Homework 5

Problem 1. Consider the following simple linear regression model

$$Y_i = \beta X_i + U_i,$$

where $\beta \in \mathbb{R}$ is the unknown parameter. The econometrician is interested in constructing a $1-\alpha$ asymptotic confidence interval for β, where $0 < \alpha < 1/2$. Assume that the data are i.i.d. and the following assumptions hold, A-i. $\mathbb{E}\left(X_i U_i\right) = 0$. A-ii. $0 < \mathbb{E}X_i^2 < \infty$, $j = 1, \ldots, k$. A-iii. $\mathbb{E}\left(U_i^2 | X_i\right) = \sigma^2$.

Define

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_n X_i\right)^2,$$

where $\hat{\beta}_n$ is the OLS estimator of β. For each confidence interval listed below indicate if it is asymptotically valid. Carefully justify your answers.

1. $\left[\hat{\beta}_n - z_{1-\alpha/2}\hat{\sigma}_n, \hat{\beta}_n + z_{1-\alpha/2}\hat{\sigma}_n\right]$.
2. $\left[\hat{\beta}_n - z_{1-\alpha/2}\hat{\sigma}_n/\sqrt{n}, \hat{\beta}_n + z_{1-\alpha/2}\hat{\sigma}_n/\sqrt{n}\right]$.
3. $\left(-\infty, \hat{\beta}_n - z_\alpha\sqrt{\hat{\sigma}_n^2/\sum_{i=1}^{n} X_i^2}\right]$.
4. $\left\{b \in \mathbb{R} : \left(\hat{\beta}_n - b\right)^2 \leq \chi^2_{1,1-\alpha} v_n\right\}$, where

$$v_n = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_n X_i\right)^2 X_i^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2}.$$

Solution. First write the asymptotic distribution of $\hat{\beta}_n$:

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \to_d N\left(0, \frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}\right).$$

Since $\hat{\sigma}_n^2 \to_p \sigma^2$, by WLLN and continuous mapping theorem (CMT),

$$\frac{\hat{\sigma}_n^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2} \to_p \frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}.$$

Then by CMT and Slutsky's theorem,

$$\frac{\sqrt{n}\left(\hat{\beta}_n - \beta\right)}{\sqrt{\frac{\hat{\sigma}_n^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}} \to_d \left(\frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}\right)^{-1/2} \cdot N\left(0, \frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}\right) \sim N\left(0, 1\right).$$

Or,

$$\frac{\hat{\beta}_n - \beta}{\sqrt{\frac{\hat{\sigma}_n^2}{\sum_{i=1}^{n} X_i^2}}} \to_d N\left(0, 1\right)$$

and by CMT,

$$\left|\frac{\hat{\beta}_n-\beta}{\sqrt{\frac{\hat{\sigma}_n^2}{\sum_{i=1}^n X_i^2}}}\right| \to_d |Z|,$$

where $Z \sim N(0,1)$.

1.

$$\begin{aligned}
\Pr\left(\beta \in \left[\hat{\beta}_n - z_{1-\alpha/2}\hat{\sigma}_n, \hat{\beta}_n + z_{1-\alpha/2}\hat{\sigma}_n\right]\right) &= \Pr\left(\left|\frac{\hat{\beta}_n-\beta}{\hat{\sigma}_n}\right| \le z_{1-\alpha/2}\right) \\
&= \Pr\left(\left|\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\sqrt{n}\cdot\hat{\sigma}_n}\right| \le z_{1-\alpha/2}\right).
\end{aligned}$$

By CMT,

$$\frac{1}{\sqrt{n}\cdot\hat{\sigma}_n} \to_p 0 \cdot \frac{1}{\sigma} = 0.$$

By Slutsky's theorem,

$$\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\sqrt{n}\cdot\hat{\sigma}_n} \to_d 0 \cdot N\left(0, \frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}\right) = 0.$$

So

$$\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\sqrt{n}\cdot\hat{\sigma}_n} \to_p 0$$

and

$$\Pr\left(\left|\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\sqrt{n}\cdot\hat{\sigma}_n}\right| \le z_{1-\alpha/2}\right) \to 1.$$

It does not converge to $1-\alpha$, therefore the confidence interval is not valid.

