# Effects of Educational Reforms in the Turkish context


- Mustafa Yağız Kılıçarslan
- Joel Gomez
- Timothy Hanson

# Table of contents

# 01

# Introduction

Background of study

# Introduction

In 2012 the Compulsory schooling mandate of Turkey increased schooling from 8 to 12 years. This policy change has impacted numerous amount of people when implemented. At the time many families argued that their children would not benefit from this additional years of education. Their argument was that their children are better off working at an earlier age to gain **on-the-job training** which would earn them more money **compared to additional schooling**. Similarly these families argued that **their daughters should marry** and start a family **instead of having education**. While education and earnings have a multifaceted relationship, we investigate the causal aspects of this relationship using highschool completion as our primary treatment variable

# Where did we take the sample from?

**Turkey**

Turkish Labor data
(2021-2022)

# Dataset Description

Labour Force Statistics Micro Data Set, 2022:
Compiled by the **Turkish Statistical Institute (TurkStat)**, this dataset enables **in-depth analysis** of Turkey's labor market dynamics through the **Household Labour Force Survey (HLFS).**

- Scope and Methodology:
  - Geographic Coverage: All regions of Turkey, reported at NUTS1 and NUTS2 levels.
  - Sample Size: Quarterly surveys involving 58,560 households.
  - Collection Methods: Computer-Assisted Personal Interviewing (CAPI) and Computer-Assisted Telephone Interviewing (CATI).
- Key Variables:
  - Demographics: Age, gender, education level.
  - Employment: Job type, hours worked, labor force participation.
  - Unemployment: Job search activities, duration, and reasons.
  - Income: Monthly earnings and sources.
- Data Format: Delivered in CSV for use in statistical and econometric analyses.

# Background on Educational Reforms in Turkey

Turkey's educational reforms aim to strengthen the alignment between **workforce skills and labor market demands**, fostering human capital development and economic growth.

- Key Reforms:
  1. Compulsory Education Expansion:
     - Increased **mandatory schooling years**, boosting educational attainment and labor force participation among secondary and tertiary graduates.
  2. Vocational Education Modernization:
     - Improved **alignment of technical education with market demands**, reducing skill mismatches.
  3. Gender Equity in Education:
     - Enhanced **female participation in education**, leading to higher female labor force engagement.
  4. Lifelong Learning Initiatives:
     - Promoted reskilling and **continuous education** to meet evolving economic needs.

# Research question

Did the 2012 reform extending compulsory schooling by four years in Turkey affect the wages in the labour market during 2021–2022 of students who completed their education under this system?

If so, what is the effect of this treatment?

