# Replication Package

### Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment

by **John Horton**

## 1 Overview

This document describes the organization of code, data, and software associated with "Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment" (Horton, 2025a). The replication package for this project is available at the Github repository johnjosephhorton/minimum_wage (Horton, 2025b) as well as the openICPSR repository 208551V1 (Horton, 2025c).

This project is largely self-documenting in the sense that the key recipes for building the project are captured in code. The key pieces for replicating this project are:

1. A Makefile that gives recipes for how each figure, table, and called-out number is constructed in the paper.
2. A Dockerfile that describes the system set-up needed with respect to computing resources and packages.

A comprehensive list of files, programs, and software dependencies is provided in the current document for the reader's convenience.

## 2 Folder Structure

The main directory contains three folders:

```
-analysis
-codebooks
-writeup
```

The `analysis` folder contains 23 R scripts that generate all outputs used in the paper. (A complete list of programs as well as software requirements and dependencies is given in Section 3.1.)

The `codebooks` folder contains 10 Markdown files that give an overview of the contents of each dataset used in the analysis (one Markdown file for each dataset). (A complete list of datasets used in this project is given in Section 3.2.)

The `writeup` folder contains the main writeup of the paper (see `minimum_wage.tex`). The figures, tables, and parameter files generated during replication will be stored in `writeup/plots`, `writeup/tables`, and `writeup/parameters`, respectively. In the repositories hosting the replication package, these locations are empty by design. When the project is built, these folders are populated with the necessary files.

When the project is built (see Section 2 below), a `data` folder is automatically created in the main directory and populated with the relevant datasets. (There is no need to create an empty `data` folder manually.)

## 3 Replication

A replicator will obtain a `.env` file from the author and can reproduce the entire writeup of the paper using one of three available options.

### 3.1 Building the Paper

The building of this project is orchestrated by a software called make. From the `writeup` folder at the command line, simply type:

```
make minimum_wage.pdf
```

which will then build the paper. Below is a short tutorial on `make`. Familiar readers may skip the tutorial and go to Section 2.2 directly.

#### 3.1.1 A Short Make Tutorial

A Makefile lists recipes for how a particular output used in the paper is constructed. The Makefile entry for the `first_stage.pdf` in the `plots` folder looks like this:

```
plots/first_stage.pdf: ../analysis/first_stage.R  data/df_mw_first.csv
 cd ../analysis && Rscript first_stage.R
```

Note that the target, or output, is `plots/first_stage.pdf`. After the colon are the dependencies—here, the R file `first_stage.R` in the `analysis` folder and the dataset `data/df_mw_first.csv`. The indented line is the recipe that constructs the target by running `Rscript` on `first_stage.R`.

There is code shared across multiple figures or tables. For example, `utilities_outcome_experimental_plots.R` contains helper functions used in several plots. To capture this dependency, the Makefile has entries like:

```
analysis/plot_any_exper.R: analysis/utilities_outcome_experimental_plots.R
 touch analysis/plot_any_exper.R
```

The command `touch` updates the timestamp of `analysis/plot_any_exper.R`, which in turn causes the recipe for `any_exper.pdf` to be re-run because `plot_any_exper.R` is a dependency.

### 3.2 Getting the Data

To get the data, obtain a `.env` file from the author and place it in the `~/minimum_wage/writeup` folder. The `.env` file contains: (1) a private URL to the zipped and encrypted data, and (2) a key to decrypt the data. An example `.env` file is:

```
GPG_PASSPHRASE='<password>'
project_name="minimum_wage"
DROPBOX_URL="<url of dropbox link hosting the data>"(base)
```

These values are used by the bash script `fetch_data.sh` in the main directory.

There are three options to get the data:

#### 3.2.1 Non-Docker, Local Approach (Recommended)

Run the following in the command line:

```
$ cd writeup
$ make docker
```

The script `system_update.sh` in the main directory will automatically install the right dependencies on a Linux machine and build the project.

```
git clone git@github.com:johnjosephhorton/minimum_wage.git
cd minimum_wage
cp ~/Downloads/.env writeup
sudo ./system_update.sh
```

#### 3.2.2 Docker Approach

Clone the repository and (assuming the `.env` file is in your Downloads folder):

```
git clone git@github.com:johnjosephhorton/minimum_wage.git
cd minimum_wage/writeup
cp ~/Downloads/.env .
```

Then run:

```
make docker
```

The final PDF will be inside the Docker container, which will provide a localhost URL to access the PDF.