2.

$$\Pr\left(\beta \in \left[\hat{\beta}_n - z_{1-\alpha/2}\hat{\sigma}_n/\sqrt{n}, \hat{\beta}_n + z_{1-\alpha/2}\hat{\sigma}_n/\sqrt{n}\right]\right) = \Pr\left(\left|\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\hat{\sigma}_n}\right| \le z_{1-\alpha/2}\right).$$

Note

$$\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\hat{\sigma}_n} \to_d \sigma^{-1}\cdot N\left(0, \frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}\right) \sim N\left(0, \frac{1}{\mathbb{E}\left(X_i^2\right)}\right).$$

So $\Pr\left(\left|\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\hat{\sigma}_n}\right| \le z_{1-\alpha/2}\right)$ does not converge to $1-\alpha$, unless $\mathbb{E}X_i^2 = 1$.

3. By CMT,

$$-\frac{\hat{\beta}_n-\beta}{\sqrt{\frac{\hat{\sigma}_n^2}{\sum_{i=1}^n X_i^2}}} \to_d (-1)\cdot N(0,1) \sim N(0,1).$$

Then,

$$\Pr\left(\beta \in \left(-\infty, \hat{\beta}_n - z_\alpha\sqrt{\hat{\sigma}_n^2/\sum_{i=1}^n X_i^2}\right]\right) = \Pr\left(-\frac{\hat{\beta}_n-\beta}{\sqrt{\frac{\hat{\sigma}_n^2}{\sum_{i=1}^n X_i^2}}} \le -z_\alpha\right) \to \Pr\left(Z \le -z_\alpha\right) = 1-\alpha,$$

where $Z \sim N(0,1)$.

4. We have

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_nX_i\right)^2X_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)^2}\to_p\frac{\mathbb{E}\left(U_i^2X_i^2\right)}{\mathbb{E}\left(X_i^2\right)^2}=\frac{\sigma^2\mathbb{E}\left(X_i^2\right)}{\mathbb{E}\left(X_i^2\right)^2}=\frac{\sigma^2}{\mathbb{E}\left(X_i^2\right)}.$$

(Why?) Therefore, by CMT and Slutsky's theorem,

$$\sqrt{n}\left(\hat{\beta}_n-\beta\right)\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_nX_i\right)^2X_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)^2}\right)^{-1}\sqrt{n}\left(\hat{\beta}_n-\beta\right)=\left(\frac{\sqrt{n}\left(\hat{\beta}_n-\beta\right)}{\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_nX_i\right)^2X_i^2}{\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)^2}}}\right)^2\to_d Z^2\sim\chi_1^2,$$

where $Z\sim N(0,1)$. So

$$\frac{\left(\hat{\beta}_n-\beta\right)^2}{v_n}\to_d\chi_1^2$$

and

$$\Pr\left(\frac{\left(\hat{\beta}_n-\beta\right)^2}{v_n}\le\chi_{1,1-\alpha}^2\right)\to1-\alpha.$$

The confidence interval is valid.

Problem 2. (a) Let $X_m\sim t_m$, i.e. X_m is a t-distributed random variable with m degrees of freedom. Show that $X_m\to_d N(0,1)$ as $m\to\infty$. Hints: Use the definition of the t-distribution and the WLLN. (b) Let $X_{q,m}\sim F_{q,m}$, i.e. $X_{q,m}$ is an F-distributed random variable with q and m degrees of freedom. Find the limiting distribution of $X_{q,m}$ as $m\to\infty$. Hints: Use the definition of the F-distribution and the WLLN.