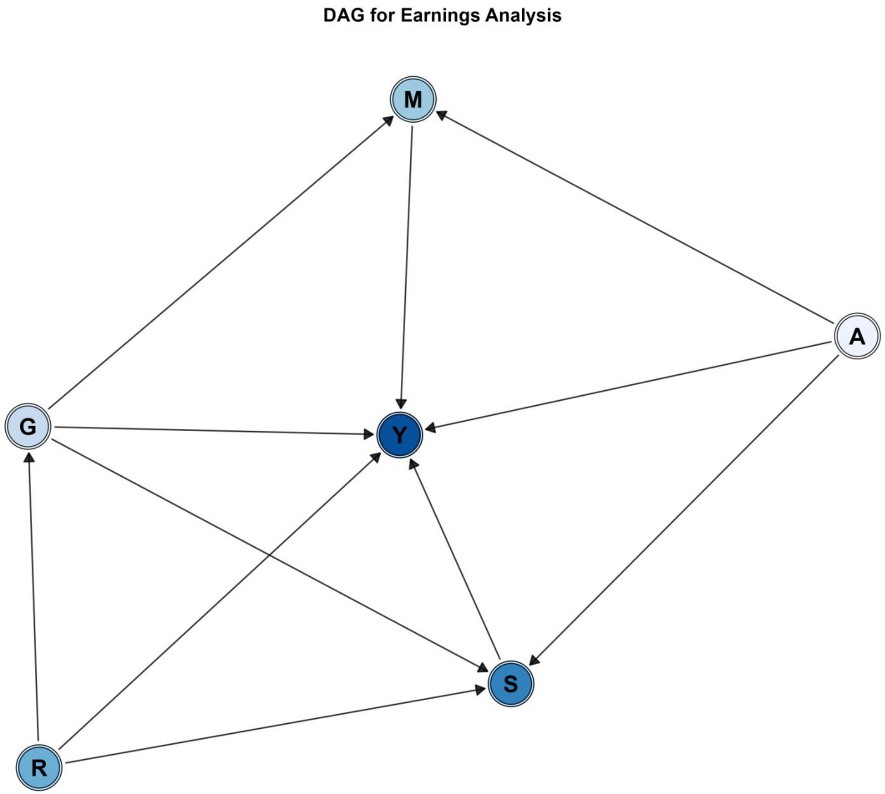
# Contribution of the project

The effect of schooling on wages is **endogenous**, as can be seen in the DAG. However, using **Double Machine Learning (DML)** we are able to overcome this problem. Thus, it is **possible to find** out whether there is a **causal effect** of schooling on wages.

Moreover, this study is the **first to address** the Turkish education reform of 2012 -as well as education reforms in general- **using** an innovative method such as **DML**.

# DAG

Y: Log hourly earnings
R: Region
G: Gender
A: Age
M: Marital Status
S: Schooling



# 02

# Methods

Step taken

# DATA CLEANING & FEATURE ENGINEERING

## DATA CLEANING

- *COMBINED THE DATASETS {2021,2022}*
- *RENAMED OUR VARIABLES*

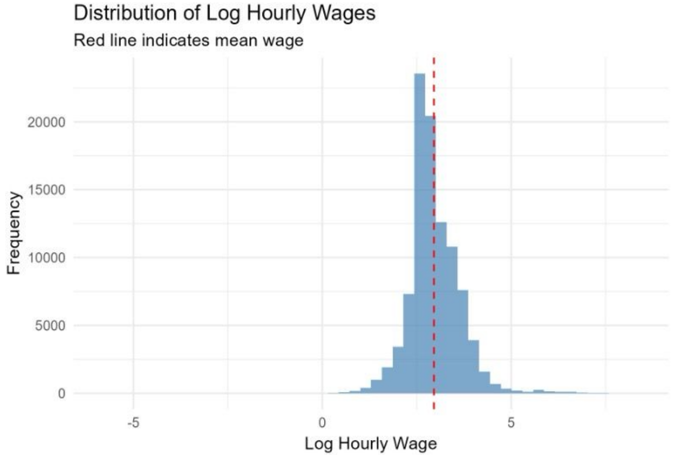
## FEATURE ENGINEERING

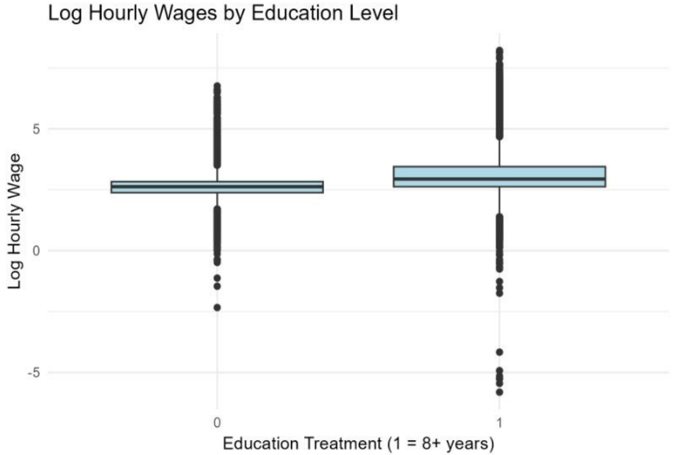
- *ENCODING CATEGORICAL VARIABLES(MARITAL STATUS, REGION)*
- *NON-LINEAR CONTINUOUS VARIABLES (AGE, AGE-SQUARED)*
- *CREATING LOG-HOURLY WAGE OUTCOME VARIABLES*
- *CREATING TREATMENT VARIABLES FOR COMPULSORY SCHOOLING*