#### 3.2.3 Replit Approach (Convenient)

A public Replit "repl" is available at `https://replit.com/@johnhorton/minimumwage`. Fork this repository, add the `.env` file to the `writeup` folder, and press the big green "Run" button. The dependencies are specified in the replit.nix file in the main directory. The PDF will be generated and stored in the `writeup` folder.

# 4 Code, Datasets, and Outputs

This section describes software requirements, datasets used, and outputs of the analysis.

## 4.1 Program and Software

The entire project has been implemented in R (version 4.4.1). The `analysis` folder contains 23 R scripts that utilize 10 datasets to generate all outputs reported in the paper.

#### 4.1.1 R Scripts

Table below provides a list of all 23 R scripts located in the `analysis` folder.

| R Script | R Script |
|---|---|
| avg_wages_by_cat.R | first_stage.R |
| jjh_misc.R | parameters.R |
| parameters_country_selection.R | parameters_effects.R |
| plot_any_exper.R | plot_application_event_study.R |
| plot_composition.R | plot_did_all_outcomes.R |
| plot_event_study_hired_admin.R | plot_event_study_hourly_rate_hired.R |
| plot_feedback.R | plot_fill_and_hours.R |
| plot_follow_on_openings.R | plot_hours_zero.R |
| plot_organic_applications.R | quantile_hours_worked.R |
| randomization_check.R | realized_wage_distro.R |
| settings.R | table_any_prior.R |
| utilities_outcome_experimental_plots.R | |

#### 4.1.2 R Packages

The complete list of R packages used in the scripts above, along with the versions last used to run the code, is given below.

| R Package (version) | R Package (version) |
|---|---|
| cowplot (1.1.3) | data.table (1.16.2) |
| directlabels (2024.1.21) | dplyr (1.1.4) |
| ggplot2 (3.5.1) | ggrepel (0.9.6) |
| gridExtra (2.3) | gt (0.11.1) |
| lfe (3.1.1) | lmtest (0.9-40) |
| lubridate (1.9.4) | magrittr (2.0.3) |
| plyr (1.8.9) | purrr (1.0.2) |
| quantreg (6.00) | sandwich (3.1-1) |
| scales (1.3.0) | stargazer (5.1) |
| tidyr (1.3.1) | |

## 4.2 Datasets

The analysis uses 10 different datasets, all automatically stored in the `data` folder after initiating the data download procedure. The `codebooks` folder contains 10 Markdown files that document the contents of each dataset. The table below lists all datasets along with a brief description.

| Dataset | Description |
|---|---|
| df_mw_first.csv | Main experimental outcomes at job post level (first observation) |
| df_mw_all.csv | Main experimental outcomes at job post level (all observations) |
| df_mw_admin.csv | Main experimental outcomes at job post level for admin data |
| df_mw_lpw.csv | Main experimental outcomes at job post level for low wage positions |
| df_exp_results.csv | Aggregated experimental results and summary statistics |
| hires_country_composition.csv | Geographic composition of hires |
| event_study_windows.csv | Time windows and intervals used in the event study analysis |
| did_panel.csv | Data used for the DiD analysis |
| event_study_hired.csv | Detailed hiring outcome data for event study analyses |
| event_study_windows_hr_v_fp.csv | Composition of fixed price and hourly jobs over time |

## 4.3 Outputs

### 4.3.1 Figures

The table below maps each figure used in the paper to the generating R script and its dataset dependencies. All figures are stored in `writeup/plots` after the project is built.

| Figure | R Script | Data Dependency | Location in Paper |
|---|---|---|---|
| Figure 1 | analysis/realized_wage_distro.R | data/df_mw_first.csv | Page 125 |
| Figure 2 | analysis/first_stage.R | data/df_mw_first.csv | Page 126 |
| Figure 3 | None (TikZ Diagram) | None | Page 126 |
| Figure 4 | analysis/plot_fill_and_hours.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Page 130 |
| Figure 5 | analysis/plot_composition.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Page 134 |
| Figure 6 | analysis/plot_event_study_hourly_rate_hired.R | data/event_study_hired.csv | Page 138 |
| Figure 7 | analysis/plot_did_all_outcomes.R | data/did_panel.csv | Page 140 |
| Figure 8 | analysis/plot_application_event_study.R | data/event_study_windows.csv | Page 143 |
| Figure A1 | analysis/plot_organic_applications.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Appendix |
| Figure A2 | analysis/plot_follow_on_openings.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Appendix |
| Figure A3 | analysis/avg_wages_by_cat.R | data/df_mw_first.csv | Appendix |
| Figure B1 | analysis/plot_hours_zero.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Appendix |
| Figure B2 | analysis/plot_any_exper.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Appendix |
| Figure B3 | analysis/plot_feedback.R | data/df_mw_all.csv, df_mw_admin.csv, df_mw_lpw.csv | Appendix |
| Figure B4 | analysis/plot_event_study_hired_admin.R | data/event_study_hired.csv | Appendix |

