USER's GUIDE FOR the STATA AND MATLAB CODES

The files in this folder allow one to replicate the empirical results reported in the paper "An Empirical Analysis of Labor Income Processes" by Fatih Guvenen published in RED (2008). In addition to the instructions below, each STATA and MatLab program contains instructions at the top of the file for inputs/outputs when applicable as well as comments throughout the programs that explain the intent of each piece of code.

ESTIMATION of HIP AND RIP PROCESSES:

Following the steps below allows the user to perform the estimation of the income process (both HIP and RIP versions) for the whole population. At the end of the instructions I explain how to do the estimation for the college and high-school samples separately.

STEP 1:

The STATA code **PREP_COVMAT.DO** takes as input the Stata data file named **"ready_newdata.dta"** and creates the covariance matrix of income residuals that is used in the estimation and to construct the figures and tables in the paper. Simply run the code in Stata version 7 or later to obtain the covariance matrices used in step 2 below.

Details of the Stata Code:

The input STATA data file **"ready_newdata.dta"** essentially contains variables from the Panel Study of Income Dynamics downloaded from PSID's website. Only individuals in the SEO (non-random poverty sample) and the Latino subsample are excluded when downloading the data. Essentially no other sample cleaning or selection has been done up to this point. The **PREP_COVMAT.DO** program also does the sample cleaning and selection as described in the appendix of the paper. This program generates a separate covariance matrix for each cohort (each matrix has agemax by agemax elements) that will be merged in the next step below. Corresponding to each covariance matrix, the code also generates a matrix whose elements report the number of observations that are used to compute the

covariance. The code contains further comments that explain the purpose of different parts of the program.

STEP 2:

Run the MatLab program **LoadCovNew96.m** which essentially stacks the covariance matrices for each cohort to create a 3 dimensional matrix that has (agemax x agemax x #of cohorts) elements. It also stacks the number of observations matrix. The resulting two matrices are the inputs to the main estimation in the next step.

STEP 3:

The main estimation in the paper uses three MatLab programs: **EstimARMA_NEW_April_04.m**, **MinDistOBJ_NEW_March04.m**, **minDistOBJ_DV_NEW_March04.m**

The first program is the main code that the user calls, which in turn calls the latter two codes internally. The inputs to the main program is further explained in the comments section at the top of the program. Essentially, the two matrices created in step 2 above are the main inputs.

The program estimates the model and outputs the estimated parameter vector, the minimized objective value and a flag variable that indicates if the minimization of the objective function yielded an acceptable solution.

Estimation for College and High-school subsamples:

Follow the exact same steps as above with ONE change. Before running the STATA code in step 1, COMMENT OUT line 151 and UNCOMMENT line 152. Then follow all remaining steps. This performs the estimation for the college sample. For high-school sample, proceed as above, but now COMMENT OUT line 151 and UNCOMMENT line 153.

GENERATING MACURDY's TEST:

To calculate the auto-covariances of income growth at different lags (reported in Table 3) the user needs to simply run the MatLab program named **GENERATE_MaCurdy_TEST.m**. This program takes as input a panel of simulated labor income observations (dimensions: #of individuals by #of time periods). This panel is generated by simply simulating equation (2) in the paper by setting $g()=0$, $f()=\alpha^i + \beta^i h$, and $\phi_t = 1$. The code also allows the user to set the highest order of autocovariances to be computed (called L; set to 18 in the paper) and the number of times the exercise should be repeated (called Nsim; set to 1000 in the paper). The outputs are the first L autocovariances (called "covx") and autocorrelations (called "corrx") across Nsim replications. This empirical distribution is used to calculate the mean and the standard deviations in Table 3 of the paper.