Worker's Age, Sex, Education on Average Hourly Earnings

Dong Hyun Yi

August 15, 2016

Worker's Age, Sex, Education on Average Hourly Earnings

Dong Hyun Yi

Abstract:

This paper will investigate the relationship between a Worker's Age and Earnings--with respect to variables such as Sex and Education status. Data have been compiled for the year of 2008 (from a March 2009 survey) from *Handbook of Labor Statistics* published by the United States Bureau of Labor Statistics website. These data will examine full-time workers described by the documentation provided as "employed more than 35 hours per week for at least 48 weeks in the previous year" (see CPS08 Documentation).

Method:

a) A linear regression model will be utilized to account for Age and Earning variables. The model will be used for the provided data to find a linear function between the dependent and independent variables. This model will be used to regress and attempt to fit parameters that best predict the relationship between Worker's Age and Worker's Earnings.

i) Dependent Variable is Average Hourly Earnings (AHE) and will be presented on the Y-axis.

ii) Independent Variables are Age of Worker, Sex of Worker, and Highest Earned Education (High School Diploma vs. Bachelor's Degree) obtained by each Worker.

b) Expectations of this model have been partially predicted by Stock and Watson beforehand: "Generally, older workers have more job experience leading to higher productivity and earnings" (Page 136, Stock and Watson).

i) The first independent variable is Age which as aforementioned, will be expected to have a positive correlation with Average Hourly Earnings.

ii) The second independent variable is Education Status. I expect those with Bachelor's Degree to earn more than those with High School Diplomas.

iii) The third independent variable is Sex of Worker which I expect to have no significant difference between each Sex.

Empirical Exercise 4.1 from *Introduction to Econometrics* by Stock & Watson

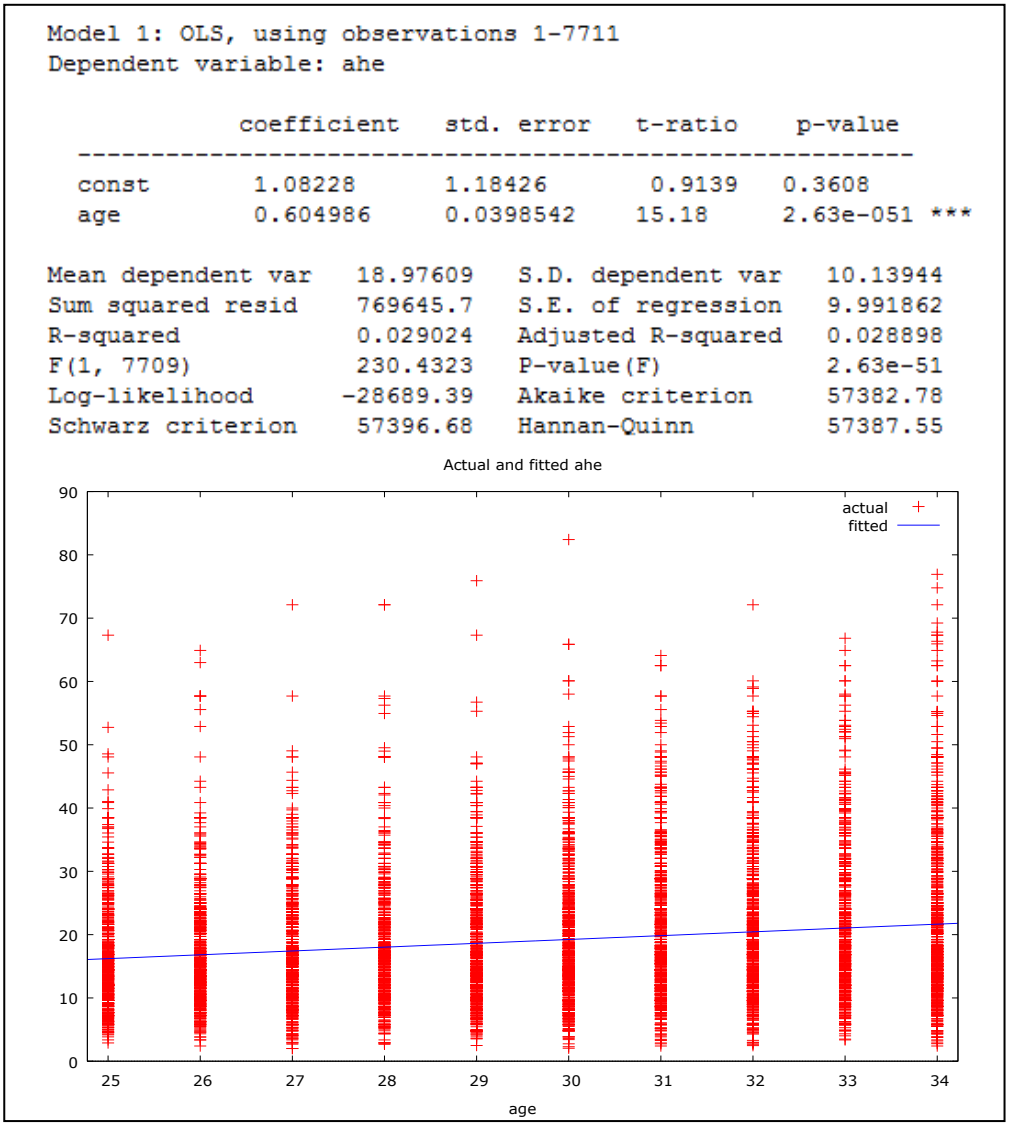
a. Stock and Watson first want us to run a linear regression of average hourly earnings **(AHE)** on age (**Age**) and find the estimated intercept and estimated slope. S&W ask how much do earnings increase as workers age by 1 year. The two variables used will be **AHE** as the **dependent variable** (y) and **Age** as the **independent variable** (x).

