

The Rise in Returns to Skill?

A Modern Regression Analysis of Wage Inequality in the Current Population Survey (CPS)

Senan Hogan-H.

ECON 190, Pomona College

senan.hoganhennessy@pomona.edu

25 April 2018

Overview

- 1 Research Motivation
 - Research Question
 - Recent Trends
 - Literature review
- 2 Methodology
 - Data Section
 - Decomposition Methods
 - Prediction Methods
- 3 Results
 - Decomposition
 - Variable Importance
- 4 Summary
 - Overview of Results
 - Discussion and Limitations

Research Question

Research Question

- Income inequality has been rising rapidly since the 1960s.

Research Question

- Income inequality has been rising rapidly since the 1960s.
- What is driving this rise in inequality?

Research Question

- Income inequality has been rising rapidly since the 1960s.
- What is driving this rise in inequality?
- Can standard statistical methods explain rising inequality?

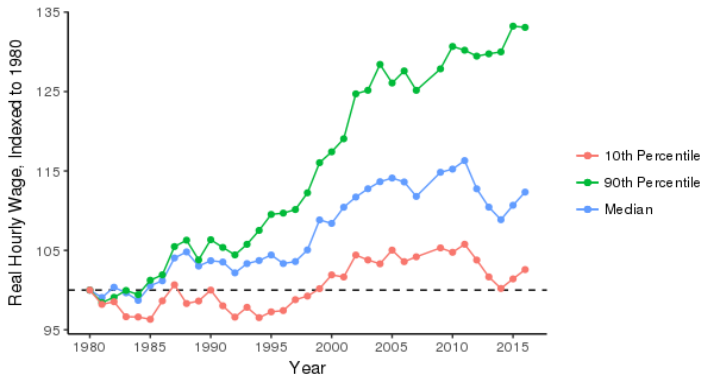
Research Question

- Income inequality has been rising rapidly since the 1960s.
- What is driving this rise in inequality?
- Can non-standard statistical methods explain rising inequality?

Recent Trends

Recent Trends

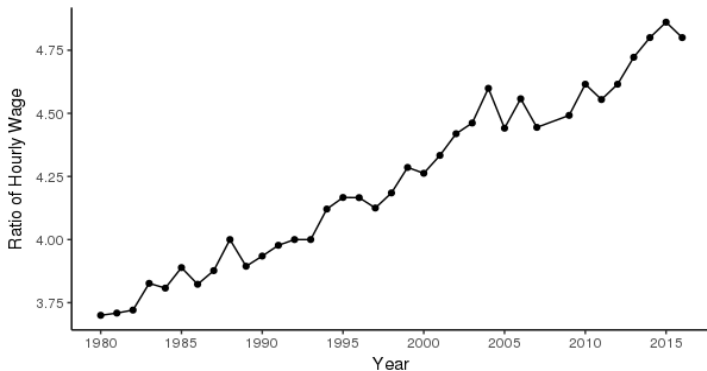
Figure: Indexed Real Hourly Wage by Percentile, 1980-2016.



Recent Trends

Recent Trends

Figure: Ratio of Wage Between 90th and 10th Percentiles, 1980-2016.



Previous Research

Previous Research

- Juhn, Murphy, Pierce (1993, JMP 1993 henceforth) document rising wage inequality

Previous Research

- Juhn, Murphy, Pierce (1993, JMP 1993 henceforth) document rising wage inequality
- They present a decomposition method for wage inequality that is extremely influential

Previous Research

- Juhn, Murphy, Pierce (1993, JMP 1993 henceforth) document rising wage inequality
- They present a decomposition method for wage inequality that is extremely influential
- Attribute most of the rise in inequality to 'rise in return to skill

Previous Research – My Research's Place

Previous Research – My Research's Place

- My research looks to decompose rising inequality with adjusted version of the JMP method

Previous Research – My Research's Place

- My research looks to decompose rising inequality with adjusted version of the JMP method
- Aim to show how components for rising inequality change with prediction methods