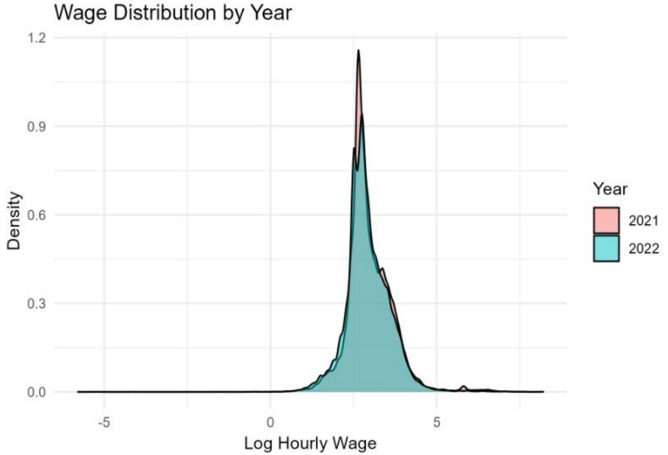
# EXPLORATORY DATA ANALYSIS

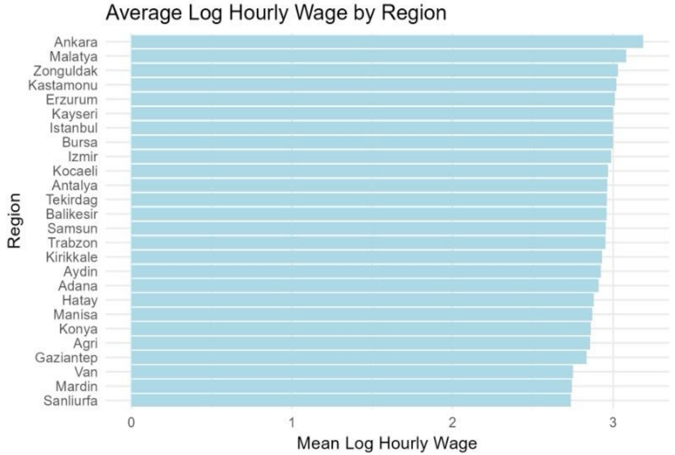
- *FILTER NA'S AND INTERMEDIATE VARIABLES*
- *CHECK IF MISSING VALUES ARE MORE COMMON IN CERTAIN GROUPS*
- *CREATE PLOTS TO VISUALIZE THE ASPECTS OF THE DATA*

