

RISKY MATCHING

HECTOR CHADE AND ILSE LINDENLAUB

READ-ME FILE

April 29, 2021

Contents

1	Overview	2
2	Data Availability	2
3	Replication Instructions	3
3.1	Data Preparation Instructions	3
3.2	Statistical Decomposition and Estimation Instructions	3
3.3	Instructions to Reproduce Final Tables and Figures	4
4	List of Programs and Produced Output	5
4.1	Data Preparation	5
4.2	Statistical Decomposition	6
4.3	Estimation	6
4.4	Descriptive Statistics and Figures (Online Appendix)	10
4.5	Simulation (Online Appendix)	10
5	Computational Requirements	11
6	References	12

1 Overview

This appendix explains the codes and data to reproduce our results. For an overview, the replication of the results can be divided into five different parts and seven main folders:

- Data Preparation
 - Folder: `Occupational_Tasks`
 - * Sub-folder: `DOT`
 - * Sub-folder: `ONET`
 - Folder: `Test_Scores`
 - * Sub-folder: `Original_Data_Altonji_et_al.(2009)`
 - Folder: `Wage_Data`
- Statistical Decomposition
 - Folder: `Statistical-Decomp`
- Estimation
 - Folder: `Estimation`
 - * Sub-folder: `Counterfactuals_Code`
 - * Sub-folder: `Estimation_Code`
 - * Sub-folder: `Output`
 - * Sub-folder: `Weighting_Matrix`
- Descriptive Statistics and Figures
 - Folder: `Appendix_Figures`
- Simulation
 - Folder: `Simulation`

2 Data Availability

We use four data sources, all available for public use:

1. Current Population Survey (CPS). We rely on the Annual Social and Economic Supplement (ASEC) of the CPS. We downloaded the CPS raw data from IPUMS CPS.
2. Dictionary of Occupational Titles (DOT). We downloaded this data from David Autor’s website (Autor, Levy, and Murnane (2003), QJE) (follow the links: ‘DOT means by occupation’ and ‘Consistent occupation crosswalk files’ on that website).
3. O*NET. We downloaded this data (Version 22.2.) from the Onetcenter website.
4. National Longitudinal Survey of Youth 1979 and 1997 (NLSY79 and NLSY97). We use the AFQT scores prepared by Altonji, Bharadwaj, and Lange (2009), and downloaded their data here.

3 Replication Instructions

3.1 Data Preparation Instructions

Here is an overview of the steps to re-construct our different datasets, see below for the details.

1. We first prepare the CPS data for our analysis by applying the restrictions specified in the paper and Online Appendix 2.1.1. Our CPS dataset is saved in `cps_data.dta` in Folder `Wage_Data`.
2. We construct a cognitive skill requirement for each occupation. In the early period, we do so based on DOT data, where we use `dot77-8090.dta` and `occ90.dta` in Folder `Occupational_Tasks/DOT/Original_Data_Autor_et_al.(2003)`. In the later period, we do so based on the O*NET data-files `Abilities.xlsx`, `Skills.xlsx` and `Work_Activities.xlsx` in Folder `Occupational_Tasks/ONET/Original_Data_O*NET`. See Online Appendix 2.1.2. for details. We run the do-files `cognitive_skill_dot.do` and `cognitive_skill_onet.do` in Folders `Occupational_Tasks/DOT` and `Occupational_Tasks/ONET`, respectively, to produce the datasets we are working with: `dot_final.dta` and `onet_final.dta`.
3. We then construct the empirical G (job) and Q (ability) distributions. For the G distribution, we use the constructed cognitive job attributes, see point 2. For the Q distribution, we use AFQT test scores from the NLSY, made comparable across periods by Altonji et al, see Online Appendix 2.1.3. for details. We use their file `afqt_adjusted_final.dta` (which they produced using their code `AFQT MATCHING with weights.do`) in Folder `Test_Scores/Original_Data_Altonji_et_al.(2009)`. To construct these distributions, run the do-file `data_distributions.do` in Folder `Wage_Data`.
4. We calculate 10 moments from the CPS data which will serve as inputs (targets) in the estimation, using CPS, DOT and O*NET data. To do so, run do-file `data_moments.do` in Folder `Wage_Data`.

3.2 Statistical Decomposition and Estimation Instructions

Here is an overview of the steps to reproduce the paper's main results. Below are more details.