GRETL OLS:

```
Model 1: OLS, using observations 1-7711
Dependent variable: ahe

              coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       1.08228      1.18426       0.9139   0.3608
  age         0.604986     0.0398542    15.18     2.63e-051 ***

Mean dependent var   18.97609   S.D. dependent var   10.13944
Sum squared resid    769645.7   S.E. of regression   9.991862
R-squared            0.029024   Adjusted R-squared   0.028898
F(1, 7709)           230.4323   P-value(F)           2.63e-51
Log-likelihood      -28689.39   Akaike criterion     57382.78
Schwarz criterion    57396.68   Hannan-Quinn         57387.55
```



Average Hourly Earnings **=** 1.08228 + 0.604986(***Age)*** **R-squared**= .029024

i) The estimated **intercept** of this linear regression is: **1.08228**

ii) The estimated **slope** for variable "**Age**" = **0.604986**

iii) Workers earn approximately **$0.60** in Average Hourly Earnings **more** per one year increased **Age**.

b. Next Stock and Watson ask us to Predict **Bob's** earnings (**Age 26**) using the above regression as well as **Alexis's** earnings (**Age 30).**

i) For **Bob** we use the given age, **26,** and insert the number in place of X in our equation:

AHE = 1.08228 + 0.604986(**26**)

which gives us estimated **AHE** of 16.811916. **AHE = $16.81** for **Bob**

ii) For **Alexis** we use the given age, **30,** and insert the number in place of X again:

AHE = 1.08228 + 0.604986(**30**)

which gives us estimated **AHE** of 19.23186. **AHE = $19.23** for **Alexis**

c. Next S&W ask if age accounts for a large fraction of the variance in earnings across individuals. (R-squared ranges from **0** and **1**).

The OLS data presents the R-squared value as **0.029024**. This low R-squared value shows that the variable **Age** explains only a minute fraction of the variability across individuals.

Conclusion from Exercise 4.1

The textbook exercise only took into account one of the independent variables, **Age**. Using Gretl I ran a linear regression with the dependent variable **Average Hourly Earnings** and the independent variable **Age**.

