FINM3123 Introduction to Econometrics

Chapter 5 Exercises

1. In the simple regression model under MLR.1 through MLR.4, we argued that the slop estimator, $\hat{\beta}_1$, is consistent for β_1. Using $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$, show that plim$\hat{\beta}_0 = \beta_0$. [You need to use the consistency of $\hat{\beta}_1$ and the law of large numbers, along with the fact that if $\text{plim}(T_n) = \alpha$ and $\text{plim}(U_n) = \beta$, then $\text{plim}(T_n + U_n) = \alpha + \beta$ and $\text{plim}(T_n U_n) = \alpha\beta$.]

2. Use the data WAGE1.RData (or WAGE1.xls) for this exercise.
   (a) Estimate the equation
   $$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$
   Save the residuals and plot the histogram.
   (b) Repeat part (a), but with $\log(wage)$ as the dependent variable.
   (c) Would you say that Assumption MLR.6 is closer to being satisfied for the level-level model or the log-level model?

3. Several statistics are commonly used to detect nonnormality in underlying population distribution. Here we will study one that measures the amount of skewness in a distribution. Recall that any normally distributed random variable is symmetric about its mean; therefore, if we standardize a symmetrically distributed random variable, say $z = (y - \mu_y)/\sigma_y$, where $\mu_y = E(y)$ and $\sigma_y = sd(y)$, then z has mean zero, variance one, and $E(z^3) = 0$. Given a sample of data $\{y_i : i = 1, ..., n\}$, we can standardize y_i in the sample by using $z_i = (y_i - \hat{\mu}_y)/\hat{\sigma}_y$, where $\hat{\mu}_y$ is the sample mean and $\hat{\sigma}_y$ is the sample standard deviation. (We ignore the fact that these are estimates based on the sample.) A sample statistic that measures skewness is $n^{-1}\sum_{i=1}^{n} z_i^3$, or where n is replaced with $(n-1)$ as a degrees-of-freedom adjustment. If y has a normal distribution in the population, the skewness measure in the sample for the standardized values should not differ significantly from zero.
   (a) First use the data set 401KSUBS.RData (or 401KSUBS.xls), keeping only observations with $fsize = 1$. Find the skewness measure for inc. Do the same for $log(inc)$. Which variable has more skewness and therefore seems less likely to be normally distributed?
   (b) Next use BWGHT2.RData (or BWGHT2.xls). Find the skewness measures for $bwght$ and $log(bwght)$. What do you conclude?
   (c) Evaluate the following statement: "The logarithmic transformation always makes a positive variable look more normally distributed."
   (d) If we are interested in the normality assumption in the context of regression, should we be evaluating the unconditional distribution of y and $log(y)$? Explain.