Previous Research – My Research's Place

- My research looks to decompose rising inequality with adjusted version of the JMP method
- Aim to show how components for rising inequality change with prediction methods
- Expose the decomposition's sensitivity to specific prediction methods used

Data - March CPS

Data - March CPS

Current Population Survey

- Current Population Survey (CPS)
 - Serves as the nation's primary source of statistics on labor force characteristics.
 - The Annual Social and Economic Supplement provides the official annual statistics on the nation's poverty levels as well as statistics on income, health insurance coverage, marital status, educational attainment, employee benefits, work schedules, school enrollment, noncash benefits and migration.

[CPS Table Creator](#)

The CPS Table Creator gives you the ability to create customized tables from the Current Population Survey's Annual Social and Economic Supplement (CPS ASEC).

United States
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Used the CEPR March CPS extracts. (accessed 2018)

JMP Decomposition

JMP Decomposition

Standard linear model for predicting a measure of wages individual years:

$$Y_{it} = \mathbf{X}_{it}\beta_t + \varepsilon_{it} \quad (1)$$

JMP Decomposition

Standard linear model for predicting a measure of wages individual years:

$$Y_{it} = \mathbf{X}_{it}\beta_t + \varepsilon_{it} \quad (1)$$

Consider returns to observed characteristics constant over the entire time period:

$$Y_{it} = \mathbf{X}_{it}\bar{\beta} + \varepsilon_{it} \quad (2)$$

JMP Decomposition

Standard linear model for predicting a measure of wages individual years:

$$Y_{it} = \mathbf{X}_{it}\beta_t + \varepsilon_{it} \quad (1)$$

Consider returns to observed characteristics constant over the entire time period:

$$Y_{it} = \mathbf{X}_{it}\bar{\beta} + \varepsilon_{it} \quad (2)$$

Note:

JMP Decomposition

Standard linear model for predicting a measure of wages individual years:

$$Y_{it} = \mathbf{X}_{it}\beta_t + \varepsilon_{it} \quad (1)$$

Consider returns to observed characteristics constant over the entire time period:

$$Y_{it} = \mathbf{X}_{it}\bar{\beta} + \varepsilon_{it} \quad (2)$$

Note: $\bar{\beta} \neq \beta_t$, for $t = 1980, \dots, 2016$

JMP Decomposition

JMP Decomposition

(observed income distribution)

$$Y_{it} =$$

JMP Decomposition

(observed income distribution)

$$Y_{it} =$$

(predicted income distribution, under fixed returns)

$$\mathbf{X}_{it}\bar{\beta}$$

+

JMP Decomposition

(observed income distribution)

$$Y_{it} =$$

(predicted income distribution, under fixed returns)

$$\mathbf{X}_{it}\bar{\beta}$$

+

(difference to predicted income distribution, under variable returns)

$$\mathbf{X}_{it}(\beta_t - \bar{\beta})$$

+

JMP Decomposition

So that the distribution of income is the sum of three components:

$$Y_{it} = \mathbf{X}_{it}\bar{\beta} + \mathbf{X}_{it}(\beta_t - \bar{\beta}) + \varepsilon_{it} \quad (3)$$

JMP Decomposition

So that the distribution of income is the sum of three components:

$$Y_{it} = \mathbf{X}_{it}\bar{\beta} + \mathbf{X}_{it}(\beta_t - \bar{\beta}) + \varepsilon_{it} \quad (3)$$

⇒ This approach relies heavily on the variables in \mathbf{X}_{it} or even prediction method used.

JMP Decomposition

So that the distribution of income is the sum of three components:

$$Y_{it} = \mathbf{X}_{it}\bar{\beta} + \mathbf{X}_{it}(\beta_t - \bar{\beta}) + \varepsilon_{it} \quad (3)$$

⇒ This approach relies heavily on the variables in \mathbf{X}_{it} or even prediction method used.

