Replication Package

Environment Setup

All code in this replication package assumes that an environment variable is correctly configured for both Stata and R. This setup ensures consistent file paths across different systems, enabling seamless execution of the scripts.

For both R and Stata, the environment variable `US_Ineq_Repl` must be defined and should point to the Analysis folder. The following sections outline how to set up this environment variable for each statistical package.

Setting Up Stata Environment

Stata automatically executes a `profile.do` script upon startup (if found). This script should be stored in a directory that Stata searches automatically. You can view the eligible paths by typing:

```
adopath
```

A common practice is to define global directory variables inside `profile.do`, specifying system-specific settings. The example below demonstrates defining the path for Dropbox, which is commonly used for file synchronization across multiple computers:

```
* Settings specific to local environment
global DROPBOX "C:/Users/YourName/Dropbox"

* Run a script with shared settings across environments
run "$DROPBOX/stata_profile.do"
```

In turn, the `stata_profile.do` file located inside the Dropbox directory contains project-specific definitions:

```
set varabbrev off
set reshape_favor speed
set doeditbackup off
global US_Ineq_Repl "$DROPBOX/Replication/Analysis"
```

This method ensures all computers connected to Dropbox share the same project directory structure.

Setting Up R Environment

Similar to Stata's `profile.do`, R runs an `.Rprofile` file upon startup (if found). To locate your home directory where `.Rprofile` should be stored, run the following command in R:

```
normalizePath(path.expand("~"), winslash="/")
```

An `.Rprofile` file may contain environment-specific settings, such as defining project paths:

```
cat("Running .Rprofile...\n\n")

# Settings specific to local environment
Sys.setenv(DROPBOX = "C:/Users/YourName/Dropbox")

# Load shared settings from Dropbox
source(file.path(Sys.getenv("DROPBOX"), "R_profile.R"))
```

A secondary script stored in Dropbox, `R_profile.R`, then defines paths for projects:

```
Sys.setenv(US_Ineq_Repl = file.path(Sys.getenv("DROPBOX"), " Replication
/Analysis"))
```

This method ensures that all systems running the code use the same directory structure, improving consistency across different environments.

Folder Structure and Organization

This replication package follows a structured directory setup to facilitate reproducibility and efficient data processing. The root directory is `Analysis`, which contains four main subdirectories: `Data`, `Processed`, `Results`, and `Scripts`. Each serves a distinct purpose, ensuring a clear separation between raw data, processed outputs, final results, and executable scripts.

Directory Overview

```
Analysis/
├── Data/
│   ├── CEPR/
│   │   ├── CPS_Basic/
│   │   │   ├── CEPR/
│   │   │   ├── NBER/
│   │   ├── CPS_ORG/
│   │   │   ├── CEPR/
│   │   │   ├── NBER/
│   │   ├── Software/
├── Processed/
│   ├── CEPR/
├── Results/
│   ├── Figures/
│   ├── Tables/
├── Scripts/
│   ├── CH/
│   ├── MM/
│   ├── FFL/
```

Directory Purpose

- **`Data/`**: Contains raw input data. This directory is read-only and includes datasets from CEPR, NBER, and CPS, as well as relevant software.
- **`Processed/`**: Stores intermediate processed data files. These are generated by scripts and should not be modified manually.
- **`Results/`**: Contains final outputs, including figures and tables.
- **`Scripts/`**: Holds all code files needed to run the analysis, divided into subdirectories for different methodologies (e.g., `CH` for my proposed approach, `MM` for Machado and Mata, and `FFL` for the Firpo, Fortin and Lemieux method).

Environment Configuration

Your `US_Ineq_Repl` environment variable should be set to the directory containing the `Analysis` folder. This ensures that all file paths resolve correctly across different systems.