Solution. By the definition of a t-distribution, $X_m=Z_0/\sqrt{(Z_1^2+\ldots+Z_m^2)/m}$, where $Z_0,Z_1,\ldots,Z_m$ are iid $N(0,1)$. As $m\to\infty$,

$$\frac{1}{m}(Z_1^2+\ldots+Z_m^2)\to_p\mathbb{E}Z_1^2=1,$$

and therefore by Slutsky's theorem,

$$X_m\to_p Z_0\sim N(0,1).$$

Using the definition of $F_{q,m}$ distribution, we can write:

$$X_{q,m}=\frac{Y/q}{m^{-1}\sum_{j=1}^{m}Z_j^2},$$

where $Z_1,\ldots,Z_m$ are iid $N(0,1)$, and

$$Y\sim\chi_q^2$$

and independent from Z's. Since $\mathbb{E}Z_j^2=1$, by WLLN,

$$m^{-1}\sum_{j=1}^{m}Z_j^2\to_p 1.$$

It follows that

$$X_{q,m}\to_d Y/q.$$

The result can be alternatively stated as

$$qF_{q,m}\to_d\chi_q^2.$$

Problem 3. Let $\hat{\theta}_n$ be a consistent estimator of the scalar parameter θ. Suppose further that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \omega^2)$, for some $0 < \omega^2 < \infty$. Assume that $\theta \neq 0$. Use the delta method or CMT to find the non-degenerate (i.e., not a constant) asymptotic distributions of

$$\frac{1}{\hat{\theta}_n}$$

after a suitable normalization (where a "suitable normalization" means subtraction of a constant and/or multiplication by a constant, and a constant can be any non-random term).

Solution. Using the delta-method,

$$\sqrt{n}\left(\frac{1}{\hat{\theta}_n} - \frac{1}{\theta}\right) \to_d -\frac{1}{\theta^2} N(0, \omega^2) =^d N(0, \omega^2/\theta^4).$$

Problem 4. Aggregate demand Q_D for a certain commodity is determined by its price P, aggregate income Y, and population, POP,

$$Q_D = \beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + U^D$$

and aggregate supply is given by

$$Q_S = \alpha_1 + \alpha_2 P + U^S$$

where U_D and U_S are independently distributed error terms: U_D and U_S are independent from all other variables and they are also independent from each other. Remember that the quantity and the price are determined simultaneously in the equilibrium $Q_S = Q_D = Q$. We observe only the equilibrium values Q so that the observed price must satisfy the equation (demand = supply):

$$\beta_1 + \beta_2 P + \beta_3 Y + \beta_4 POP + U^D = \alpha_1 + \alpha_2 P + U^S.$$

1. Show that the OLS (ordinary least squares) estimator of α_2 will be inconsistent if OLS is used to fit the supply equation.
2. Show that a consistent estimator of α_2 is
$$\widetilde{\alpha}_2 = \frac{\sum_{i=1}^n \left(Y_i - \overline{Y}\right)\left(Q_i - \overline{Q}\right)}{\sum_{i=1}^n \left(Y_i - \overline{Y}\right)\left(P_i - \overline{P}\right)}.$$
($\overline{Y} = n^{-1}\sum_{i=1}^n Y_i$, $\overline{Q} = n^{-1}\sum_{i=1}^n Q_i$, $\overline{P} = n^{-1}\sum_{i=1}^n P_i$.)
3. Explain how to construct a bootstrap percentile confidence interval for α_2, using $\widetilde{\alpha}_2$.

Solution.

1. The reduced form equation (which expresses P as a function of the explanatory variables and the error terms) for P is
$$P = \frac{1}{\alpha_2 - \beta_2}\left(\beta_1 - \alpha_1 + \beta_3 Y + \beta_4 POP + U^D - U^S\right).$$
Therefore in the supply equation
$$Q_S = \alpha_1 + \alpha_2 P + U^S,$$
P is correlated with U^S. The OLS estimator is
$$\begin{aligned}\hat{\alpha}_2^{OLS} =& \frac{\sum_{i=1}^n \left(P_i - \overline{P}\right)\left(Q_i - \overline{Q}\right)}{\sum_{i=1}^n \left(P_i - \overline{P}\right)^2} \\ =& \alpha_2 + \frac{\sum_{i=1}^n \left(P_i - \overline{P}\right)\left(U_i^S - \overline{U}^S\right)}{\sum_{i=1}^n \left(P_i - \overline{P}\right)^2}\end{aligned}$$

