

Assignment 1: Predicting Hourly wage of Financial Managers using Linear Regressions

Talgat Ilimbek uulu

Motivation

The goal of this assignment is to predict hourly wage of financial managers in US market. The starting point of our study will be a human capital theory. In labor economics, human capital is generally referred to the stock of skills and characteristics possessed by workers that increase their productivity. Human capital research has traditionally used educational attainment as a way to measure human capital. The work by Mincer (1974) proposes a well-known single-equation model that establishes a link between wages, schooling, and experience in the labor market.

$$\ln w = f(s, x) = \beta_0 + \beta_1 s + \beta_2 x^2$$

where w is wage, s is a years of schooling and x is a potential experience in a labor market.

The notable advantage of the Mincer equation is that it has a robust theoretical foundations. Mincer equation will serve as a basis upon which I plan to develop and refine additional three models to explore and analyze the data.

Data

In the sample, we have 1399 individuals in the “financial managers” occupation, of whom 71% hold a college degree. The dependent variable is the natural logarithm of earnings per hour, calculated as the ratio of weekly earnings to usual hours worked. Regarding the education variable, the data employs a credential-oriented measure. To align with the model, I transformed the variable into years of schooling, using standard duration for the obtained degrees. For example, a PhD degree is considered equivalent to 22 years of schooling, MA degree to 18 years of schooling, etc. Years of potential experience are computed as the difference between current age and the age of potential entry into the labor market. This calculation assumes

that individuals without a high school degree enter the labor force at age 17, those with a high school degree at 19, those with some college at 21, and those with a college degree at 23. see Borjas (2003)

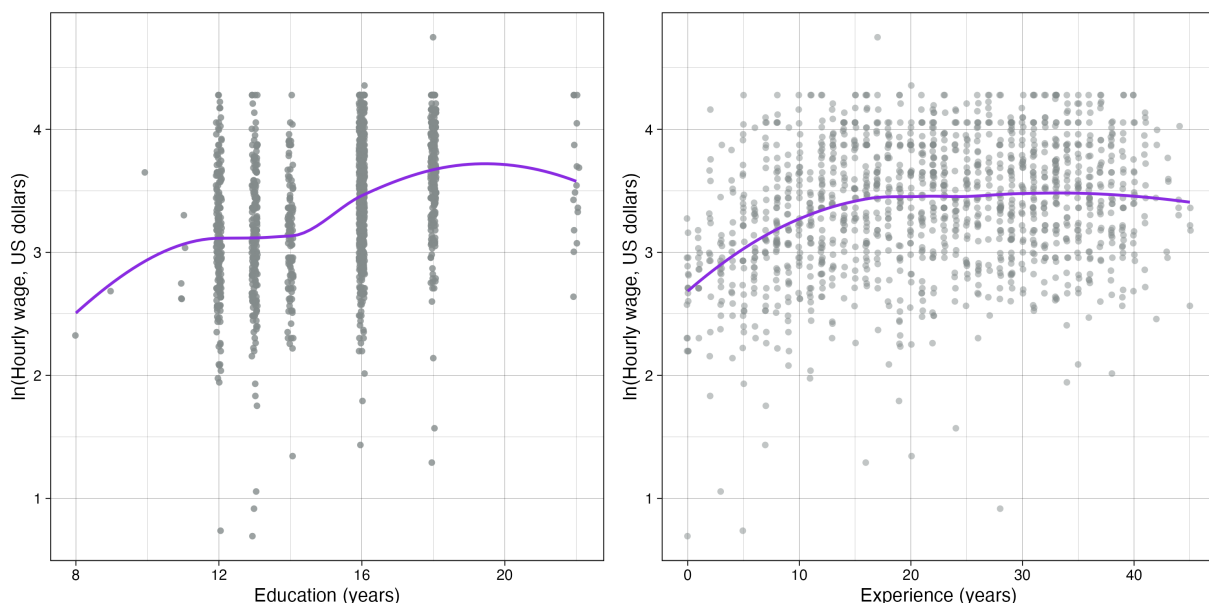


Figure 1: Log hourly wage, education and experience

To check whether functional form suggested by theory is consistent with the data, I estimate a lowess regression and plot together with the scatterplot (see Figure 1). According to the pattern uncovered by lowess, average earnings increase with each additional year of schooling and experience shows the concave pattern. The standard regression table of the model is presented in Table 1.

