Econ 21020 - Problem Set 3

Due 11/3 at 11:59 PM. Submit to Canvas. May be completed in groups of up to 6 students. Only one submission is required per group. Note that this assignment will be graded for COMPLETION and not for accuracy.

Problem 1

We learned that one interpretation of the linear regression equation is that the coefficients β_0 and β_1 determine the function of X that is the best linear approximation to $E[Y|X]$. However, "best" need not always imply "good." Consider $X \sim N(0,1)$ and $Y = X^2$. Then, $E[Y|X] = X^2$.

(a) We said that under the (equivalent) best linear predictor of Y given X interpretation, β_0 and β_1 will satisfy the following first order conditions:

$$E[Y - \beta_0 - \beta_1 X] = 0$$
$$E[X(Y - \beta_0 - \beta_1 X)] = 0$$

Solve this system of equations for β_0 and β_1. It may be useful to note that for $X \sim N(0,1)$, $E[X^2] = 1$ and $E[X^3] = 0$.

(b) Draw pictures of the best linear approximation to $E[Y|X]$, $\beta_0 + \beta_1 X$, and the actual $E[Y|X]$ on the same graph. The graph need not be extremely precise - it just needs to capture the major features of the functions. Interpret the result in light of the best linear approximation interpretation.

(c) Say that our econometrician has a theory about the functional form of the conditional expectation of Y, and runs the regression

$$\sqrt{Y} = \beta_0 + \beta_1 X$$

on the interval $Y \geq 0$ instead. How will the performance of the linear approximation to the conditional expectation for this regression compare to that of the original estimating equation? There is no need to show any math for this part if you do not feel that you need to.

Problem 2

Say we are interested in studying the effect of sentencing on recidivism of juvenile offenders. Consider the all causes model:

$$Y = g(X, U)$$

where $Y = 1$ if the offender committed another crime and $Y = 0$ otherwise. Let $X = 1$ if the offender's sentence included prison time and $X = 0$ if otherwise. U has the typical meaning it would take if we are claiming that g is a causal model. We have observational data on Y and X.

(a) Give two examples of unobserved determinants you think would be a part of U.

(b) Under what assumptions would we be able to interpret the results of the regression model

$$Y = \beta_0 + \beta_1 X + U$$

causally? Does that assumption(s) seem plausible in this context? Why or why not?

(c) Define and interpret Y_1 and Y_0 as potential outcomes, in the sense discussed in class.

(d) Define and interpret the Average Treatment Effect (ATE) between $X = 1$ and $X = 0$. Based on your response to part (b), do you think we can recover this ATE using a linear regression?

Problem 3

In class, we made extensive use of the fact that our OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_0$ satisfy:

$$\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

without fully justifying this. Let's walk through a demonstration that this will be the case. Assume that this minimization problem can be solved by taking first-order conditions (the second-order condition is easy to check, if desired). Then, our estimates will satisfy the first-order conditions:

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{1}$$

$$\frac{1}{n} \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{2}$$

(a) Solve FOC (1) for $\hat{\beta}_0$. This should yield the OLS estimator for β_0, $\hat{\beta}_0 = \overline{Y}_n - \hat{\beta}_1 \overline{X}_n$

(b) Plug the result from part (a) into FOC (2). Solve for $\hat{\beta}_1$. This should yield the OLS estimator for β_1, $\hat{\beta}_1 = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}^2_{X,n}}$

Problem 4

Show that the R^2 from a regression of Y on X is equal to the R^2 from a regression of X on Y. Provide some intuition for this result.

Problem 5

Say we are comfortable assuming homoskedasticity of U. Let's call

$$\hat{\sigma}^2_{NR} = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2}{\hat{\sigma}^2_X}$$

the non-heteroskedasticity robust estimator of σ_1^2 (the variance of the limiting distribution of $\sqrt{n}(\hat{\beta}_1 - \beta_1)$). Demonstrate that this estimator is consistent for σ_1, assuming homoskedasticity, our normal maintained assumptions, and $E[Y^4], E[X^4] < \infty$. (Hint 1: I would deal with the numerator on its own first and bring the denominator back in later.) (Hint 2: Remember that $\hat{\beta}_0$, $\hat{\beta}_1$, and sample variances and covariances are consistent under the assumptions that have been made.)

Problem 6

This exercises uses observational (non-experimental) data on the years of schooling and the (log) weekly wage of 329,509 observations of American men born between 1930-1939, as was used in the paper Angrist and Krueger (199). This data is available on Canvas, under Modules PSet Data, as "ak91.csv". We will consider the variables "education", which represents the years of education completed by the man and "log_weekly_wage" which represents the log weekly wage. Consider these variables to be an iid sample $(Y_i, X_i) \sim (Y, X)$ where Y is years of education and X is log wage of American men born between those years. Assume that $E[X^4], E[Y^4] < \infty$.

(a) Which of the three interpretations of linear regression from class do you think would be most appropriate for the regression equation:

$$Y = \beta_0 + \beta_1 X + U$$

where Y and X are defined as above. Why?

(b) Perform the indicated regression using the data, without heteroskedasticity-robust standard errors. Interpret the coefficients you get in light of your response to (a).

(c) Re-run the regression with heteroskedasticity-robust standard errors. Did the coefficients and/or standard errors change? Why?

(d) In our setting, which type of standard errors would you consider appropriate? Why?