# VISUALIZATION OF THE ASPECTS OF THE DATA



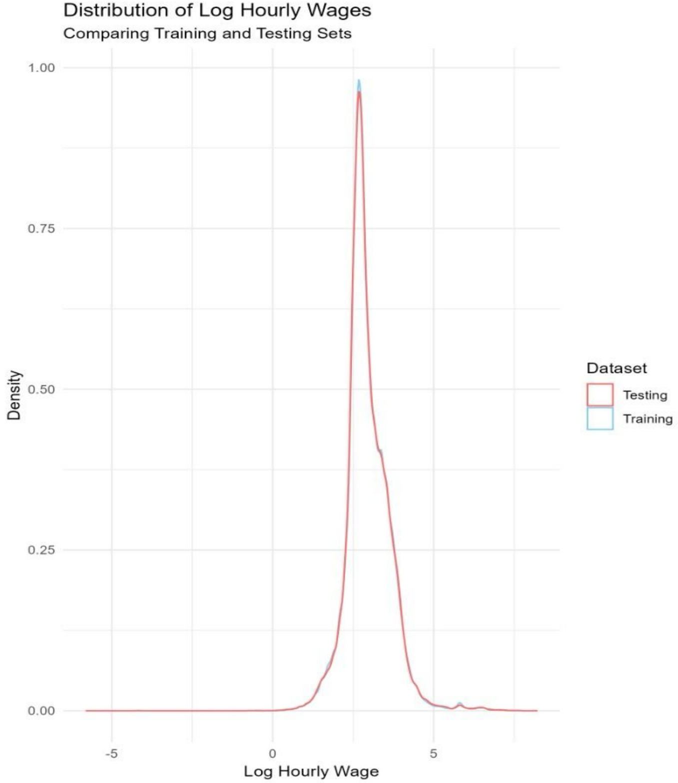



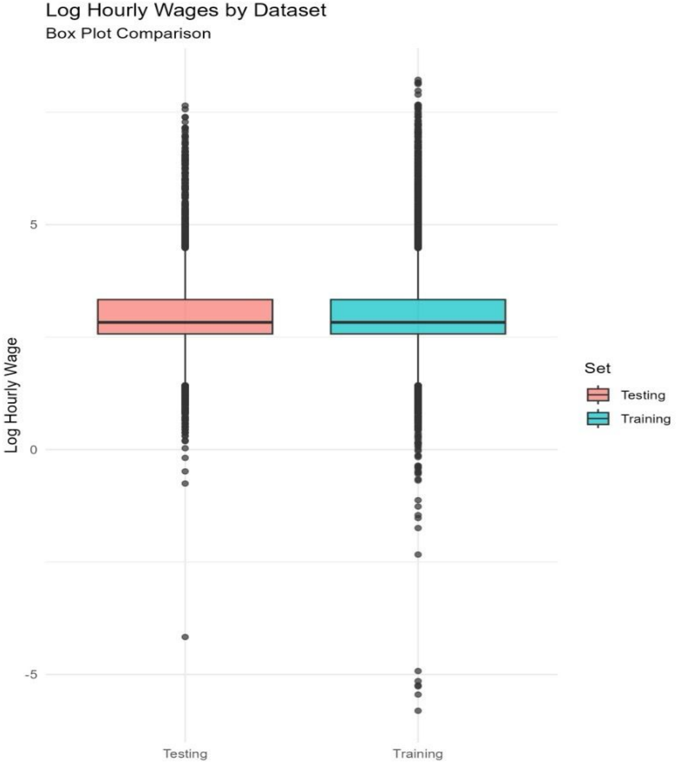



# MACHINE LEARNING MODELS

- *CHECK THE BALANCE OF THE DISTRIBUTION OF THE DATA*
- *REGRESSION MODEL*
- *DOUBLE MACHINE LEARNING*

# BALANCE OF DATA SPLITS



# REGRESSION MODEL

$$\log\_hourly\_wage = \beta_0+\beta_1 age+\beta_2 age^2+\beta_3 marital\_status+\beta_4 region+\beta_5 gender+ \beta_6 working\_hours + \beta_7 highschool\_completion$$

# DOUBLE MACHINE LEARNING

- Prepare data for DML
- Outcome variable (Y):log_hourly_wage,
- Treatment variable (D):Highschool education,
- Control variables (X): age, age_sq, gender, marital_status, region, working_hours
- Create Double ML data object
- Create learners for both stages
- For predicting High-school education (classification task)
- For predicting wages (regression task)
- Create and fit DML model
- Fit the model