1. Run `statistical_decomp.do` in Folder `Statistical-Decomp` to obtain the results from the statistical decomposition of wage inequality.
2. Run, in that order, `gmm_estimation_inputs_fixedtheta_final.m` (see further instructions in the code) and `gmm_call_fixedtheta_final.m` in Folder `Estimation/Counterfactual_Code` to obtain the estimation results from the counterfactual model with exogenous investment. Then, run `Decomp_fixedtheta_final.m` in the same folder to obtain the decomposition of inequality changes into changes of this model's primitives.
3. Run, in that order, `gmm_estimation_inputs_final.m` (see further instructions in the code) and `gmm_call_final.m` in Folder `Estimation/Estimation_Code` to obtain the estimation results from the baseline model. Then, run `Decomp_final.m` in the same folder to obtain the decomposition of inequality changes into changes of model primitives.

4. Run `post_estimation_final.m` in Folder `Estimation/Estimation_Code` to produce the paper's main tables and figures.

3.3 Instructions to Reproduce Final Tables and Figures

1. `post_estimation_final.m` reproduces Figures 3-6, Tables 1-13 of the paper.
2. `simulation_multiplicity.nb` reproduces Figure 1 in the Online Appendix.
3. `cognitive_skill_dot.do` reproduces Table 2 and Figure 2 in the Online Appendix.
4. `appendix_figures.do` reproduces Tables 3-5 and Figure 3 in the Online Appendix.

Note: In each code, the default paths need to be adjusted by the user.

4 List of Programs and Produced Output

4.1 Data Preparation

Folder: Occupational_Tasks

Sub-folder: DOT

Code	Description	Output
cognitive_skill_dot.do	Computes the cognitive task measure per census occupation in 1990 for the early period, based on data from the DOT. Inputs: Original_Data_Autor_et_al.(2003)/dot77-8090.dta Original_Data_Autor_et_al.(2003)/occ90.dta	dot_final.dta

Sub-folder: ONET

Code	Description	Output
cognitive_skill_onet.do	Runs a PCA to compute a single cognitive task measure per census occupation in 1990 for the later period, based on data from O*NET. Inputs: Original_Data_O*NET/Skills.xlsx Original_Data_O*NET/Work_Activities.xlsx Original_Data_O*NET/Abilities.xlsx Crosswalks/Crosswalk_census1990_soc2010.dta	onet_final.dta Figure2_OnlineAppendix.eps Table2_OnlineAppendix.txt

Folder: Test_Scores

Sub-folder: Original_Data_Altonji_et_al.(2009)

Contains test score data from NLSY. We directly use the data of this subfolder, `afqt_adjusted_final.dta`, in do-file `Wage_Data/data_distributions.do`.

Folder: Wage_Data

Code	Description	Output
data_moments.do	Calculates the moments that serve as inputs in the estimation. Inputs: cps_data.dta ./Occupational_Tasks/DOT/dot_final.dta ./Occupational_Tasks/ONET/onet_final.dta	college_share.csv college_share.dta college_premium.csv college_premium.dta percentiles.csv percentiles.dta moments.csv moments.dta inc_var.csv inc_var.dta

Code	Description	Output
data_distributions.do	Builds on the codes in the folders Occupational_Tasks and Test_Scores to construct the empirical G and Q distributions for both periods. These distributions will serve as inputs in the estimation.	G_early.csv G_later.csv G_expanded_early.csv G_expanded_later.csv G_expanded_sample_early.csv G_expanded_sample_later.csv Q_early.csv Q_later.csv
	Inputs:	
	cps_data.dta	
	./Occupational_Tasks/DOT/dot_final.dta	
	./Occupational_Tasks/ONET/onet_final.dta	
	./Test_Scores/Original_Data_Altonji_et_al.(2009)/afqt_adjusted_final.dta	

4.2 Statistical Decomposition

Folder: Statistical_Decomp

Code	Description	Output
statistical_decomp.do	Computes statistical decomposition of moments of the wage distribution in both periods.	counterfactuals.csv
	Inputs:	
	./Wage_Data/cps_data.dta	

4.3 Estimation

Folder: Estimation

Sub-folder: Weighting_Matrix

Contains the weighting matrices for the GMM estimation.

For baseline model with endogenous investment:

Weight_early.mat

Weight_later.mat

For counterfactual model with exogenous investment:

Weight_early_fixedtheta.mat

Weight_later_fixedtheta.mat

To construct them, we performed an ‘m out of n’ bootstrap on our set of moments and then define the weighting matrix as the inverse of the diagonal of the covariance matrix of the bootstrapped moments.

