ECON 704
Problem Set 1

All random variables in this problem set are scalar unless otherwise stated.

1. Suppose we observe $\{X_i\}_{i=1}^n$. Find the probability limit of:

   a. $\frac{1}{n}\sum_{i=1}^n X_i^2$

   b. $\frac{1}{n}\sum_{i=1}^n \ln(X_i)$

   c. $\frac{\sum_{i=1}^n \ln(X_i)}{\sum_{i=1}^n X_i^2}$

2. Suppose we observe $\{X_i\}_{i=1}^n$. Using the Central Limit Theorem and Slutsky's Theorem, find the asymptotic distribution of :

   a. $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(X_i^2 - \mathbb{E}[X^2]\right)$

   b. $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\ln(X_i) - \mathbb{E}[\ln(X)]\right)$

   c. $\sqrt{n}\cdot\frac{\sum_{i=1}^n (\ln(X_i) - \mathbb{E}[\ln(X)])}{\sum_{i=1}^n X_i^2}$

3. Suppose we observe $\{(X_{i1}, Y_{i1}, X_{i2}, Y_{i2})\}_{i=1}^n$ which are i.i.d. across i. Consider the model

$$Y_{it} = X_{it}\beta + u_i + \varepsilon_{ot}$$

   where X_{it} is strictly exogenous with respect to ε_{it}.

   (a) Express $\dot{X}_{i1}$ and $\dot{X}_{i2}$ in terms of X_{i1} and X_{i2}. Do the same for $\dot{Y}_{i1}$ and $\dot{Y}_{i2}$.

   (b) Show that the fixed effects estimator can be written as:

$$\hat{\beta}^{\text{FE}} = \frac{\sum_{i=1}^n (Y_{i1} - Y_{i2})(X_{i1} - X_{i2})}{\sum_{i=1}^n (X_{i1} - X_{i2})^2}$$

   (c) Find the probability limit of $\hat{\beta}^{\text{FE}}$.

   (d) Find the asymptotic distribution of $\sqrt{n}(\hat{\beta}^{\text{FE}} - \beta)$.

   (e) How would you test the hypothesis that $\beta = 0$?

   Hint: For (c) and (d), apply Steps 1 to 4 of the appropriate proofs in the lecture slides.

4. (Wooldridge, Introductory Econometrics Chapter 14 Exercise C1) Use the data in `RENTAL` for this exercise. The data on rental prices and other variables for college towns are for the years 1980 and 1990. The idea is to see whether a stronger presence of students affects rental rates. The model is

$$\log(rent_{it}) = \beta_0 + \delta y90_t + \beta_1 \log(pop_{it}) + \beta_2 \log(avginc_{it}) + \beta_3 pctstu_{it} + e_{it}$$

Here, $y90$ is a dummy variable for the year 1990, *pop* is city population, *avginc* is average income, and *pctstu* is student population as a percentage of city population (during the school year).

(a) Estimate the equation by pooled OLS and report the results. Report the standard error under both homoskedasticity and heteroskedasticity. Interpret $\hat{\delta}$ and $\hat{\beta}_3$.

(b) Re-estimate the model with city fixed effects. Report standard errors that are clustered at the city level (this is $\hat{V}^{\text{FE}}$ in the lecture notes. Also see example R code). Interpret $\hat{\delta}$ and $\hat{\beta}_3$.

(c) Compare your estimates in (a) and (b). Suggest reasons for any differences. Is either estimate the causal effect of student population on rent?

(d) Re-estimate the model without $y90$ but with both city and year fixed effects. Verify that they are the same as in (b).

Note: you can complete this exercise using any software/language. `R` is recommended.