AHE*=** 1.08228 + 0.604986Age***

The Y-intercept of **1.08228** was found as well as the slope of the line, .**604986**.

R-squared value = **0.029024**

Evidence shows that the independent variable **Age**, although positively correlated with dependent variable **Average Hourly Earnings**, only explains a small portion of the variability across individuals.

Further Analysis:

a) Summary Statistics of CPS08 Data:

Table 1.

Summary Statistics of CPS08				
Variable	**Mean**	**Median**	**Minimum**	**Maximum**
ahe	18.9761	16.8269	2.00321	82.4176
year	2008.00	2008.00	2008.00	2008.00
bachelor	0.481001	0.000000	0.000000	1.00000
female	0.432629	0.000000	0.000000	1.00000
age	29.5772	30.0000	25.0000	34.0000
Variable	**Std. Dev.**	**C.V.**	**Skewness**	**Ex. kurtosis**
ahe	10.1394	0.534327	1.44071	3.07687
year	0.000000	0.000000	undefined	undefined
bachelor	0.499671	1.03882	0.0760503	-1.99422
female	0.495472	1.14526	0.271965	-1.92603
age	2.85526	0.0965357	-0.0268934	-1.20391
Variable	**5% Perc.**	**95% Perc.**	**IQ range**	**Missing obs.**
ahe	7.05288	38.4615	11.5385	0
year	2008.00	2008.00	0.000000	0
bachelor	0.000000	1.00000	1.00000	0
female	0.000000	1.00000	1.00000	0
age	25.0000	34.0000	5.00000	0

i) **Table 1** presents statistics summarizing the **CPS08** Dataset. From the above we can several observations about the variables.

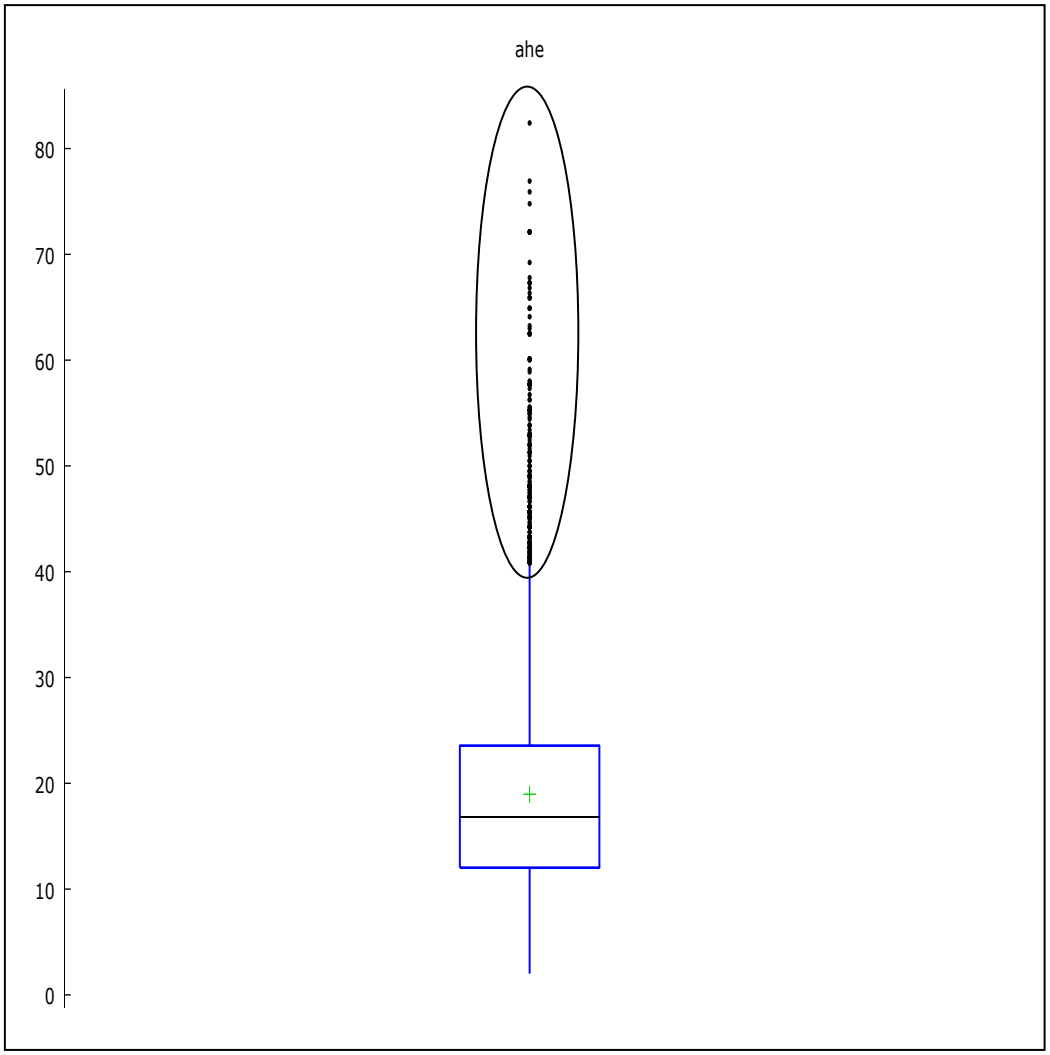
- The **Mean** of the Average Hourly Earnings per Worker is **$18.98**
- The **Median** of the Average Hourly Earnings per Worker is **16.83**
- One **Standard Deviation** is **~$10.14** from the mean
- The **Skewness** is shown as **1.44071**, a positive skewness showing that **a number of individuals makes more than the average**.
- The **Minimum AHE** is **$2.00** and the Maximum **AHE**, $**82.42**

