Econ 3040 - Assignment 3: Wages and education

Ryan T. Godwin

Due date: November 12th, 2022. Worth 3% of your final grade.
Instructions: Submit your assignment in the "Assignment 3" drop box on UM Learn. Include your name and student number. Do not copy and paste output from `R`. Format your results nicely. Submit the `R` code that you used for each question in your assignment.

What are the returns to education? In this assignment, you will estimate a multiple regression model, and comment on omitted variable bias (OVB). The data is from the National Longitudinal Survey of Young Men (NLSYM). National longitudinal surveys, from the U.S. Bureau of Labor Statistics (), are famous data sources for economists. The particular data set for this assignment uses some NLSYM data from 1976. Load the data using:

```
school <- read.csv("https://rtgodwin.com/data/school.csv")
```

Table 1: Description of the variables in the "school" data set.

wage	hourly wage measured in (1976) cents
ed	number of years of education
exp	number of years of work experience
iqscore	worker's score on a standard IQ test
black	a dummy variable indicating if the worker is African-American
sinmom14	a dummy variable equal to 1 if the worker's mother was single when the worker was age 14
momed	number of years of education of the worker's mom
daded	number of years of education of the worker's dad

1. Using least squares, estimate the model: $wage = \beta_0 + \beta_1 ed + \epsilon$. What are the estimated returns to education?
2. Explain why it is a bad idea to estimate a model with *only* education as an explanatory variable? What is the problem with leaving the other variables out of the model?
3. Estimate a model with "wage" as the dependent variable, and all other variables as explanatory variables (use all the other variables as regressors). Explain the differences between the estimated results of this new model, and the model that you estimated in Question 1.
4. Use the model from Question 3 to calculate the predicted wage for a *representative* case (choose values for all of the "x" variables and calculate a prediction).
5. Three of the variables from the model in Question 3 appear to be statistically insignificant (according to the t-statistics and p-values reported in the `summary()` output). Use an F-test to see if these 3 variables may be dropped from the model.
6. What are the estimated returns to education?