README

Ewout Verriest

June 2025

This is the README file for the paper titled "Parenting with Patience: Parental Incentives and Child Development" by Del Boca, Flinn, Verriest and Wiswall (2025). Any further questions or requests can be sent to Ewout Verriest via email (ewoutverriest@gmail.com or ewout@psu.edu).

1 Directories

After downloading the replication package on a hard drive, it should have the following structure of folders and subfolders, some of which may be empty prior to running the code. If not, you should create the directories yourself (see also the first few lines of the master MATLAB file "**master_DFVW.m**").

- A folder titled "**paper**".

  This folder may be empty initially, but will later contain all the tables and figures reported in the paper. All tables will be MATLAB-generated .txt files containing LaTeX code. All figures will be MATLAB-generated .png and .eps files.

- A folder titled "**matlab**".

  This contains three subfolders:

  - A folder titled "**data**".

    This folder should contain all the data files necessary to run the master MATLAB file titled "**master_DFVW.m**", including the data file "child_data_PSID.txt" which is available in the replication package.

    The other data sets used for estimation (from the Osaka Preference Parameter Study and from Prof. Laurence Steinberg, respectively) are proprietary and therefore not included in the replication package (see Section 2 below on how to request access).

    In order to allow the code to compile even without having access to these proprietary data files, the master MATLAB file "**master_DFVW.m**" contains some code that loads only the necessary summary statistics and data moments which were obtained from the two raw data sets, and which are used in the estimation. This way, the entire code can be run even without having access to either of the two proprietary data sets. We note that all of these summary statistics and data moments (from the Osaka and Steinberg data sets, respectively) are also reported in the paper itself.

For more details, we refer to the code section of the master file titled "Load and clean Osaka Preference Parameter Study data" and "Load and clean Steinberg et al. data on children's discount factors", respectively.

- A folder titled "**mfiles**".

  This folder should contain all the MATLAB .m-files which are called upon at some point by the master file (see also the overview below in Section 3).

- A folder titled "**output**".

  This output folder is initially empty, but will save some temporary data files (e.g. parameter estimates, MSM weighting matrices, bootstrap results, etc.) later on as the code runs.

- A folder titled "**osaka**".

  This folder should contain a STATA do-file called "**master_osaka_betas.do**" which loads the (proprietary) raw data from the 2010 U.S. wave of the Osaka Preference Parameter Study.

  - If you have requested access to the raw data set from the 2010 U.S. survey through the website, you should have obtained a data file titled "2010_USA.dta", and it should be stored in this "**osaka**" folder.

    Then, running the STATA do-file creates a cleaned version of the Osaka data called "osaka_2010_final.txt", and stores it in the "**data**" folder where it can be used by the MATLAB master code.

  - Conversely, if the raw data set is not available (e.g. because you did not request access to it), then the STATA .do-file will not compile. However, the MATLAB master code .m-file will still work, by loading in the required summary statistics calibrated from the raw data.

2 Data

The MATLAB master file uses three pieces of data, only one of which is publicly accessible, and two of which require permission to access (but which are not strictly required to successfully compile most of the code).

1. The first data file is called "child_data_PSID.txt", and is available in the replication package.

   This file contains the household panel data set from the merged PSID-CDS data, and was obtained after using STATA to load, transform, clean and merge various raw data files which were freely obtained from the PSID website.[1] This STATA code and the raw PSID-CDS data files are not necessary to complete the replication of the paper, but are available upon request.

2. The second data file is called "osaka_2010_final.txt", and is not available in the replication package.

   This file contains the cleaned survey data from the 2010 wave of the Osaka Preference Parameter Study. This data can be requested through this website: link, and can

[1] See https://simba.isr.umich.edu/data/data.aspx to download original PSID and CDS files.

be obtained after fulfilling the qualifications and requirements of the data application process.

Note: the 2010 U.S. survey questionnaire is accessible here: link. Our main question of interest was "Question 5", which asked participants to compare receiving either 100 dollars in one month, or X dollars in thirteen months. We refer to our Appendix for more information on how we transformed the responses to this question into a distribution of discount factors for U.S. adults.

