Aizhan Borubaeva
Prediction with Machine Learning for Economists
2023/24 Fall
Assignment 1

Many people are interested in issues that can have some correlation with their salaries. As a young person with high opportunity costs of potential investment of time, money and effort I also thought about it. Moreover, as a student of the Economic Policy Department, I am also interested in this issue from the view of potential policy implications to prevent any potential discrimination in the job market. In the regression analysis below, I will check some factors that can influence wages.

The data for this analysis is taken from the Current Population Survey made by the US government monthly to monitor the labor market. There is a dataset which contains data collected from approximately 30,000 people (Feenberg, Daniel, and Jean Roth, 2007).

The analysis is based on the occupational code "10 - Chief executives" because this is the top level of management that can show how the given factors can influence the wage of the person after years of a successful career and hard work. Chief executives can be considered the most successful people in their spheres and companies, so their salary can be considered a long-run output taking all other factors, such as hard work, soft skills, ongoing training and other non-initial factors equal.

I used the log-level model where ln_earnings_per_hour is a dependent variable, while «female», «age», «Masters», «PhD», "Married», No_children" are independent variables. ln_earnings_per_hour – is calculated as a natural logarithm of the weekly wage divided by the number of hours per week, so I calculated the wage per hour first. «Female» – is a dummy variable, which identifies if the gender of the respondent is female. Age is a numeric variable of the age of respondents. «Masters» and «PhD" are the dummy variables showing if the final degree of the respondent is at the master's or PhD level. Thus, for «PhD» it is 1 when a respondent has a PhD level and 0 in any other case, «Master» is 1 when the respondent has a master's degree and 0 otherwise. It should be noted here that for people with a PhD, the «Master" variable is equal to 0. "Married" is a dummy variable which is equal to 1 when a person is married nowadays, and 0 if he or she is single, divorced or widowed. "No_children" is a dummy which is 1 when a person does not have children and 0 otherwise.

I chose these predictors because there is a theory that gender influences earnings[1]. Age can be a good predictor because older age on average means more years of experience, while the level of education should demonstrate the level of knowledge and expertise. Such predictors as marital status and the presence of children can be good predictors because people with families usually spend more money, especially for funding their children and because of the theories that married men get higher salaries[2] and there is a child penalty (R. Bartlett et al.1984).

The descriptive statistics are listed in Table 1:

Table 1. Descriptive statistics

	earnwke	uhours	ln_earnings_per_hour	female	age	Masters	PhD	Married	No_children
count	1274.000000	1274.000000	1274.000000	1274.000000	1274.000000	1274.000000	1274.000000	1274.000000	1274.000000
mean	2013.636028	47.460754	3.642047	0.279435	49.047096	0.293564	0.030612	0.821036	0.563579
std	815.175259	9.654216	0.577918	0.448898	9.244313	0.455573	0.172332	0.383473	0.496136
min	1.000000	5.000000	-3.036554	0.000000	22.000000	0.000000	0.000000	0.000000	0.000000
25%	1346.000000	40.000000	3.372275	0.000000	43.000000	0.000000	0.000000	1.000000	0.000000
50%	2086.460000	45.000000	3.781341	0.000000	50.000000	0.000000	0.000000	1.000000	1.000000
75%	2884.610000	50.000000	4.055122	1.000000	56.000000	1.000000	0.000000	1.000000	1.000000
max	2884.610000	99.000000	6.175386	1.000000	64.000000	1.000000	1.000000	1.000000	1.000000

[1] UNDP (United Nations Development Programme). 2023. 2023 Gender Social Norms Index (GSNI): Breaking down gender biases: Shifting social norms towards gender equality. New York.
[2] R. Bartlett et al. "Wage determination and marital status: another look". Industrial Relations. 1984

I considered the following fixed-effect OLS models which are listed in Table 2:

Table 2. Models

Models	Dependent variable	Predictors
Model_1	ln_earnings_per_hour	female
Model_2	ln_earnings_per_hour	female, age
Model_3	ln_earnings_per_hour	female, age, Masters, PhD
Model_4	ln_earnings_per_hour	female, age, Masters, PhD, Married, No_children

The results of the regressions are the following:

Table 3. Results

Models	Model_1	Model_2	Model_3	Model_4
Intercept	3.692	3.483	3.425	3.365
female	-0.178	-0.173	-0.155	-0.138
age	-	0.004	0.004	0.006
Masters	-	-	0.217	0.209
PhD	-	-	0.160	0.149
Married	-	-	-	0.022
No_children	-	-	-	-0.113
R-squared	0.018	0.021	0.05	0.057
RMSE	0.6669	0.6667	0.6591	0.6369
BIC Scores	-167.7366	-167.6177	-173.0480	-177.9539

The table above shows that according to R-squared the best models are the third (0.05) and the fourth one (0.057) because they have the highest R-squared. According to RMSE, the best models are again the third (0.659) and the fourth (0.637) ones with the lowest cross-validated RMSE. BIC Scores demonstrate the same result with the lowest values (-173 and -178). However, according to all three tests, the fourth model is the best one.

Figure 1. Relationship between Model Complexity and RMSE

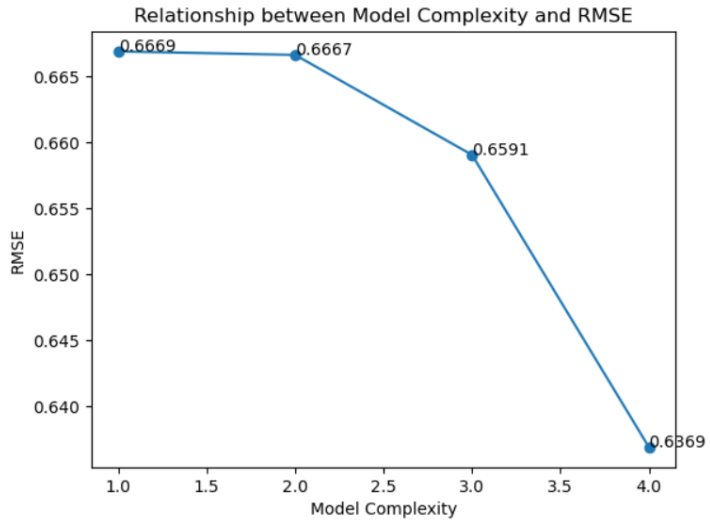

Figure 1 also shows that the more complex the model the lower RMSE, so the better the model.

Therefore, based on all information received there is a conclusion that the fourth model with the maximum number of predictors is the best one. However, I would suggest the following adding of relevant predictor to check if there is an even better model.

Bibliography

Feenberg, Daniel, and Jean Roth. "CPS Labor Extracts 1979 - 2006." NBER, January 2007. http://www.nber.org/data/morg.html.

UNDP (United Nations Development Programme). 2023. 2023 Gender Social Norms Index (GSNI): Breaking down gender biases: Shifting social norms towards gender equality. New York.

R. Bartlett et al. “Wage determination and marital status: another look”. Industrial Relations. 1984

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2019. "Child Penalties across Countries: Evidence and Explanations." AEA Papers and Proceedings, 109: 122-26.