Sub-folder: Counterfactuals_Code

Code	Description	Output
<code>gmm_estimation_inputs_fixedtheta_final.m</code>	<p>Specifies the pre-set parameters, and inputs the empirical moments, the weighting matrix and the empirical distributions; and it defines the estimation problem. Code calls function <code>moments_fixedtheta_final.m</code>, which produces the model moments and evaluates the GMM objective function – but no need to run it separately.</p> <p>Inputs:</p> <ul style="list-style-type: none"> <code>../Wage_Data/college_share.csv</code> <code>../Wage_Data/college_premium.csv</code> <code>../Wage_Data/percentiles.csv</code> <code>../Wage_Data/moments.csv</code> <code>../Wage_Data/inc_var.csv</code> <code>../Wage_Data/G_expanded_sample_early.csv</code> <code>../Wage_Data/G_expanded_sample_later.csv</code> <code>../Wage_Data/Q_early</code> <code>../Wage_Data/Q_later</code> <code>./Weighting_Matrix/Weight_early_fixedtheta.mat</code> <code>./Weighting_Matrix/Weight_later_fixedtheta.mat</code> <code>moments_final_fixedtheta.m</code> 	<code>inputs_gmm_fixedtheta.mat</code>
<code>gmm_call_fixedtheta_final.m</code>	<p>Runs the GMM estimation.</p> <p>Inputs:</p> <ul style="list-style-type: none"> <code>inputs_gmm_fixedtheta.mat</code> 	<p><code>gmm_output_2015original_fminsearch_fixedtheta.mat</code></p> <p><code>gmm_output_1980original_fminsearch_fixedtheta.mat</code></p>
<code>Decomp_fixedtheta_final.m</code>	<p>Performs a decomposition of wage inequality into the different sources. The code is essentially the same as <code>Decomp_final.m</code> in the subfolder <code>Estimation_Code</code>, but with exogenous educational investment (exog. θ^*).</p> <p>Inputs:</p> <ul style="list-style-type: none"> <code>gmm_output_1980original_fminsearch_fixedtheta.mat</code> <code>gmm_output_2015original_fminsearch_fixedtheta.mat</code> <code>../Wage_Data/Q_early.csv</code> <code>../Wage_Data/Q_later.csv</code> <code>../Wage_Data/G_expanded_sample_early.csv</code> <code>../Wage_Data/G_expanded_sample_later.csv</code> <code>../Wage_Data/college_share.csv</code> 	<p><code>Decomp_7525_sep_fixedtheta.mat</code></p> <p><code>Decomp_skillpremium_sep_fixedtheta.mat</code></p>

Sub-folder: Estimation_Code

Code	Description	Output
gmm_estimation _inputs_final.m	<p>Specifies the pre-set parameters, and inputs the empirical moments, the weighting matrix and the empirical distributions; and it defines the estimation problem. Code calls functions <code>moments_1980_final.m</code> and <code>moments_2015_final.m</code>, which produce the model moments and evaluate the GMM objective function in each period – but no need to run those separately.</p> <p>Inputs:</p> <p> <code>../Wage_Data/college_share.csv</code> <code>../Wage_Data/college_premium.csv</code> <code>../Wage_Data/percentiles.csv</code> <code>../Wage_Data/moments.csv</code> <code>../Wage_Data/inc_var.csv</code> <code>../Wage_Data/G_expanded_sample_early.csv</code> <code>../Wage_Data/G_expanded_sample_later.csv</code> <code>../Wage_Data/Q_early.csv</code> <code>../Wage_Data/Q_later.csv</code> <code>./Weighting_Matrix/Weight_early.mat</code> <code>./Weighting_Matrix/Weight_later.mat</code> <code>moments_1980_final.m</code> <code>moments_2015_final.m</code> </p>	<code>inputs_gmm.mat</code>
gmm_call_final.m	<p>Runs the GMM estimation.</p> <p>Inputs:</p> <p><code>inputs_gmm.mat</code></p>	<p><code>gmm_output_2015original_fminsearch.mat</code></p> <p><code>gmm_output_1980original_fminsearch.mat</code></p>
Decomp_final.m	<p>Performs a decomposition of wage inequality into the different sources.</p> <p>Inputs:</p> <p> <code>gmm_output_1980original_fminsearch.mat</code> <code>gmm_output_2015original_fminsearch.mat</code> <code>../Wage_Data/Q_early.csv</code> <code>../Wage_Data/Q_later.csv</code> <code>../Wage_Data/G_expanded_sample_early.csv</code> <code>../Wage_Data/G_expanded_sample_later.csv</code> </p>	<p><code>Decomp_7525_sep.mat</code></p> <p><code>Decomp_skillpremium_sep.mat</code></p> <p><code>Decomp_theta_sep.mat</code></p>

