

The Rise in Returns to Skill?

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

Senan Hogan-H.*

Senior Seminar in Economics, Pomona College[†]

May 2018

Abstract

Abstract goes here.

*This paper was completed in accordance with requirements for the Pomona College Department of Economics Senior Seminar, Spring 2018. I am grateful for the comments and guidance of Michael Kuehlwein, Pomona College Department of Economics.

[†][This project's Github repository, which hosts all contributing materials, is available here.](#)

1 Expanded Proposal

Wage inequality has been documented as rising for many years in the economic literature, yet evidence that attributes this to a rise in return to skill use out-dated and orthodox regression approach to predicting wages. Multiple significant studies decompose wage inequality in order to attribute a large portion of the rise in inequality to result from a rise in returns to skill, including most notably ?.

The statistical analysis of these studies rely heavily on the approach of applying Ordinary Least Squares regression algorithm to the Mincer earnings function, which dates back to some of the first labour economics studies that focus on wage inequality in ? and later ?. The function to be estimated by this approach takes the following form.

$$\ln w_i = \ln w_0 + \rho s_i + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad (1)$$

Here, w_i represents the wage for individual i , x_i years of potential labour market experience, s_i years of education, and w_0 the standard intercept, ρ , β_1 , β_2 standard coefficients to be estimated in the OLS framework with error term ε_i . This model is extremely influential in labour economics to describe and predict inequality in wages in the US population. However, empirical analysis that looks to estimate the relevance of higher returns to skill in explaining rising inequality in the labour market crucially rely on running this regression in separate years noting that the model has significantly less predictive power (and importantly greater error) in later decades of the 20th century.

This research paper will analyse wage inequality and returns to skill by applying newer regression approaches to the same data set (and years since), and comparisons to the orthodox OLS approach described above. The data set used will be the March CPS, available on the BLS website.¹ This is a data set that is representative of the US population, and is available as a cross-sectional data set for individuals separately in every year.

¹This data set is readily available and I am in the processing of accessing in the format needed.

Estimation techniques are OLS regression applied to the Mincer wage equation. Next other regression methods are explored in the context of recent research on the determinants of rising inequality. Firstly, a regression tree design demonstrates a method of predicting wages in the US economy that is organic by design. The method is applied to a sub-sample of the data set in the middle of the 20th century and to early in the 21st century, demonstrating the organic properties of this organic regression technique in order to contrast it with the strict and non-organic approach that has dominant labour economics for many decades.

Lastly, the centre-piece of the analysis will be random forest training and estimation across the entire data-set (in appropriate year increments). This regression technique has many statistical properties which, for the sake of brevity, I will not go in to here. But it is important to note that training of these models may be conducted in a manner that does not over-fit any training sample by using a technique called k-fold validation. The variables selected in the model, of course, will be only those reasoned to be economically significant, taking note of points raised by ?. The variable importance – a very well defined concept in this machine learning approach – will also be documented year on year, going on to assess returns to skill across time by importance of years of education in this model.

1.1 So What?

The Mincer wage has dominated labour economics research and the way of thinking about wage inequality for many decades. However, it is only one model of predicting wages and documenting returns to skill; and importantly it is a model that predicts wages today much worse than it did when first published in a research journal. This analysis expands analysis of wage inequality beyond the simple OLS framework, to newer regression techniques. The newer regression techniques minimise similar error functions to that of the OLS algorithm, yet have crucial difference in that they are formed to fit the data and not according to a rigid framework. Since the 1950s, the US economy has become much larger, much more complicated and much more unequal. Entire regions and industries have stagnated, the

global economy has integrated massively, and the internet has risen to great prominence in every-day life. It is clear that any analysis that looks at wages and wage differences over this time period needs to better acknowledge these issues. This analysis moves away from the rigid framework of labour economics in the 20th century towards regression techniques with no such rigidity, while continuing to acknowledge the role of economic theory – especially as related to variable selection. It is possible that the results will show that returns to skill are estimated to have risen even more than by previous methods, less, or even that another variable shows much greater significance in these newer regression techniques.