Understanding Managers' Trade-offs between Exploration and Exploitation — Replication notebook

Alina Ferecatu and Arnaud De Bruyn

April 2020

Contents

Overview

This replication notebook accompanies the paper **Understanding Managers' Trade-offs between Exploration and Exploitation**, forthcoming in *Marketing Science*. We outline how to estimate the EEI/EWA model and the benchmarks proposed in the paper, and how to replicate the results reported.

- The most recent version of this package can be found at: https://alinaferecatu.com.

The analyses for this project were conducted on 1. a Macbook Pro with an Intel Core i7 processor, speed 3.5 GHz, and memory 16 GB RAM, and 2. a Linux virtual machine running ubuntu 20.04, with 10 cores and memory 128 GB, hosted by Surfsara (https://userinfo.surfsara.nl/). The estimation and out-of-sample prediction codes run in approximately 50 hours per model. We suggest you run them separately, save the output as instructed in each R file, and import these outputs to compute summary measures.

This README-file contains step-by-step instructions for running the analyses in the R files. For each analysis, we describe the following:

- **Data**: The input data and variable names,

- **Output**: Describes the output files and variables,

- **Code Structure and Instructions**: How to run the codes.

This README-file assumes familiarity with the paper.

Structure of each R file

- **Prerequisites**: The necessary packages are listed in the beginning of each file.

- **Set paths**: Set local paths necessary to load the data and the save the outputs. The paths are PATH_FUNCTION, PATH_DATA, PATH_RESULTS, PATH_PLOTS.

- **Source functions**: source the R file *EEtradeoffs_functions.R*, that includes functions relevant for all analyses. The file containing the functions is loaded in each R file.

- **Data**: Loads the input data for each analysis. We describe below the data and variable names per R file.

- **Output**: Save the output for each analysis. We describe below the output of each R file.

Source functions used in all R files

R code *EEtradeoffs_functions.R*

Data

None

Output

None

Code Structure and Instructions

- Code structure:
  - Basic functions.
  - Functions used for the computation of the optimal path
  - Estimation functions written in *stan* for the EEI/EWA model and the benchmark models. There are two versions of each code, a version that contains generated quantities, and a simplified version labeled '_oos' that does not contain generated quantities. The latter is used for out-of-sample predictions. Generated quantities are not needed there, and would just increase the size of the stanfit object.
  - Model comparison functions, with in-sample fit functions, followed by the functions built for out-of-sample predictions.
  - The out-of-sample predictions are organized by model, followed by type of out-of-sample predictions (temporal forecasting and cross-sectional validation).
  - Functions for posterior predictive checks.
  - Function that helps report the parameter estimates.
- Instructions:
  - Source the functions file in every R file used for replication.

Summary statistics for all studies

R code *Summary_stats_all_studies.R*

The r code *Summary_stats_all_studies.R* computes summary statistics for Studies 1, 2, and 3. Comments in the file point to where the measures are reported in the paper.

Data

Six datasets are used in this file:

1. *businessTask_Study1.csv*, which records choices of subjects in Study 1. It contains the following variables:
   - *userid*: The subject ID, from 1 to 89.
   - *usercode*: Unique user code
   - *rounds*: The number of the round, from 0 to 99.
   - *picks*: Subjects' choice of arm: arm 1 (A), 2 (B), or 3 (C).
   - *results*: the reward received from choosing the arm at each round.
   - *time_spent*: time spent per round.
   - *Set*: the experimental set of draws, with levels from set1 to set5.
2. *psychVars_Study1.csv*, which records the psychometrics for each subject in Study 1. It contains the following variables:
   - *user_id*: Unique user code, matches the usercode variable in *businessTask_Study1.csv*.
   - *bomb_test*: number of boxes collected by each subject in the risk aversion task.
   - *crra*: constant relative risk aversion coefficient, computed based on the bomb test results.
   - *SSR*: Subject-level average score of the situation-specific rational sub-scale.
   - *SSE*: Subject-level average score of the situation-specific experiential sub-scale.
   - *maximiz*: Subject-level average score of the maximization scale.
   - *Age*: Subjects' age.
   - *Gender*: Subjects' gender.
   - *rewardRA*: reward earned for the RA task.
   - *gameReward*: reward earned from the business simulation task, in euros.
3. *roundsData_xtreme2.csv*, which records choices of subjects in Study 2. It contains the following variables:
   - *Business_sim*: Factor with 2 levels: business_sim_1 - first business simulation, business_sim_2 - second business simulation.
   - *usercode*: Unique user code.
   - *Rounds*: The number of the round, from 1 to 100.
   - *picks*: Subjects' choice of arm: arm 1 (A), 2 (B), or 3 (C).
   - *results*: the reward received from choosing the arm at each round.
   - *time_spent*: time spent per round.
4. *psychVars_xtreme2.csv*, which records the psychometrics for each subject in Study 2. It contains the following variables:
   - *usercode*: Unique user code, matches the usercode variable in *roundsData_xtreme2.csv*.
   - *bomb_test*: number of boxes collected by each subject in the risk aversion task.
   - *crra*: constant relative risk aversion coefficient computed based on the bomb test.
   - *SSR*: Subject-level average score of the situation-specific rational sub-scale.
   - *SSE*: Subject-level average score of the situation-specific experiential sub-scale.
   - *maximiz*: Subject-level average score of the maximization scale.
   - *gameReward*: Reward earned from the business simulation, in pounds.
5. *roundsData_xtreme3.csv*, which records choices of subjects in Study 2. It contains the following variables:
   - *Business_sim*: Factor with 2 levels: business_sim_1 - first business simulation, business_sim_2 - second business simulation.
   - *usercode*: Unique user code.
   - *Rounds*: The number of the round, from 1 to 100.
   - *picks*: Total number of purchases in the data
   - *results*: The reward received from choosing the arm at each round.
   - *time_spent*: Time spent per round. Note that there were six data points with errors recording time spent (negative). We substitute them with mean time spent.
6. *psychVars_xtreme3.csv*, which records the psychometrics for each subject in Study 2. It contains the

following variables:

- *usercode*: Unique user code, matches the usercode variable in *roundsData_xtreme3.csv*.
- *bomb_test*: Number of boxes collected by each subject in the risk aversion task.
- *crra*: constant relative risk aversion coefficient, computed based on the bomb test.
- *SSR*: Subject-level average score of the situation-specific rational sub-scale.
- *SSE*: Subject-level average score of the situation-specific experiential sub-scale.
- *maximiz*: Subject-level average score of the maximization scale.
- *gameReward*: Reward earned from the business simulation, in pounds.

Output

The following output types:

1. The summary statistics reported in Section 3.2 (Study 1), and Web Appendices WA1.2.2 (Study 2), and WA1.3.2 (Study 3).

2. Figures displayed in Section 3.2 (Study 1).

3. The last part of the code shows the EWA plots following the reported in Web Appendix WA5.

Code Structure and Instructions

- Code structure:
  - Summary stats for study 1.
  - Summary stats for Study 2.
  - Summary stats for study 3.
  - Plots of the EWA example (WA5).
- Instructions:
  - None.

Estimation of the EEI/EWA model using Study 1 data set

R code *Study1_analysis.R*

The r code estimates the EEI/EWA model using the data set from Study 1. It reports the model parameters, saves the output, performs various analyses based on estimation results, and reports in-sample fit statistics and posterior predictive checks.

Data

Two datasets are used in this file:

1. *businessTask_Study1.csv*, which records choices of subjects in Study 1 (see variable description in section **Summary statistics for all studies**).

2. *psychVars_Study1.csv*, which records the psychometrics for each subject in Study 1 (see variable description in section **Summary statistics for all studies**).

Output

The following output types are reported:

1. A *stanfit* object, which contains the estimation results of the EEI/EWA model. We used two HMC chains, with 4000 iterations per chain. The first 3000 iterations are used for warmup. The two chains run in parallel. The estimation takes approx. 50 hours to run. The object is approx. 9GB.