ii) **Table 1** reveals that the dataset has several outliers (data points that exist outside of the interquartile range) shown by the **Standard Deviation** and **Minimum-Maximum** range**.**

Further Analysis:

b) BOXPLOT of **Average Hourly Earnings data column:**

Graph 1.



i) This boxplot of the data column for Average Hourly Earnings shows us that the there are several **outlier** data points (highlighted by the oval). These data points exist outside of the **Interquartile Range**. There is a great degree of variability with a number of workers with **Average Hourly Earning**s past **3σ** (std. dev.) above the mean.

ii) This many outliers that lie above the **interquartile range** makes clear that our data in column **AHE** is positive skewed.

c) Examining **Independent Variable, EDUCATION,** of Worker :

i) Summary Statistics of Workers with **High School Diploma:**

Table 2

	Mean	Median	Minimum	Maximum
ahe	15.332	13.942	2.0513	72.115
bachelor	0.00000	0.00000	0.00000	0.00000
	Std. Dev.	**C.V.**	**Skewness**	**Ex. kurtosis**
ahe	7.7010	0.50229	1.7198	5.5227
bachelor	0.00000	NA	NA	NA
	5% perc.	**95% perc.**	**IQ range**	**Missing obs.**
ahe	5.9538	29.083	8.8580	0
bachelor	0.00000	0.00000	0.00000	0

Table 2 shows Workers with a **High School Diploma** having:

- **a mean** Average Hourly Earnings **of $15.33**
- **a median** Average Hourly Earnings **of $13.942**
- **a standard deviation of $7.70**

ii) Summary Statistics of Workers with **Bachelor's Degree**

Table 3

	Mean	Median	Minimum	Maximum
ahe	22.908	20.673	2.0032	82.418
bachelor	1.0000	1.0000	1.0000	1.0000
	Std. Dev.	**C.V.**	**Skewness**	**Ex. kurtosis**
ahe	10.953	0.47814	1.1803	2.0381
bachelor	0.00000	0.00000	NA	NA
	5% perc.	**95% perc.**	**IQ range**	**Missing obs.**
ahe	9.2308	43.269	13.462	0
bachelor	1.0000	1.0000	0.00000	0