3. The third data file is called "steinberg_data_nomiss.csv", and is not available in the replication package.

   This file contains the proprietary survey data measuring forward-lookingness and delayed gratification for over 900 children and adolescents. This data was collected by Professor Laurence Steinberg at Temple University (and his co-authors on the Steinberg et al. (2009) paper), who generously shared the data with us after an email request outlining the purpose of our study. The data set may still be available via email request: laurence.steinberg@temple.edu.

3 Overview of MATLAB code

All MATLAB .m-files are contained in the folder titled "mfiles". The Master file is called "**master_DFVW.m**", which calls all the others. The code has lots of clarifying commentary, and the discussion below briefly summarizes how the code works in broad strokes.

3.1 "master_DFVW.m" - Master MATLAB file

1. Set path directories - adjust the main directory as needed.

2. Choose what to do: decide whether to (re-)estimate the model; simulate the model given a set of parameters; run the bootstrap standard error procedure; run the comparative statics exercises.

3. Define fixed model parameters which are not to be estimated (e.g. the number of grid points for the endogenous state variables, measurement parameters for the cognitive skill process, MATLAB solver settings, etc.)

4. Load the PSID-CDS panel data file, and give names to all the variables (i.e. the columns of "child_data_psid.txt"). Each row corresponds to a household-year observation. We created a balanced panel with 17 rows per household (with 247 households, so 4199 rows total). We also define some additional variables (e.g. variables which are not observed in the data but will be simulated in the model).

5. Utilize the Osaka Preference Parameter Study data set from the 2010 U.S. survey.

   - If the raw data set is available (i.e. after you have requested and obtained access to the raw data, and have used the STATA do-file called "master_osaka_betas.do" to clean it), then the master file calls a function called "**data_load_osaka.m**". This function loads the cleaned data set called "osaka_2010_final.txt", and uses it to calibrate the exogenous, discretized distribution of parents' discount factors,

conditional on their educational attainment level. The code also creates some figures used in the paper showing the conditional distribution in the data.

- Alternatively, if the raw data set is not available, the code simply defines the output necessary to run the rest of the code (i.e. summary statistics and moments of the raw data which would otherwise be calculated inside "**data_load_osaka.m**").

Then, the master file also uses a function called "**sim_init_betap.m**" to generate a matrix of random parent discount factors for each simulated household, drawn from the calibrated conditional distribution.

6. Utilize the Steinberg et al. (2009) survey data on children's and adolescents' discount factors.

   - If the raw data set is available (i.e. after you have requested and obtained access to the raw datafrom Prof. Steinberg), then the master file calls a function called "**data_load_steinberg.m**" which calibrates the exogenous grid points for the children's discount factors based on the observed data, and defines some external data moments which will be targeted in the estimation later.
   - Alternatively, if the raw data set is not available, the code simply defines the output necessary to run the rest of the code (i.e. summary statistics and moments of the raw data which would otherwise be calculated inside "**data_load_steinberg.m**").

7. Define various random number draws which will be used in estimation and simulation later on (e.g. for parental wage offers, non-labor income, children's realized test scores, children's Markov transitions and initial discount factor realizations).

8. Pre-draw initial latent child quality level drawn from beta distribution. We refer to the MATLAB code and the paper Appendix for more information on this process.

9. Calculate data moments. This calls a function called "**build_moments.m**" which takes in a data set (either real or simulated data) and returns a vector of moments, all of which are targeted in the estimation.

10. Bootstrap the Method of Simulated Moments weighting matrix, using a block bootstrap routine at the household-level, and use the inverse of the moments covariance matrix as weights.

11. Define vectors of initial guesses for the main model parameters, as well as the exogenous non-labor income process. Our default approach is to define the current "best guess" as the set of final converged estimates we obtained after estimating the model. This way, replicating the full code does not necessarily require re-estimating the whole model (which is time-consuming). It allows you to skip the estimation step (by setting "fp.estimate=0"), and simply simulate the model using our final set of estimates (by setting "fp.simulate=1").

12. Estimate the (truncated, non-negative) non-labor income process. This calls a function called "**run_income_process.m**" which separately estimates the parameters governing the exogenous non-labor income process.

    This file in turns calls several other m-files which respectively

(a) build the set of targeted moments related to the non-labor income distribution ("**NLI_moment_funct.m**");

(b) build the Method of Simulated Moments objective function ("**NLI_obj_funct.m**");

(c) simulate a vector of non-labor income draws conditional on a set of parameter estimates and household observables ("**NLI_sim_funct.m**").