# 03

# Results

Outcomes and Limitations

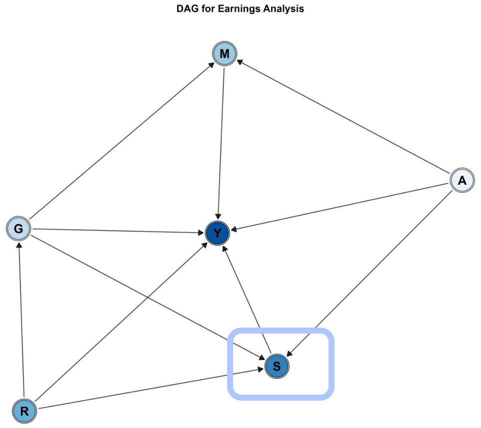
# REGRESSION MODEL

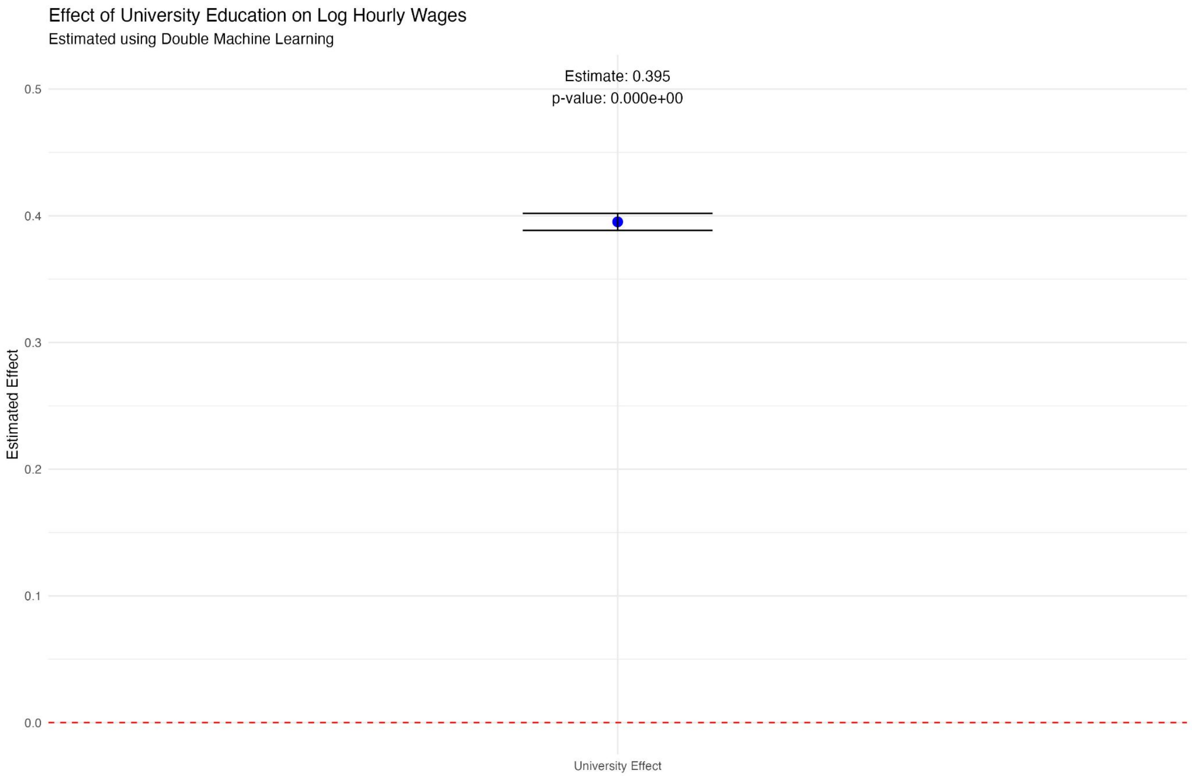
| term <chr> | estimate <dbl> | std.error <dbl> | statistic <dbl> | p.value <dbl> |
|---|---|---|---|---|
| (Intercept) | 2.68271750 | 0.003142203 | 853.76957 | 0.000000e+00 |
| age | 0.61094475 | 0.014177513 | 43.09252 | 0.000000e+00 |
| age_sq | -0.54032359 | 0.013586607 | -39.76884 | 0.000000e+00 |
| treatment2 | 0.43566841 | 0.004074671 | 106.92112 | 0.000000e+00 |
| working_hours | -0.35597757 | 0.001917370 | -185.65932 | 0.000000e+00 |
| gender_male | -0.08450761 | 0.001910236 | -44.23935 | 0.000000e+00 |
| marital_status_Ma... | 0.03672317 | 0.002668978 | 13.75926 | 5.074654e-43 |
| region_Manisa | -0.03546040 | 0.002193526 | -16.16594 | 1.110789e-58 |
| region_Adana | -0.03265749 | 0.002156648 | -15.14271 | 1.015723e-51 |
| region_Hatay | -0.02934735 | 0.002129216 | -13.78317 | 3.647950e-43 |

# PREDICTED VS ACTUAL


| .metric <chr> | .estimator <chr> | .estimate <dbl> |
|---|---|---|
| rmse | standard | 0.4917730 |
| rsq | standard | 0.4651916 |
| mae | standard | 0.3499310 |

# DOUBLE MACHINE LEARNING



# Interpretation of Results

1. **Comparing the estimates of Linear Regression and DML**
   - LinReg estimated that completing highschool increases wages by 43% compared to people who did not.
   - DML estimated that same treatment has 39% effect on wages.

2. **Statistical significance**
   - From ~0 p-values we can argue that they are statistically significant

3. **Causal Inference**
   - DML removes the confounding effects on treatment variable hence its estimates can be used for causal arguments

# Limitations

1. Our model validation relies on a single train-test split, which may not fully capture the model's performance variability. Implementing k-fold cross-validation would provide more robust performance estimates and better account for potential overfitting.

2. We face potential selection bias by only observing wages for employed individuals in our dataset. This could skew our estimates since we cannot observe the potential wages of those who chose not to work.

3. Our analysis assumes uniform returns to education across different demographics. However, educational benefits likely vary by gender, region, and age cohort. Including interaction terms or conducting subgroup analyses would provide a more nuanced understanding of the policy's impact across different population segments.

# 04

# CONCLUSION

INTERPRETATION OF RESULTS

Research Question: Was “4+4+4” policy effective?

Baseline Model Linear Regression showed promising results with ~50% accuracy (RMSE) and it predicted completing highschool increased wages by 43%

Using the DAG we argued that there would be endogenous variation in our treatment variable (highschool completion, specifically gender and region)

Then we used a state-of-the-art Double Machine Learning Technique for eliminating this endogenous variation.

Our DML model predicted that highschool completion increases wages by 39% - 4% less than baseline prediction.

We argue that this 4% was the overestimation of baseline model coming from endogenous variation.

Hence we argue that real effect of highschool completion is 39%, statistically significant and notably large.

Hence we conclude that policy change was a effective.

# Thank you!