Code	Description	Output
post_estimation_final.m	This file inputs parameter estimates and empirical distributions to conduct post estimation exercises and plot the empirical distributions in both periods.	Figure3_1.eps Figure3_2.eps Figure4_1.eps Figure4_2.eps Figure5_1.eps Figure5_2.eps
	Inputs:	Figure6.eps
	gmm_output_1980original_fminsearch.mat	Table1
	gmm_output_2015original_fminsearch.mat	Table2
	Decomp_theta_sep.mat	Table3
	Decomp_skillpremium_sep.mat	Table4
	Decomp_7525_sep.mat	Table5
	latexable.m	Table6
	../Wage_Data/college_share.csv	Table7
	../Wage_Data/college_premium.csv	Table8
	../Wage_Data/percentiles.csv	Table9
	../Wage_Data/moments.csv	Table10
	../Wage_Data/inc_var.csv	Table11
	../Wage_Data/G_expanded_sample_early.csv	Table12
	../Wage_Data/G_expanded_sample_later.csv	Table13
	../Wage_Data/G_expanded_early.csv	
	../Wage_Data/G_expanded_later.csv	
	../Wage_Data/Q_early.csv	
	../Wage_Data/Q_later.csv	
	./Counterfactuals_Code/gmm_output_1980original_fminsearch_fixedtheta.mat	
	./Counterfactuals_Code/gmm_output_2015original_fminsearch_fixedtheta.mat	
	./Counterfactuals_Code/Decomp_skillpremium_sep_fixedtheta.mat	
	./Counterfactuals_Code/Decomp_7525_sep_fixedtheta.mat	
	./Weighting_Matrix/Weight_early.mat	
	./Weighting_Matrix/Weight_later.mat	
	./Weighting_Matrix/Weight_early_fixedtheta.mat	
	./Weighting_Matrix/Weight_later_fixedtheta.mat	
	../Statistical_Decom/counterfactuals.csv	

Sub-folder: Output

All output from post_estimation_final.m is saved here.

4.4 Descriptive Statistics and Figures (Online Appendix)

Folder: Figures_Appendix

Code	Description	Output
appendix_figures.do	Generates the descriptive figures and tables for the Online Appendix.	Table3_OnlineAppendix.csv Table4_OnlineAppendix.csv Table5_OnlineAppendix.csv Figure3_OnlineAppendix.eps
	Inputs: ./Occupational_Tasks/DOT/dot_final.dta ./Occupational_Tasks/ONET/onet_final.dta ./Wage_Data/cps_data.dta ./Wage_Data/Q_early.csv ./Wage_Data/Q_later.csv ./Wage_Data/G_expanded_sample_early.csv ./Wage_Data/G_expanded_sample_later.csv	

4.5 Simulation (Online Appendix)

Folder: Simulation

Code	Description	Output
simulation_multiplicity.nb	Constructs a graph showing multiple equilibria based on a parametric example of the model.	Figure1_OnlineAppendix.eps

5 Computational Requirements

Software:

- `Stata Version 15.1`
 - `gquantiles`, which is part of package `gtools`, installation instructions are here
- `Matlab Version R2018b`
 - `latextable.m` (Andrew E. Slaughter, 2009)
- `Mathematica Version 12.1.0.0`

Memory and Runtime Requirements: The code was last run on a Intel Core i5 laptop with MacOS version 10.13.1.

6 References

Altonji, Bharadwaj and Lange (2009). Constructing AFQT Scores that are Comparable Across the NLSY79 and the NLSY97. Data retrieved from here.

Autor, Levy, and Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics*, 118(4), 1279-1334. Data retrieved from here.

Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1979 cohort, 1979-2016 (rounds 1-27). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2019.

Flood, King, Rodgers, Ruggles and Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 8.0 [Annual Social and Economic Supplement (ASEC)]. Minneapolis, MN: IPUMS, 2020. <https://doi.org/10.18128/D030.V8.0>

U.S. Department of Labor, Employment and Training Administration. O*NET 22.2 Database. <https://www.onetcenter.org/database.html>