Instructions for the Replication of the Results in "Careers in Firms: The Role of Learning about Ability and Human Capital Acquisition"

Elena Pastorino
Stanford University, Hoover Institution, and NBER

September 2023

1 Overview

This replication package contains the code used to estimate the baseline version of the model discussed in the main text of the paper. The Supplementary Appendix is available from the author's website at `https://sites.google.com/site/elenapastorino1econ/work` and provides details about the construction of the data sample used in estimation.

2 Artificial Data

Since the data used in estimation is proprietary, the code provided for the estimation of the model uses as input artificial data, which is supplied in the text file *fortran_sample_artificial.txt*. The artificial data is generated by the Stata program file *artificial_data_generation.do*, also provided, through the following six steps:

1. Create an empty data set containing the same variables (with empty values) and the same number of observations as the original sample used in estimation.

2. Generate random salaries by drawing person-year observations from a univariate uniform distribution over $[33,846, 64,014]$, which corresponds to the 10th-90th percentile range of the original pooled salary data from the estimation sample across all managers, years of employment, and job levels in the BGH firm's internal hierarchy of

jobs. Each draw is independent across individuals and within individuals over time. By construction, all possible salaries within the range described have equal probability.

3. Randomly draw high and low performance ratings for each manager and year of tenure at the firm. Each observation is independently drawn across individuals and within individuals over time from a uniform distribution over [0,1]—any draw above $p_{high} = 0.5$ is interpreted as a realization of a high performance rating.

4. Create random job level paths for each manager. Initialize all managers at Level 1 in the first year at the firm and then increase a manager's job level in each subsequent year by one with probability $p_{level} = 0.5$, assuming that no demotions occur and that the highest possible job level is Level 3. Once Level 3 is reached, an individual remains at this level until the end of the artificial observation period.

5. Randomly shorten careers at the firm by drawing a random separation year. Specifically, draw a random number of "failures" (no separation) until the first "success" (separation) from a binomial distribution with probability of success $p_{success} = 0.5$. If the separation year is greater than 11, assume that the worker is employed at the firm at least for the first 10 years. This procedure leads to relatively short careers with an average tenure length of two years.

6. Draw random demographic characteristics corresponding to managers' education at entry, age at entry, and year of entry into the firm from a range for each of these variables that is consistent with the observed one in the estimation sample, with each value having equal probability. Specifically, education takes any of the values (rounded to the next integer) drawn from a uniform distribution over [16,23], age takes any of the values (rounded to the next integer) drawn from a uniform distribution over [30,40]—this range has been shortened for reasons of data protection—and the entry of entry takes any of the values (rounded to the next integer) drawn from a uniform distribution over [1970,1979]. Each demographic characteristic is drawn independently.

Note that the only moments from the estimation sample that are used in this procedure are the truncated range of managers' salaries and the range of values of managers' education at entry and year of entry into the firm. Please contact Michael Gibbs at the University of Chicago Booth School of Business to inquire about how to obtain access to the original data.

3 Estimation Program

The model has been estimated using Fortran 90 (Intel compiler) and the Microsoft Visual Studio 2015 suite for it on a 64-bit Windows Server 2012 R2 with two 6-core processors Intel(R) Xeon(R) CPU E5-2643 v3 of 3.40GhZ and 768 GB RAM. The program consists of a solution file *Console1.sln*, a project file *Console1.vfproj*, the Fortran 90 source files, and three auxiliary text files. These text files are input files to the program: the file of the artificial data sample, *fortran_sample_artificial.txt*, a file for data trimming, *ind_delete.txt*, and the file listing the starting values of the model parameters, *param_values.txt*.

The suggested compilation order of the Fortran 90 files in Visual Studio is: *E_mod_types.f90*; *E_mod_nrutil.f90*; *E_mod_param_initial.f90*; *E_mod_parameters.f90*; *E_modwage.f90*; *E_amoeba.f90*; *E_sub_interp.f90*; *E_mod_grid.f90*; *E_mod_classerror.f90*; *E_mod_update.f90*; *E_sub_min_phi.f90*; *E_value_infinite.f90*; *E_outputfiles.f90*; *E_likegaussj.f90*; and *E_main.f90*. Outside of the Visual Studio suite, this order is also the suggested compilation and linking order for the Fortran 90 compiler—other orders are possible as long as they are consistent with the use statements of the component modules of the program.

The original program has been augmented to allow for alternative robustness exercises. The program that produces the estimates reported in the main text of the paper can be obtained by preserving the equality constraints listed in the Fortran 90 file *E_mod_parameters.f90*. The program has also been modified so as to prevent the identification of any individuals from the original data.