2. Figures displayed in Section 6.1 and 6.2.

3. Stargazer tables, reported in sections 6.1. and 6.2.

4. A *.csv* file, with intermediary results necessary for posterior predictive checks. These results include switching behavior per subjects.

Code Structure and Instructions

- Code structure:
    - Estimation of EEI/EWA model.
    - Estimation results of EEI/EWA model: parameter estimates, transition matrices, heterogeneity plots.
    - Posterior predictive checks.
    - In-sample fit statistics.
- Instructions:
    - The code estimates the proposed EEI/EWA model and reports the results for this model.
    - To replicate the model comparison results, first estimate all models discussed in Section 6.2. Save the *stanfit* objects (lines 140-141). Save the *.csv* files containing the posterior predictive checks (lines 718-733). Import all *.csv* files and compute posterior predictive checks (lines 736 - 813).
    - Use the model specific stanfit obejcts for in-sample fit statistics (lines 815-823).

Estimation of the EEI/EWA model using Study 2 data set

R code *Study2_analysis.R*

The r code estimates the EEI/EWA model using the data set from Study 2. It saves the output for each business simulation, Bsim1 and Bsim2. The estimation for Bsim2 is based on informative priors, that result from the estimation of the EEI/EWA model using the parameter estimates from the first business simulation.

Data

Seven datasets are used in this file:

1. *rounds_xtreme2.csv*, which records choices of subjects in Study 2 (see variable description in section **Summary statistics for all studies**).

2. *psychVars_xtreme2.csv*, which records the psychometrics for each subject in Study 2 (see variable description in section **Summary statistics for all studies**).

3. *gamma_prior_mean.csv*, mean of the normal prior for the (untransformed) group-level parameters for Bsim2.

4. *gamma_prior_sd.csv*, standard deviations of the normal prior for the (untransformed) group-level parameters for Bsim2.

5. *betaD_prior_mean.csv*, means of the normal prior for the aggregate-level parameters showing the impact of disappointing outcomes, used for Bsim2.

6. *betaD_prior_sd.csv*, standard deviations of the normal prior for the aggregate-level parameters showing the impact of disappointing outcomes, used for Bsim2.

7. *lambda_prior* vector with mean and standard deviation of the normal prior of the (untransformed) lambda_explore parameter, used for Bsim2.

Output

The following output types are reported:

1. Two *stanfit* objects, which contain the estimation results of the EEI/EWA model for each business simulation (Bsim1 and Bsim2). We used two HMC chains, with 4000 iterations per chain. The first 3000 iterations are used for warmup. The two chains run in parallel. The estimation takes approx. 50 hours to run. Each object is approx. 9GB.

Code Structure and Instructions

- Code structure:
  - Estimation of EEI/EWA model for Bsim1
  - Estimation of EEI/EWA model for Bsim2
- Instructions:
  - Change the MODEL flag between "eei_ewa_mixed" for Bsim1 and "eei_ewa_mixed_infPrior" for Bsim2.

  - Change the DATA flag between Bsim1 and Bsim2.
  - The model estimation will take the appropriate inputs, including the stan code, the data and priors (informative or not), depending on the MODEL and DATA flags.

Estimation of the EEI/EWA model using Study 3 data set

R code *Study3_analysis.R*

The r code estimates the EEI/EWA model using the data set from Study 3. It saves the output for each condition (R20, R50, R100, and R200).

Data

Two datasets are used in this file:

1. *rounds_xtreme3.csv*, which records choices of subjects in Study 3, across the four conditions (see variable description in section **Summary statistics for all studies**).

2. *psychVars_xtreme3.csv*, which records the psychometrics for each subject in Study 3 (see variable description in section **Summary statistics for all studies**).

Output

The following output types are reported:

1. Four *stanfit* objects, which contain the estimation results of the EEI/EWA model for each condition (R20, R50, R100, and R200).

