STA 302f13 Assignment Twelve[1]

These questions are practice for the final exam, and are not to be handed in. Material like this may or may not be on the final.

This assignment explores the *consistency* of estimated regression coefficients when independent variables are missing from the model or measured with error. Roughly speaking, an estimator is consistent if it converges to the parameter it's estimating as the sample size $n \to \infty$. This kind of large-sample accuracy is pretty much the least you can ask.

1. In the following regression model, the independent variables X_1 and X_2 are random variables. The true model is
$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i,$$
independently for $i = 1, \ldots, n$, where $\epsilon_i \sim N(0, \sigma^2)$.

   The mean and covariance matrix of the independent variables are given by
$$E\left(\begin{array}{c} X_{i,1} \\ X_{i,2} \end{array}\right) = \left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right) \quad \text{and} \quad cov\left(\begin{array}{c} X_{i,1} \\ X_{i,2} \end{array}\right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{array}\right)$$
   Unfortunately $X_{i,2}$, which has an impact on Y_i and is correlated with $X_{i,1}$, is not part of the data set. Since $X_{i,2}$ is not observed, it is absorbed by the intercept and error term, as follows.
$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i \\ &= (\beta_0 + \beta_2\mu_2) + \beta_1 X_{i,1} + (\beta_2 X_{i,2} - \beta_2\mu_2 + \epsilon_i) \\ &= \beta_0' + \beta_1 X_{i,1} + \epsilon_i'. \end{aligned}$$
   The primes just denote a new β_0 and a new ϵ_i. It was necessary to add and subtract $\beta_2\mu_2$ in order to obtain $E(\epsilon_i') = 0$. And of course there could be more than one omitted variable. They would all get swallowed by the intercept and error term, the garbage bins of regression analysis.

   (a) What is $Cov(X_{i,1}, \epsilon_i')$?

   (b) Calculate the variance-covariance matrix of $(X_{i,1}, Y_i)$ under the true model. Is it possible to have non-zero covariance between $X_{i,1}$ and Y_i when $\beta_1 = 0$?

   (c) Suppose we want to estimate β_1. The usual least squares estimator is
$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (X_{i,1} - \overline{X}_1)(Y_i - \overline{Y})}{\sum_{i=1}^n (X_{i,1} - \overline{X}_1)^2}.$$
   You may just use this formula; you don't have to derive it. You may also use the fact that like sample means, sample variances and covariances converge to the corresponding Greek-letter versions as $n \to \infty$ (except possibly on a set of probability zero) like ordinary limits, and all the usual rules of limits apply. So for example, defining $\widehat{\sigma}_{xy}$ as $\frac{1}{n-1}\sum_{i=1}^n (X_{i,1} - \overline{X}_1)(Y_i - \overline{Y})$, we have $\widehat{\sigma}_{xy} \to Cov(X_i, Y_i)$.

   So finally, here is the question. As $n \to \infty$, does $\widehat{\beta}_1 \to \beta_1$? Show your work.

[1] Copyright information is at the end of the last page.

2. Consider simple regression through the origin in which the independent variable values are random variables rather than fixed constants. In addition, the independent variable values cannot be observed directly. Instead, we observe X_i plus a piece of random noise. The model is this: Independently for $i = 1, \ldots, n$, let

$$\begin{aligned} Y_i &= X_i\beta + \epsilon_i \qquad (1) \\ W_i &= X_i + e_i, \end{aligned}$$

where

- X_i has expected value μ and variance σ^2_x,
- ϵ_i has expected value 0 and variance σ^2_ϵ, and
- e_i has expected value 0 and variance σ^2_e
- X_i, e_i and and ϵ_i are all independent.

Again, the X_i values are unavailable. All we can see are the pairs (W_i, Y_i) for $i = 1, \ldots, n$.

(a) Following common practice, we ignore the measurement error and apply the usual regression estimator with W_i in place of X_i. The parameter β is estimated by

$$\widehat{\beta}_{(1)} = \frac{\sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i^2}$$

Does $\widehat{\beta}_{(1)} \rightarrow \beta$? Answer Yes or No and show your work.

(b) Now consider instead the estimator

$$\widehat{\beta}_{(2)} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n W_i}.$$

Does $\widehat{\beta}_{(2)} \rightarrow \beta$? Answer Yes or No and show your work.