Running the Scripts

To execute the full analysis, navigate to the `Analysis` folder and run the scripts in the following order:

1. **`0_run.R`** – Downloads necessary data, including CPIs and CPS data from CEPR, NBER, and the Census. The `Data` directory will contain the raw data files, while the `CEPR` folder within `Data` holds the `CPS_Basic` and `CPS_ORG` subdirectories, each containing their respective original CEPR scripts in the `CEPR` folder. The corresponding `NBER` directories will store the downloaded data after running this script.
2. **`1_run.do`** – Processes wage data using the CEPR scripts to construct the main dataset. It also generates Figures E.1a and E.1b, which display kernel density estimates of wages over time, and Table 1 in the main text, which presents summary statistics.
3. **`2_run.R`** – Estimates the conditional quantile regression model for 1986 and 2015 and produces Figure 1. This script also implements the Machado and Mata methodology and generates Table 3a in the main text.
4. **`3_run.R`** – Implements my proposed methodology to decompose the conditional Gini index. It generates Tables 2 and 3b, as well as all figures and tables in Appendix C**.**
5. **`4_run.do`** – Executes the analysis in Appendix F, implementing the Firpo, Fortin, and Lemieux (FFL) methodology for unconditional quantile regressions and decomposition.

Complete Folder Structure and Organization

The `Analysis` directory is the root folder of the replication package, containing all necessary subdirectories and scripts for data processing, analysis, and results generation. Below is a detailed breakdown of the directory structure:

1. Data (Raw Data)

This folder contains all original data sources, organized by dataset origin and format:

- **`CEPR`** (Data from CEPR's uniform extracts)
  - **`CPS_Basic`** – Basic monthly CPS files
    - **`CEPR`** – Contains CEPR processing scripts
      - **`DoFiles`** – CEPR's original `.do` scripts
    - **`NBER`** – Will contain downloaded CPS Basic data after running `0_run.R`
  - **`CPS_ORG`** – Outgoing Rotation Group (ORG) files
    - **`CEPR`** – Contains CEPR processing scripts
      - **`Do`** – CEPR's original `.do` scripts
    - **`NBER`** – Will contain downloaded CPS ORG data after running `0_run.R`
  - **`Software`** – Any additional software dependencies
- **Consumer Price Index (CPI) Files**
  - `CPI_allitems.pdf`
  - `CPI_CUSR0000SA0.xlsx`
  - `CPI_CUUR0000SA0.xlsx`
  - `cpi_mwage.xlsx`

2. Processed (Intermediate Data)

Contains processed datasets derived from the raw CPS data:

- **`CEPR`** – Processed versions of the CEPR datasets

3. Results (Final Output)

Stores the generated figures and tables from the analysis:

- **`Figures`** – Graphical outputs from all analyses
- **`Tables`** – Tables produced for the paper

4. Scripts (Code)

Houses all scripts used in the analysis, organized into subdirectories based on methodology:

- **`CH`** – Scripts related to the conditional quantile regression method
- **`MM`** – Scripts implementing the Machado and Mata (2005) method
- **`FFL`** – Scripts for the Firpo, Fortin, and Lemieux (FFL) methodology
  - **`Data`** – Data processed specifically for the FFL approach
  - **`Figures`** – Figures generated from the FFL methodology
  - **`Tables`** – Tables produced from the FFL decomposition

5. Root Scripts

These scripts should be run in sequence inside the **`Analysis`** folder:

- `0_run.R` – Downloads all required data
- `1_run.do` – Processes wage data and generates summary statistics
- `2_run.R` – Implements conditional quantile regression and the Machado-Mata method
- `3_run.R` – Implements the conditional Gini index decomposition
- `4_run.do` – Runs the FFL decomposition in **Appendix F**

Full Directory Overview

```
Analysis/
├── Data/
│   ├── CEPR/
│   │   ├── CPS_Basic/
│   │   │   ├── CEPR/
│   │   │   │   ├── DoFiles/
│   │   │   ├── NBER/
│   │   ├── CPS_ORG/
│   │   │   ├── CEPR/
│   │   │   │   ├── Do/
│   │   │   ├── NBER/
│   │   ├── Software/
│   ├── CPI_allitems.pdf
│   ├── CPI_CUSR0000SA0.xlsx
│   ├── CPI_CUUR0000SA0.xlsx
│   ├── cpi_mwage.xlsx
├── Processed/
│   ├── CEPR/
├── Results/
│   ├── Figures/
│   ├── Tables/
├── Scripts/
│   ├── CH/
│   ├── MM/
│   ├── FFL/
│   │   ├── Data/
│   │   ├── Figures/
│   │   ├── Tables/
├── 0_run.R
├── 1_run.do
├── 2_run.R
├── 3_run.R
├── 4_run.do
```