Code Structure and Instructions

- Code structure:
  - Import Study 3 data.
  - Create condition-level dataset, filtering the business simulation and the psychometrics for each condition (lines 56-69).
  - Run estimation per condition.
- Instructions:
  - Change conditions to 20, 50, 100, or 200 rounds at line 58.
  - Change the name of stanfit object at line 111, and the name of the saved object at line 117.
  - For condition R20, run the code for 8000 iterations, with 6000 iterations as warmup. For the R50 condition, run the code for 6000 iterations, with 4000 iterations as warmup. For the R100 and R200 conditions, run the code for 4000 iterations, with 3000 iterations as warmup. Mind the objects' size (14GB for the R200 stanfit object) and the run time (over 80h for the R200 estimation).

Optimal path analysis

R code *Optimal_path_analysis.R*

The r code *Optimal_path_analysis.R* computes the optimal path and compares it with subjects' actual behavior for Studies 1, 2, and 3.

Data

Seven datasets are used in this file:

1. *exp_draws_study1.csv*, which records the five sets of experimental draws used in Study 1. It contains the following variables:

   - *Round*: The number of the round, from 0 to 99.
   - *A*, *B*, and *C*: The rewards for choosing option 1, 2, or 3.
   - *Set*: The experimental set, with levels from set1 to set5.

2. *nu.csv*, record the exact values of the Gittins index with unknown mean and known variance, as reported in Gittins and Jones (2011). It contains the following variables:

   - *round*: number of the round. Note that after 10 rounds,
   - *nu_95*: value of the index for known variance and a discount factor of 0.99.
   - *nu_99*: value of the index for known variance and a discount factor of 0.95.
   - *nu_sigma*: value of the index for unknown variance and a discount factor of 0.95.
   - *nu_sigma99*: value of the index for known variance and a discount factor of 0.99.

3. *businessTask_Study1.csv*, *roundsData_xtreme2.csv*, *roundsData_xtreme3.csv*, *psychVars_xtreme3.csv*, recording subjects' choices in the business simulation in Studies 1, 2, and 3, and the psychometrics in

Study 3. The latter is only used for subsetting (see variable description in section **Summary statistics for all studies**).

4. *exp_draws_study3.csv*, which records the five sets of experimental draws used in Study 1. It contains the following variables:

   - *Round*: The number of the round, from 0 to 99.
   - *A*, *B*, and *C*: The rewards for choosing option 1, 2, or 3.
   - *Set*: The experimental sets of draws for each condition, with levels from R20 to R200. For the conditions R20 to R100, those are based on the experimental set 2 used in Study 1. For R200, the first 100 experimental draws are the ones in set2 used in Study 1. We used a new set of draws for rounds 101 to 200 in the R200 condition.

Output

The following output types:

1. Plot of comparison of approximate vs. exact Gittins indices reported in Web Appendix 3.

2. Optimal path computation and comparison between optimal path and actual behavior in Study 1, reported in Section 4 (including plots and summary statistics).

3. Optimal path computation and comparison between optimal path and actual behavior in Study 2, reported in Web Appendix 1.2.3 (summary statistics).

4. Optimal path computation and comparison between optimal path and actual behavior in Study 3, reported in Web Appendix 1.3.3 (summary statistics).

Code Structure and Instructions

- Code structure:
  - Bayesian updating for one set of draws
  - Comparison of approximate vs. exact Gittins indices.
  - Optimal path computation, followed by comparison between optimal path and actual behavior in Study 1.
  - Optimal path computation, followed by comparison between optimal path and actual behavior in Study 2.
  - Optimal path computation, followed by comparison between optimal path and actual behavior in Study 3.
- Instructions:
  - The experimental draws used in Study 2 are the second and the fourth (set2, set4) sets of draws used in Study 1. See line 300.