Note: we only need to estimate the non-labor income process once, when running the code for the first time. After that, we can save time by changing the dummy "estimate_NLI_process" to 0 and simply simulate non-labor income draws at the converged set of estimates.

13. Define model primitive parameters. These parameters include utility, cognitive skill technology, child patience and wage parameters

14. Optionally: create some graphs for the paper showing additional summary statistics. This calls a function called "**graphs_intro.m**".

15. Estimate the baseline model parameters using the Method of Simulated Moments. This requires specifying an initial guess stored in a file called "temp_estimates.mat", which is then overwritten as the estimation procedure concludes.

    This calls the function "**build_obj.m**" which builds the objective function for the Method of Simulated Moments estimation routine (more details below). We use MATLAB's built-in Nelder-Mead simplex algorithm (fminsearch) to minimize the goal function, with some minor adjustments built into the file called "**fminsearch_matt.m**".

16. Simulate data under the model, given current parameter estimates. After doing this, we pass on the simulated (and real) data sets to a function called "**write_output_paper,m**", which compiles most of the figures and tables for the final paper.

17. Calculate Standard Errors for all model parameters (including the non-labor income process) using the bootstrap method. We refer to the paper and the commentary in the MATLAB code for more details on this procedure.

18. Comparative Statics exercise 1: varying the costs of using a CCT. This code evaluates various versions of the model with altered parameters, and compiles the results into a final table for the paper, using a function called "**tables_compstat_CCT.m**".

19. Comparative Statics exercise 2: shutting down education-based heterogeneity in parental wage offers, non-labor income, parental time productivity, parental and child discount factors. This code evaluates various versions of the model with altered parameters, and compiles the results into a final table for the paper, using a function called "**tables_compstat_SES.m**".

3.2 "build_obj.m": Outer loop MSM objective function

This file constructs the Method of Simulated Moments objective function, where the inputs include the current vector of parameters, all parameters of the model, variable names, observed data, and a vector of data moments.

1. Step 1: Unpack all of the model primitive parameters which are to be estimated jointly, i.e. preferences, cognitive skill technology, child discount factor process, and wage offers. (Note: non-labor income process parameters have already been estimated in a separate MSM routine, see "**run_income_process.m**").

2. Step 2: Simulate data given these parameters. This calls a function called "**sim_choices.m**", which handles the inner loop of the problem (see discussion below).

3. Step 3: Calculate simulated moments using "**build_moments.m**", which is the same function we used to calculate the moments of the real observed data.

4. Step 4: Generate the output of "**build_obj.m**", called "obj_eval".

    - If we are estimating the model ("fp.estimate=1"), then "obj_eval" is a scalar, given by the MSM objective function which depends on the squared distance between the data moments and the simulated moments (as well as the weighting matrix we specified earlier).
    - If we are not estimating the model ("fp.estimate=0"), but simply simulating household choices given a set of model parameters, then "obj_eval" is a larger structure, which contains a number of relevant outputs to be passed on, including all (transformed) model parameters, and the simulated data set.

3.3 "sim_choices.m" - Inner loop household optimization

This function solves the "inner loop" of the model, essentially iterating over different simulation rounds (indexed by r), households (indexed by q), and time periods (indexed by t), and solving the periodic optimization problem for each simulated household. This requires various computational tools, including:

- backward induction and forward simulation methods;
- discretization of the two endogenous state variables, given by the child's discount factor and stock of cognitive skills;
- various approximation methods, including (1) running surface regressions to approximate the parents' and child's future continuation values, and then - after verifying these regressions have a very high R-squared - using the regression coefficients in each backwards induction step, and (2) updating the next-period cognitive skill level of the child conditional on current-period investment choices, and approximating it by finding the nearest point on the discretized grid;
- numerical solution methods, i.e., using the MATLAB unconstrained minimization solver "**fminunc**", to solve for the household choice variables for which we don't have closed form solutions (i.e., the parents' two time investment choices)
- analytical solution methods, calculating (conditional) closed-form solutions for a subset of household choices. This step requires a function called "**stack_inner_solver.m**", which solves the household's optimization problem conditional on the household state variables, the binary CCT choice, and the parental time investments.

We refer to the comments in the MATLAB code and the Appendix of the paper for more details on the model solution and computational implementation.