Table 3 shows Workers with a **Bachelor's Degree** having:

- **a mean** Average Hourly Earnings of **$22.91**
- **a median** Average Hourly Earnings of **$20.673**
- **a standard deviation of $10.95**

Further analysis Examining Independent Variable, EDUCATION, of Worker :

iii) The Summary Statistics reveal a discrepancy between the mean **AHE** between workers that have a bachelor's degree and those who do not. Those with a **Bachelor's** Degree make on average **$7.57** than those with a **High School Diploma.**

iii) Furthermore **Table 2** and **Table 3** reveal a larger standard deviation of $7.70 among those with high school diplomas compared to $10.95 among those with a Bachelor's Degree. This shows a greater amount scatter of data points among the latter that may allow us to postulate that those with Bachelor's Degrees tend to have more variability in Average Hourly Earnings.

d) Examining **Independent Variable, SEX,** of Worker :

i) Summary Statistics of **MALE** Workers**:**

Table 4

	Mean	Median	Minimum	Maximum
ahe	20.114	17.857	2.0032	82.418
female	0.00000	0.00000	0.00000	0.00000
	Std. Dev.	**C.V.**	**Skewness**	**Ex. kurtosis**
ahe	10.677	0.53082	1.3221	2.5147
female	0.00000	NA	NA	NA
	5% perc.	**95% perc.**	**IQ range**	**Missing obs.**
ahe	7.2115	40.865	12.500	0
female	0.00000	0.00000	0.00000	0

Table 4 shows **MALE** Workers having:

- **a mean** Average Hourly Earnings **of $20.11**
- **a median** Average Hourly Earnings **of $17.86**
- **a standard deviation of $10.68**

ii) Summary Statistics of **FEMALE** Workers:

Table 5

	Mean	Median	Minimum	Maximum
ahe	17.484	15.570	2.0055	74.786
female	1.0000	1.0000	1.0000	1.0000
	Std. Dev.	**C.V.**	**Skewness**	**Ex. kurtosis**
ahe	9.1782	0.52495	1.5906	4.0397
female	0.00000	0.00000	NA	NA
	5% perc.	**95% perc.**	**IQ range**	**Missing obs.**
ahe	6.5852	35.096	10.530	0
female	1.0000	1.0000	0.00000	0

Table 5 shows **FEMALE** Workers having:

- **a mean** Average Hourly Earnings **of $17.48**
- **a median** Average Hourly Earnings **of $15.57**
- **a standard deviation of 9.1782**

Further Analysis Examining Independent Variable, SEX, of Worker :

iii) The Summary Statistics reveal a discrepancy between the mean **AHE** between **MALE** workers **FEMALE** workers. Male workers' **mean AHE** is greater than the Female workers' by **$2.63**. Though not as significant as the difference in the mean AHE among differing Education statuses, **$2.63** is roughly **15.04%** of the average **FEMALE** worker's Average Hourly Earnings.

iii) Furthermore **Table 4** and **Table 5** reveal similar **AHE** standard deviations of **$10.68** among Males and **$9.18** among Females. This allows us to postulate that there is similar amounts of scatter among the data points in both sets.

Conclusion:

Through Stock and Watson's exercise questions I was able to determine that Age is positively correlated with Average Hourly Earnings--though the R-squared value showed high variability amongst the data points.

My hypothesis that workers with Bachelor's Degrees would have a higher AHE than those with High School Diplomas was supported by the Summary Statistics in Table 2 and 3.

Male and female workers also had a discrepancy in their earnings, with Males making approximately 15% more than their female counterparts.

Through the linear regression model we were able to more closely examine each variable and was able to interpret the dataset with more precision. As Stock and Watson predicted there was a small but positive correlation between Age and Average Hourly Earnings of Workers.

Appendix

FEMALE: **1 if female;**

0 if male

YEAR: Data collected in 2008

AHE : Average Hourly Earnings

BACHELOR: **1** if worker has a bachelor's degree;

0 if worker has a high school degree