Out-of-sample predictions - temporal forecasting

R code *OOS_predictions_temporal_forecasting.R*

The r code *OOS_predictions_temporal_forecasting.R* computes the out-of-sample predictions for temporal forcasting for Study 1. It estimates the EEI/EWA model using 80 rounds per subject, and predicts choices for the last 20 rounds. It them computes the summary measures for out-of-sample predictions, such as hit rates, mean squared errors and mean absolute errors.

Data

Two datasets are used in this file: *businessTask_Study1.csv* and *psychVars_Study1.csv*, which records choices and psychometrics of subjects in Study 1 (see variable description in section **Summary statistics for all studies**).

Output

The following output types:

1. *stanfit* object with the estimation results of the EEI/EWA model based on 80 rounds of data.

2. *.csv* file with the summary statistics for the out-of-sample hit rates and mean squared errors.

3. *.csv* file with a data recording switching behavior at the subject level, necessary to compute mean absolute errors.

4. Plots of observed vs. predicted switching behavior, reported in Section 6.2 and in Web Appendix WA4.

Code Structure and Instructions

- Code structure:
    - Estimation of EEI/EWA model for 80 rounds.
    - Out-of-sample predictions for the EEI/EWA model
    - Out-of-sample switching behavior for the EEI/EWA model
    - Plots of observed vs. predicted switching behavior across all models.
- Instructions:
    - The code estimates the proposed EEI/EWA model and reports the results for this model.
    - To replicate the model comparison results, first estimate all models discussed in Section 6.2 out-of-sample. You can do so by changing the MODEL and INDDIF flags. Save the *stanfit* objects (lines 133-134). Run the out-of-sample predictions function and save the *.csv* files containing the hit rates and mean squared errors (lines 156-157). Change the out-of-sample predictions function at line 142 according to the model (see R code *EEtradeoffs_functions.R*, section *Out-of-sample predictions*). Import all *.csv* files and compute summary statistics across all models (lines 166 - 170).
    - Based on the same out-of-sample predictions function, save the *.csv* files containing the predicted switches at the subject level (lines 192-193). Import the model-specific *.csv* files and compute mean absolute errors across all models (lines 232-237), and plot predicted switches (lines 244 - 297).

Out-of-sample predictions - cross-sectional validation

R code *OOS_predictions_cross_sectional_validation.R*

The r code *OOS_predictions_cross_sectional_validation.R* computes the out-of-sample predictions for cross-sectional validation for Study 1. It uses a K-folds cross-validation approach, with 10 folds. It estimates in parallel the EEI/EWA model using data from 80 (81 in fold 10) subjects, and predicts out of sample the full history of choices for the remaining 9 (8) subjects. It them computes the summary measures for out-of-sample predictions, such as hit rates, mean squared errors and mean absolute errors.

Data

Two datasets are used in this file: *businessTask_Study1.csv* and *psychVars_Study1.csv*, which record choices and psychometrics of subjects in Study 1 (see variable description in section **Summary statistics for all studies**).

Output

The following output types:

1. A large *list of stanfit objects*, with each *stanfit* object in the list containing the estimation results of the EEI/EWA model for each fold.

2. *.csv* file with the summary statistics for the out-of-sample hit rates and mean squared errors.

3. *.csv* file with a data recording switching behavior at the subject level, necessary to compute mean absolute errors.

4. Plots of observed vs. predicted switching behavior, reported in Section 6.2 and in Web Appendix WA4.

Code Structure and Instructions

- Code structure:
    - Estimation of EEI/EWA model for the 10 folds.
    - Out-of-sample predictions for the EEI/EWA model
    - Out-of-sample switching behavior for the EEI/EWA model
    - Plots of observed vs. predicted switching behavior across all models.
- Instructions:
    - The code estimates the proposed EEI/EWA model and reports the results for this model. The 10-fold cross-validation runs in parallel. We ran the code on a virtual machine hosting ubuntu 20.04 with 10 cores and memory of 128GB RAM.
    - To replicate the model comparison results, first estimate all models discussed in Section 6.2 out of sample. You can do so by changing the MODEL and INDDIF flags. Save the list of *stanfit* objects (lines 133-134) containing the estimation of the data sets excluding the 10 folds. Change the out-of-sample predictions function at line 203 according to the model (see R code *EEtradeoffs_functions.R*, section *Out-of-sample predictions*). Save the *.csv* files containing the hit rates and mean squared errors per model (lines 228-229). Import all *.csv* files and compute summary statistics across all models (lines 238 - 242).
    - Based on the same out-of-sample predictions function, save the *.csv* files containing the predicted switches at the subject level (lines 192-193). Import the model-specific *.csv* files and compute mean absolute errors across all models (lines 290-294), and plot predicted switches (lines 296 - 356).