Table 1: Mincer equation OLS estimates

	Ln wage
Education	.096*** (.006)
Experience	.041*** (.004)
Experience squared	-.0006*** (.0001)
Constant	1.386*** (.10)
Number of observations	1399
Adjusted R-squared	0.25
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Robust standard errors in parenthesis	

The OLS regression of log hourly wage to year of schooling and experience implies that 1 year of additional schooling is associated with 9.6% increase in monthly wage for financial man-

agers. Both figures and regression give evidence supporting the argument that education and experience may play an important role in the wage determination. Note, however, all correlations should be interpreted with caution and shouldn't be given causal interpretation. The positive correlation between wage and education can reflect unobserved family and individual heterogeneity.

Four regression models

I have specified four linear regression models for predicting hourly wage of financial managers. The Model 1 is our baseline Mincer equation capturing established factors influencing wages. Building upon this foundation, I introduce three subsequent models, each incorporating additional variables used in the literature to enhance the predictive capacity (see Appendix for additional table and graphs)

Table 2: Wage baseline regression

Model No.	Used variables
Model 1	education, experience, experience squared
Model 2	Model 1 + male, married, black, no children, native born
Model 3	Model 2 + dummies for certain states and industries
Model 4	Model 3 + interaction terms

Model 2 augments the baseline Mincer equation by introducing demographic features such as gender, race, and marital status. These variables are crucial predictors, capturing disparities like well documented gender-based wage gaps and the influence of race and marital status on career opportunities. Model 3 enhances the Mincer equation by including data on the highest and lowest paying states and industries. This addition is crucial as it captures geographic and industry-specific factors influencing wage variations. Model 4 adds interaction terms to the previous model. The results are presented at Table 3.

Table 3: Wage models and measures of fit

Model	N predictors	R-squared	Training RMSE	BIC
Model 1	3	0.25	0.456	1,810.29
Model 2	8	0.30	0.442	1,760.30
Model 3	12	0.31	0.437	1,757.15
Model 4	15	0.32	0.436	1,770.64

Referring to Table 3, both Model 3 - 4 have comparable RMSE, however the Model 3 has the lowest BIC (1757.15) which suggests that it has better prediction properties compared to other models. Notice that once we increase the complexity of the model by increasing the number of predictors, R-squared increases as well, which is expected. Potentially, we could keep adding more predictors to the model and increase the R-squared, however, this would not necessarily improve the prediction properties of the model and could lead to overfitting. It is suggested to use BIC as a criterion for model selection, which penalizes the number of parameters in the model.

Table 4: Wage models estimated and evaluated using 4-fold cross-validation and RMSE

Fold No.	Model 1	Model 2	Model 3	Model 4
Fold1	0.434	0.413	0.412	0.411
Fold2	0.459	0.450	0.446	0.446
Fold3	0.473	0.460	0.460	0.458
Fold4	0.465	0.459	0.450	0.452
Average	0.458	0.446	0.442	0.442

Referring to Table 4, both Model 3 and Model 4 have the lowest average RMSE. This finding aligns with previous studies where both BIC and cross-validation often converge on the same optimal model. Given that Model 3 not only has the lowest BIC and average RMSE but is also less complex compared to Model 4, I choose it as the best model. The table below shows the point prediction of our model

Table 5: Point and interval prediction for specific individual using Model 3

	Model in logs	Recalculated to level
Predicted	3.67	43.36
PI_low 80%	2.94	20.88
PI_high 80%	4.40	90.04

For this specific individual, the hourly wage is estimated to range from 21 to 90 dollars with an 80% probability. As anticipated, this is a broad prediction interval. Nevertheless, the lower limit of 21 dollars can serve as a benchmark for this specific individual planning to work as financial manager, enabling him to make informed adjustments to his budgetary considerations.

References

- Borjas, George J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *The Quarterly Journal of Economics* 118 (4): 1335–74. <http://www.jstor.org/stable/25053941>.
- Mincer, Jacob A. 1974. "The Human Capital Earnings Function." In *Schooling, Experience, and Earnings*, 83–96. NBER Chapters. National Bureau of Economic Research, Inc. <https://ideas.repec.org/h/nbr/nberch/1767.html>.

Appendix

Table 6: Dependent variable Ln(Hourly wage), OLS Estimates.

