediction exercise on randomly generated data.
- *stopping_rules_individual_estimation_script*: This script calculates individual student parameter estimations for each of the respective utility functions.
- *stopping_rules_func_library (Sub Folder)*: A sub folder that includes an R script for each of the alternate theories which defines all necessary functions for the leave-one-out prediction exercise. It also includes the "MaxLike_func_lib_script" which is used to simplify the loading of all necessary functions and defines all necessary functions for the individual student parameter estimations.

3. *Analysis_for_tables* (file format – Excel file or .xlsx)
   - *analysis_for_tables:* In this Excel file it is possible to find the estimation results generated by the file above and their analysis. The tables that describe the estimation can be reproduced by this file.

5. Replicating the Results

To replicate the main results of the paper, please proceed as follows:

1. Download and install R and RStudio.
2. Open a replication script (for example *stopping_rules_leave_one_out_script*)
3. The working directory must be the '*stopping_rules_replication_folder'*. On line 10 of the scripts set the path name for the folder in the place holder name "YOUR_FOLDER_PATH":

   setwd("YOUR_FOLDER_PATH/stopping_rules_replication_folder")

   Note: For windows users make sure to use forward slash (/) and not backslash (\).

4. If you have not installed the required packages, you can uncomment lines 13-16 by highlighting the lines and pressing *ctrl* + *shift* + *C* (windows users) or deleting the hashtags from each line:

   # install.packages("tidyverse")
   # install.packages("lubridate")
   # install.packages("rio")
   # install.packages("bbmle")

   Run the lines by pressing *ctrl* + *enter*.

5. You are now ready to run the script.
   a. You can run the entirety of the script by highlighting the whole script (manually or press *ctrl* + *A*) and pressing *ctrl* + *enter* or just pressing *ctrl* + *shift* + *enter*.

b. You can run the script line by line manually by highlighting a particular section and pressing *ctrl + enter*. This is useful if you want to estimate only a particular utility function.

Note: The *stopping_rules_leave_one_out_script* that estimates the main leave one out prediction exercise can take a few days to run due to the regret aversion utility function estimation which can take up to 72 hours to run. All other utility functions take anywhere from 5 minutes to a couple hours to run. If you are interested in specific utility functions, we recommend estimating them separately instead of the whole script.

6. Data Dictionary

stopping_rules_data:

studentID	Student experiment ID number
qnum	Question number
choice	Student choice out of five stopping rules: 1=ll, 2=l, 3=s,4=r, 5=rr
lower.1	The lower bound of stopping rule 1
lower.2	The lower bound of stopping rule 2
lower.3	The lower bound of stopping rule 3
lower.4	The lower bound of stopping rule 4
lower.5	The lower bound of stopping rule 5
upper.1	The upper bound of stopping rule 1
upper.2	The upper bound of stopping rule 2
upper.3	The upper bound of stopping rule 3
upper.4	The upper bound of stopping rule 4
upper.5	The upper bound of stopping rule 5
prob.1	The probability of winning in stopping rule 1
prob.2	The probability of winning in stopping rule 2
prob.3	The probability of winning in stopping rule 3
prob.4	The probability of winning in stopping rule 4
prob.5	The probability of winning in stopping rule 5
treatment	treatment=1 is the main treatment T_0 (induced winning probabilities not provided) and treatment=2 is the secondary treatment T_p (winning probabilities provided)

regret_numbers (probability of regret):

qnum	Question number
choiceJ	Choice of stopping rule j
q_{1j}^h	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 1 and winning
q_{2j}^h	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 2 and winning
q_{3j}^h	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 3 and winning
q_{4j}^h	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 4 and winning
q_{5j}^h	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 5 and winning

q_{1j}^l	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 1 and losing
q_{2j}^l	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 2 and losing
q_{3j}^l	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 3 and losing
q_{4j}^l	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 4 and losing
q_{5j}^l	The probability of finishing the game with a gain had the individual chosen rule *j* conditional on choosing rule 5 and losing