Supplemental Material for

HIP, RIP and the Robustness of Empirical Earnings Processes

by Florian Hoffmann, December 2018

The supplemental material, described in the following, includes:

- An online appendix to the paper that contains a detailed data description and theoretical and empirical results that are omitted from the main text.
- The empirical covariance structure for the main sample used in the estimation.
- Programs for the estimation of the full model and for the Monte Carlo analysis.

1. Online Appendix

The online appendix is structured in six sections, following the order in which they are mentioned in the paper:

A. Sample Construction

This section provides a detailed description of the IABS data and of the construction of the sample from which the covariance structure of earnings is computed.

B. Standard Errors

Given the size and administrative nature of the IABS data, standard errors of the EWMD-estimator cannot be calculated from individual-level data on earnings. This section discusses under which conditions cluster-robust standard errors estimated from cohort-level covariance structures are asymptotically valid. Clustering takes place on the cohort-level.

C. Formal Discussion of Implication 2

Implication 2 is the key identification result of the paper. It is stated informally in the main text. Appendix section C provides a formal treatment of this result.

D. Constrained Optimization and Computational Issues

Parameter estimation is non-trivial since the model is non-linear and since many of the parameters need to satisfy linear constraints. This section discusses these issues and how they are addressed.

E. How Robust are the Conclusions? Results from the Dropout Sample

This section, in combination with appendix table 4 and appendix figures 2, 3 and 5, reports and discusses results from the "robustness sample". The robustness sample is composed of high school dropouts. The covariance structure of earnings for this group is substantially different from the covariance structure of earnings for the "main sample". It is used as a robustness check for the main empirical results of the paper.

F. Finite Sample Performance: A Monte Carlo Simulation

This section describes the simulation protocol for and results from a Monte Carlo simulation that explores the finite-sample performance of the estimator. Results are shown in appendix table 5.

2. Empirical Covariance Structure

The covariance structure for the main sample is contained in the Matlab variable ***"autocovar_hsdeg_MATLAB"***. This variable has 56,072 rows, one for each element in the cohort-level covariance structure, and 94 columns. It can be viewed as a spread-sheet type data set containing the vectorized covariance structure in one column together with variables that describe their position in the matrix of cohort-level auto-covariances. The vectorized covariance structure, which enters the left hand side of equation (3.12) in the paper, is recorded in column 6. All other columns record variables entering the right hand side of this equation, or information from which these variables are calculated, or variables that facilitate computation of the right hand side in the estimation step. Each covariance term involves earnings measured at some time ***t*** (the "diagonal elements") and earnings measured ***k*** quarters later (the "lead elements").

The following lists the variables contained in each of the 94 columns. Notice that all these variables are also listed and named in lines 43-136 of the main Matlab program ("script_NLS") for estimation.

Col. 1: Time, varying on the quarterly level and normalized to one for 1975-q1
Col. 2: Calendar year
Col. 3: Quarter of the year
Col. 4: Year of Birth
Col. 5: Raw autocovariance
Col. 6: autocovariances, corrected for top-coding
Col. 7: Lag
Cols. 8 - 31: Year-of-Birth fixed effects (col. 8: yob = 1955; col. 31: yob = 1978)
Col. 32: A column of ones
Col. 33: A dummy variable equal to one for diagonal elements, i.e. variances.
Col. 34: The minimum of column 1, calculated by cohort
Col. 35: Potential labor market experience
Col. 36: (Col. 35)^2
Col. 37: (Col. 35)*2
Col. 38 – 41: Polynomial terms as a function of Col. 35, multiplying the parameters of the variance terms of the unit-roots component. Notice that these terms are sums of powers and can be calculated analytically. This also applies to the variables recorded in columns 42 and 43. For details, see section C in the appendix, where a number of analytical results are derived.
Col. 42: Variable multiplying sigma_ab
Col. 43: Variable multiplying sigma_b
Col. 44: The calendar year at which the lead-term is measured.
Col. 45-94: These are all dummy variables. Col. 45 is equal to one if the covariance term involves a diagonal element measured in 1980; col. 47 is the corresponding variable for 1981, etc. The even-numbered columns are the corresponding variables for the non-diagonal elements.

3. Matlab Programs

The supplemental material includes Matlab programs for the estimation of the full model. All other model specifications can be constructed from these programs. Because computing cluster-robust standard errors involves non-trivial computation, programs for reporting empirical results are contained in a separate sub-folder. These programs should be run after parameter estimation. Matlab programs for the Monte-Carlo simulation in appendix F are included as well.

All programs have detailed comments. Hence, only a brief description of the programs is offered here. Other than changing the paths, the programs should run unaltered on any regular Matlab license. A modification to using an extended KNITRO-Matlab license or a TOMLAB-license would be straightforward, as mentioned briefly below. Notice however that for estimating the models considered in this paper, the algorithms provided by a standard Matlab license perform well. For additional material and questions, please contact me directly via email (***florian.hoffmann@ubc.ca***).

A. Estimation of the Full Model

Programs for estimation of the model are contained in the "Estimation" folder. The main program is "script_NLS". The function being called from this program is "NLS_fct". Starting values come from estimation of a model without a measurement error component. This is a model for which the estimation algorithm converges very quickly (also see comments in the program). The starting values, which are estimates from this more restrictive model, are contained in the variable "estimates_fullunitrootshomosc".

B. Reporting Results

Computation of standard errors and tests for profile heterogeneity are contained in the "Reporting" folder. "Report_results" is the main program, from which "evaluate_covar" is called.

C. Monte-Carlo Simulation

This folder contains the programs for a Monte-Carlo analysis and corresponds to the experiment documented in panel "Simulation 1" of Appendix Table 5. Also see the description of the simulation protocol in appendix section F. Further detailed comments are given in the programs. Notice that the main program uses parallel processing. The Matlab parallel processing toolbox is required. "Simulation_quarterly" is the main program and simulates quarterly data from an unrestricted model. The programs "estimation_fct_unrestr", "estimation_fct_noHIP", and "estimation_fct_noInit" carry out parameter estimation. The first program imposes no restrictions and should thus give back unbiased estimates. The other two programs impose wrong restrictions in the estimation. The variable "sample_size" helps simulating individual-level samples with the same sample sizes and attrition rates as in the actual data.