⇒ Perhaps we could use more than OLS?

Prediction Methods

Prediction Methods

Mincer Wage equation, dating to Mincer (1954, 1972):

$$Y_{it} = Y_0 + \rho_t s_{it} + \beta_{1t} x_{it} + \beta_{2t} x_{it}^2 + \varepsilon_{it} \quad (4)$$

Prediction Methods

Mincer Wage equation, dating to Mincer (1954, 1972):

$$Y_{it} = Y_0 + \rho_t s_{it} + \beta_{1t} x_{it} + \beta_{2t} x_{it}^2 + \varepsilon_{it} \quad (4)$$

Adjusted Mincer wage equation, proposed by Lemieux (2006):

$$Y_{it} = Y_0 + \rho_{1t} s_{it} + \rho_{2t} s_{it}^2 + \beta_{1t} x_{it} + \beta_{2t} x_{it}^2 + \beta_{3t} x_{it}^3 + \beta_{4t} x_{it}^4 + \varepsilon_{it} \quad (5)$$

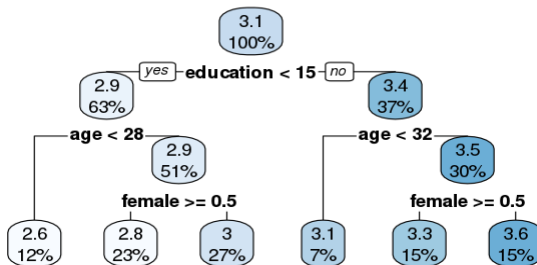
Prediction Methods

Novel prediction methods, tree-based methods.

Prediction Methods

Novel prediction methods, tree-based methods.

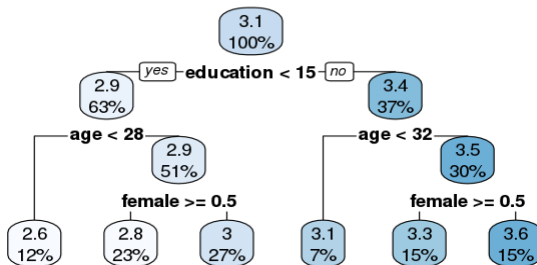
Figure: Decision tree for real log wages, 2010-2016.



Prediction Methods

Novel prediction methods, tree-based methods.

Figure: Decision tree for real log wages, 2010-2016.

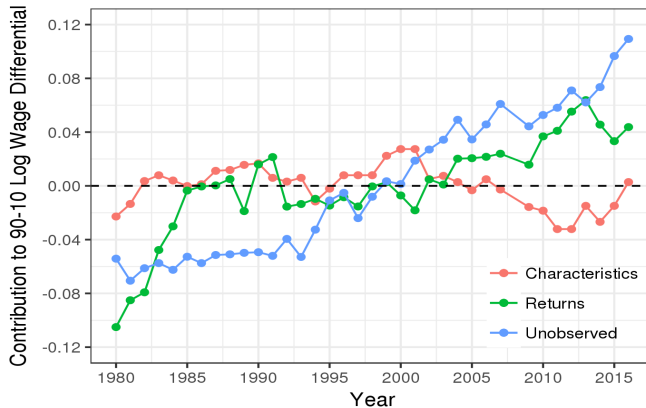


Random forest prediction uses random sampling and building a forest of decision trees to avoid over-fitting.

Results - Mincer Wage Equation

Results - Mincer Wage Equation

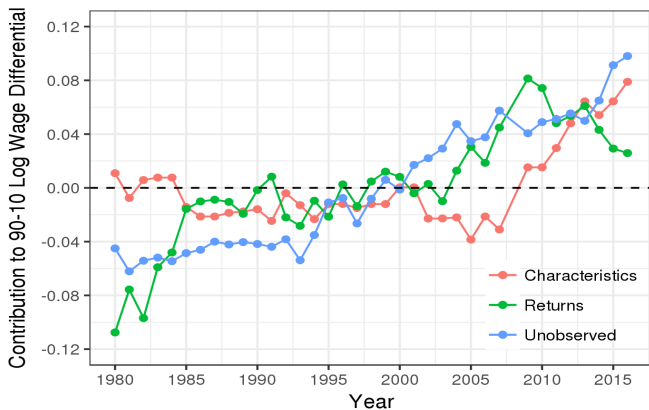
Figure: Components of 90-10 Percentile Log Wage Differential.



