Homework 3

1. Suppose that for simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where $\mathrm{E}(u_i) = 0$.

(a) Show that $\mathrm{Cov}(X_i, u_i) \neq 0 \Leftrightarrow \mathrm{E}(X_i u_i) \neq 0$.

(b) Suppose that we have found an instrument variable Z for X, write down the IV estimator for β_0 and β_1 respectively using Z.

(c) If we first run a regression with X on Z and hence have $\hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 Z_i$, where $\hat{\alpha}_0$ and $\hat{\alpha}_1$ refers to the corresponding OLS estimator. Then we run the regression with Y on $\hat{X}$ such that $Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$. If denote the corresponding OLS estimator by $\tilde{\beta}_0$ and $\tilde{\beta}_1$ respectively, show that $\tilde{\beta}_0$ and $\tilde{\beta}_1$ is equal to the IV estimator we obtained in (b).

2. A researcher is investigating the effect of income on family size in rural China. He considers the following model

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

where Y is the family size, X is the family income, and u is the classical stochastic error term that is homoskedastic, "serially" uncorrelated, and uncorrelated with X. Due to data confidentiality, the researcher can only obtain data that come from a survey of G villages but include only village average family size and village average income. However, he does know the number of families (n_g) sampled in village g for $g = 1, \ldots, G$.

(a) The researcher obtains the OLS estimator $\left(\hat{\beta}_{0,OLS}, \hat{\beta}_{1,OLS}\right)$ of (β_0, β_1) using the averaged data. What are the consequences if he uses the conventional standard error to make inference on β_1.

(b) Propose an alternative method to estimate (β_0, β_1), and explain why the alternative estimator is better than the OLS method in (a).

3. Consider the following linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where $\{u_i\}$ is IID, $\mathrm{E}(u_i \mid X_i) = 0$ and $\mathrm{Var}(u_i \mid X_i) = \sigma_u^2$. Suppose that Y_i is measured with error so that the data used in the regression is $\tilde{Y}_i$, where

$$\tilde{Y}_i = Y_i + e_i,$$

and e_i is a measurement error which is assumed to be IID mean 0 and variance σ_e^2 and independent of X_i and u_i. Consider following population regression

$$\tilde{Y}_i = \beta_0 + \beta_1 X_i + v_i$$

where v_i is the corresponding stochastic error when using mis-measured dependent variable.

(a) Are the OLS estimators still unbiased and consistent? Why or why not?

(b) Can the confidence intervals be constructed in the usual way?

(c) Evaluate the statement: measurement error in the regressor Y_i is a serious problem whereas measurement error in X_i is not.

(d) What parts of your regression output will be affected by the measurement error in Y_i.