	(1)	(2)	(3)	(4)
Education	0.097*** (0.006)	0.08*** (0.006)	0.08*** (0.006)	0.08*** (0.006)
Experience	0.04*** (0.004)	0.037*** (0.005)	0.036*** (0.005)	0.036*** (0.005)
Experience squared	-0.0006*** (0.000)	-0.0005*** (0.000)	-0.0005*** (0.000)	-0.0005*** (0.000)
Constant	1.39*** (0.09)	1.52*** (0.10)	1.52*** (0.10)	1.53*** (0.11)
Dummies				
Demographic	No	Yes	Yes	Yes
Regional	No	No	Yes	Yes
Interaction terms	No	No	No	Yes
Industry	No	No	No	Yes
Number of observations	1399	1399	1399	1399
Adj. R-squared	0.25	0.29	0.30	0.31

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ Robust standard errors in parenthesis.

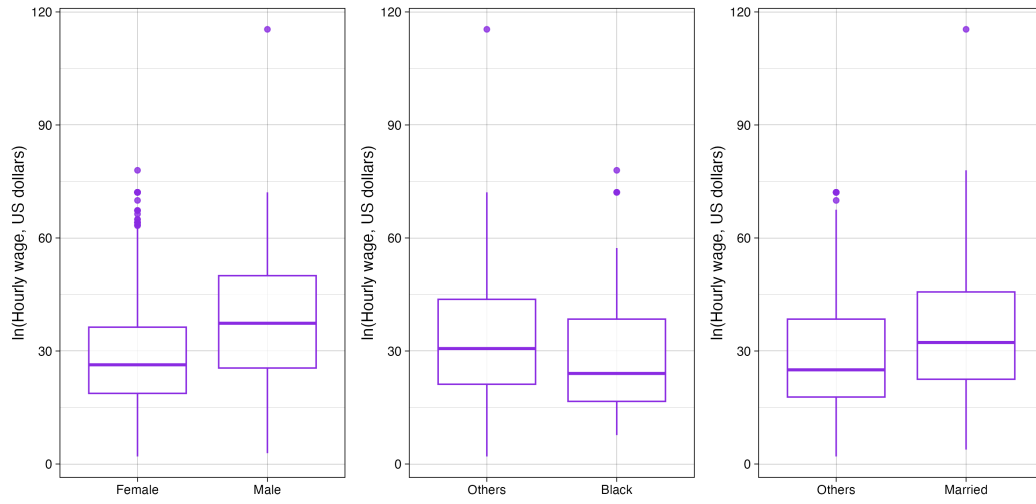


Figure 2: Hourly wage distribution by important demographic features

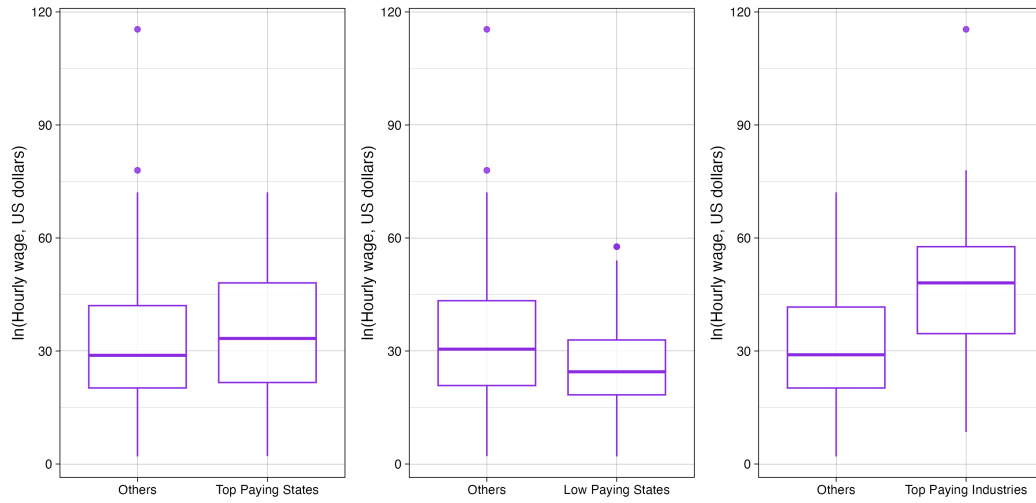


Figure 3: Hourly wage distribution by important regional and industry features