Results - Adjusted Mincer Wage Equation

Results - Adjusted Mincer Wage Equation

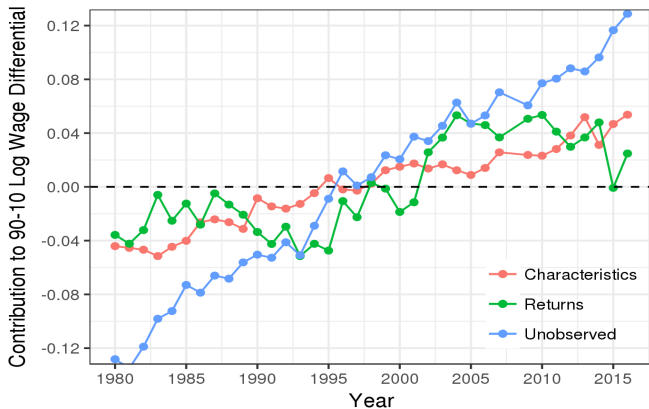
Figure: Components of 90-10 Percentile Log Wage Differential.



Results - Random Forest Prediction

Results - Random Forest Prediction

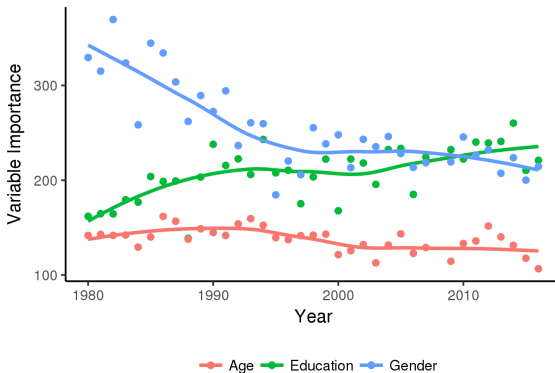
Figure: Components of 90-10 Percentile Log Wage Differential.



Results - Variable Importance

Results - Variable Importance

Figure: Variable Importance in Random Forest Prediction, 1980–2016



Summary

Summary

- Distribution of residuals and thus “unobserved” skill depends largely on prediction method used.

Summary

- Distribution of residuals and thus “unobserved” skill depends largely on prediction method used.
- Role of unobserved factors play a larger role post-1980's and more so in better prediction methods.

Summary

- Distribution of residuals and thus “unobserved” skill depends largely on prediction method used.
- Role of unobserved factors play a larger role post-1980's and more so in better prediction methods.
- Education is more and more important in random forest prediction of wages, while gender's role reduces.

Discussion and Limitations

Discussion and Limitations

Statistical properties?

Discussion and Limitations

Statistical properties?

$$E[\hat{\beta}] = ? = E[\hat{\beta}_t]$$

Discussion and Limitations

Statistical properties?

$$E[\hat{\beta}] = ? = E[\hat{\beta}_t]$$

$$\hat{\beta} \xrightarrow[n \rightarrow \infty]{p} ? \xleftarrow[n \rightarrow \infty]{p} \hat{\beta}_t$$

Discussion and Limitations

Statistical properties?

$$E[\hat{\beta}] = ? = E[\hat{\beta}_t]$$

$$\hat{\beta} \xrightarrow[n \rightarrow \infty]{p} ? \xleftarrow[n \rightarrow \infty]{p} \hat{\beta}_t$$

Why just random forests?

Thank you for listening

Thank you for listening

Any questions?