Simulation study

R code *Simulation_study.R*

The r code estimates the EEI/EWA model using simulated data. It reports the model parameters, saves the *stanfit* estimation output, performs summary statistics based on estimation results, and reports in-sample fit statistics.

Data

One dataset is used in this file:

1. *exp_draws_Study1.csv*, which records the five sets of experimental draws used in Study 1 (see variable description in section **Optimal path analysis**).

Output

The following output types are reported:

1. A *stanfit* object, which contains the estimation results of the EEI/EWA model.

2. Stargazer table for parameter estimates reported in WA6.

3. In-sample fit and state recovery statistics.

Code Structure and Instructions

- Code structure:
  - Simulate data using the EEI/EWA model.
  - Estimate the EEI/EWA model based on simulated data.
  - Report parameter estimates.
  - Report in-sample fit statistics and state recovery measures.
- Instructions:
  - The code estimates the proposed EEI/EWA model and reports the results for this model.
  - To replicate the model comparison results, first estimate all models discussed in Section 6.2. Change the MODEL flag accordingly.
  - Use the model-specific stanfit object for in-sample fit statistics and state recovery measures (lines 249-261).

Counterfactual simulation study

R code *Counterfactual_simulation_study.R*

The r code replicates the counterfactual analysis reported in Web Appendix WA8. It reports the model parameters, saves the output, performs summary statistics based on estimation results, and reports in-sample fit statistics.

Data

Four datasets are used in this file:

1. *psychVars_Study1.csv*, which records the psychometrics for each subject in Study 1 (see variable description in section **Summary statistics for all studies**).

2. *Counterfactual_analysis_heterogeneity_pars.csv*, which records the group-level parameter estimates, with intercepts and the impact of psychometrics. It contains the following variables:

   - *Pars*: The parameter estimate.
   - *Mean*: The mean of the estimate.
   - *SD*: The standard deviation of the estimate.
   - *HDI_low*: The 2.5% density bound.

- *HDI_high*: The 97.5% density bound.

3. *Counterfactual_analysis_L_Omega.csv*, which records the parameter estimates for the lower triangular of the Cholesky decomposition of the correlation matrix. It contains the following variables:

   - *Pars*: The parameter estimate.
   - *Mean*: The mean of the estimate.
   - *SD*: The standard deviation of the estimate.
   - *HDI_low*: The 2.5% density bound.
   - *HDI_high*: The 97.5% density bound.

4. *Counterfactual_analysis_tau.csv*, which records the estimates of the standard deviations of the model parameters. It contains the following variables:

   - *Pars*: The parameter estimate.
   - *Mean*: The mean of the estimate.
   - *SD*: The SD of the estimate.
   - *HDI_low*: The 2.5% density bound.
   - *HDI_high*: The 97.5% density bound.

Output

The following output types are reported:

1. Plot of the differences in simulated overall rewards and sampling strategies for different psychological profiles, reported in Figure WA10.

2. Plot of the simulated mean overall rewards and 95% confidence bounds per psychological profile, reported in Figure WA11.

3. Summary statistics reported in We Appendix WA8.1

Code Structure and Instructions

- Code structure:
  - Simulate overall rewards using the EEI/EWA model for psychological profiles differing on one psychometric only.
  - Simulate overall rewards using the EEI/EWA model for psychological profiles with all combinations of psychometrics.
- Instructions:
  - None