#### 4.3.2 Tables

The table below maps each table used in the paper to its generating R script and dataset dependencies. All tables are stored in `writeup/tables` after the project is built.

| Table | R Script | Data Dependency | Location in Paper |
|---|---|---|---|
| Table A1 | `analysis/randomization_check.R` | `data/df_mw_first.csv` | Appendix |
| Table B1 | `analysis/quantile_hours_worked.R` | `data/event_study_windows_hr_v_fp.csv` | Appendix |
| Table B2 | `analysis/table_any_prior.R` | `data/df_mw_first.csv` | Appendix |

#### 4.3.3 Parameters

Some scripts generate "parameters" (numbers automatically called out in the paper). For example, after building the project, the file `parameters/parameters.tex` will contain a line such as:

```
\newcommand{\numTotal}{159,656}
```

which relates to the total allocation to the experiment. In the `parameters.R` script, this value is created as follows:

```
addParam("\\numTotal", formatC(dim(df.mw.first)[1], big.mark = ","))
```

The table below maps each parameter file to its generating R script. All parameter files are saved in `writeup/parameters` after the corresponding R scripts are run.

| Parameter File | R Script |
|---|---|
| `parameters/parameters.tex` | `analysis/parameters.R` |
| `parameters/effects_parameters.tex` | `analysis/parameters_effects.R` |
| `parameters/parameters_fill_and_hours.tex` | `analysis/plot_fill_and_hours.R` |
| `parameters/did_parameters.tex` | `analysis/plot_did_all_outcomes.R` |
| `parameters/parameters_composition.tex` | `analysis/plot_composition.R` |
| `parameters/params_country_selection.tex` | `analysis/parameters_country_selection.R` |

### 4.4 License for Code

The code is licensed under an MIT license.

## 5 Data Availability and Provenance

### 5.1 Provenance

This data comes from a large online labor market (Anonymous Online Platform, 2025) that conducted the experiment described in the paper. The data was pulled from the company's database using SQL. The data used for this analysis is proprietary and confidential, but may likely be obtained via Data Use Agreements with the data provider. The data provider remains anonymous. Researchers interested in accessing the data may contact John Horton at `jjhorton@mit.edu`. The author will assist with any reasonable replication attempts for two years following publication.

### 5.2 Statement about Rights

- ✓ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ❍ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permissions are documented in the LICENSE.txt file in the main directory.

### 5.3 Summary of Availability

- ❍ All data **are** publicly available.
- ❍ Some data **cannot be made** publicly available.
- ✓ **No data can be made** publicly available.

## 6 Computational Requirements

The computational requirements for reproducing this paper are minimal. The project can be run on a modern laptop in about 20 minutes using only open-source software.

The machine details when last run are:

- OS: "Ubuntu 20.04.6 LTS"
- Processor: 11th Gen Intel(R) Core(TM) i7-11850H @ 2.50GHz, 16 cores
- Memory available: 31GB

The code was last run on 2025-03-27 15:16:41. Linux dependencies include LaTeX and various R package requirements.

### 6.1 Runtime

Approximate time needed to reproduce the analyses on a standard 2023 desktop machine is about 20 minutes:

- ❍ <10 minutes
- ✓ 10–60 minutes
- ❍ 1–2 hours
- ❍ 2–8 hours
- ❍ 8–24 hours
- ❍ 1–3 days
- ❍ 3–14 days
- ❍ >14 days
- ❍ Not feasible to run on a desktop machine, as described below.

## 7 References


- Anonymous Online Platform. 2025. "Data For: Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment," Unpublished Data, Accessed March 29, 2025.
- Horton, John J. 2025a. "Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment." American Economic Review 115, No. 1: 117–146. https://doi.org/10.1257/aer.20170637.
- Horton, John J. 2025b. "minimum_wage." GitHub Repository. Accessed March 27, 2025. https://github.com/johnjosephhorton/minimum_wage.
- Horton, John J. 2025c. "Code For: Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment." American Economic Association, Inter-university Consortium for Political and Social Research. http://doi.org/10.3886/E208551V1.