$$\longrightarrow_p \alpha_2 + \frac{\operatorname{Cov}\left(P_i, U_i^S\right)}{\operatorname{Var}\left(P_i\right)}$$

and

$$\begin{aligned}
\operatorname{Cov}\left(P_i, U_i^S\right) =& \operatorname{Cov}\left(\frac{1}{\alpha_2-\beta_2}\left(\beta_1-\alpha_1+\beta_3 Y_i+\beta_4 POP_i+U_i^D-U_i^S\right), U_i^S\right) \\
=& -\frac{1}{\alpha_2-\beta_2}\operatorname{Var}\left(U_i^S\right) \\
\neq& 0
\end{aligned}$$

assuming that Y and POP are exogenous and so $\operatorname{Cov}\left(U^S, Y\right) = \operatorname{Cov}\left(U^S, POP\right) = 0$. We are told that U^S and U^D are distributed independently, so that $\operatorname{Cov}\left(U^S, U^D\right) = 0$.

2. The instrument variable estimator is

$$\begin{aligned}
\hat{\alpha}_2^{IV} =& \frac{\sum_{i=1}^n\left(Y_i-\overline{Y}\right)\left(Q_i-\overline{Q}\right)}{\sum_{i=1}^n\left(Y_i-\overline{Y}\right)\left(P_i-\overline{P}\right)} \\
=& \alpha_2 + \frac{\sum_{i=1}^n\left(Y_i-\overline{Y}\right)\left(U^S-\overline{U}^S\right)}{\sum_{i=1}^n\left(Y_i-\overline{Y}\right)\left(P_i-\overline{P}\right)} \\
\longrightarrow_p& \alpha_2 + \frac{\operatorname{Cov}\left(Y_i, U_i^S\right)}{\operatorname{Cov}\left(P_i, Y_i\right)}.
\end{aligned}$$

The desired result follows from the assumptions $\operatorname{Cov}\left(Y_i, U_i^S\right) = 0$ and $\operatorname{Cov}\left(P_i, Y_i\right) \neq 0$.

3. See the lecture slides.

Problem 5. (a) Let $X_m \sim t_m$, i.e. X_m is a t-distributed random variable with m degrees of freedom. Show that $X_m \to_d N(0,1)$ as $m \to \infty$. Hints: Use the definition of the t-distribution and the WLLN. (b) Let $X_{q,m} \sim F_{q,m}$, i.e. $X_{q,m}$ is an F-distributed random variable with q and m degrees of freedom. Find the limiting distribution of $X_{q,m}$ as $m \to \infty$. Hints: Use the definition of the F-distribution and the WLLN.

Solution. By the definition of a t-distribution, $X_m = Z_0/\sqrt{(Z_1^2+\ldots+Z_m^2)/m}$, where $Z_0, Z_1, \ldots, Z_m$ are iid $N(0,1)$. As $m \to \infty$,

$$\frac{1}{m}(Z_1^2+\ldots+Z_m^2) \to_p \mathbb{E}Z_1^2 = 1,$$

and therefore by Slutsky's theorem,

$$X_m \to_p Z_0 \sim N(0,1).$$

Using the definition of $F_{q,m}$ distribution, we can write:

$$X_{q,m} = \frac{Y/q}{m^{-1}\sum_{j=1}^m Z_j^2},$$

where $Z_1, \ldots, Z_m$ are iid $N(0,1)$, and

$$Y \sim \chi_q^2$$

and independent from Z's. Since $\mathbb{E}Z_j^2 = 1$, by WLLN,

$$m^{-1}\sum_{j=1}^m Z_j^2 \to_p 1.$$

It follows that

$$X_{q,m} \to_d Y/q.$$

The result can be alternatively stated as

$$qF_{q,m} \to_d \chi_q^2.$$

Problem 6. Suppose that the linear model

$$\text{PS} = \beta_0 + \beta_1 \text{Funds} + \beta_2 \text{Risk} + U$$

satisfies $\mathrm{E}[U] = \mathrm{E}[U \cdot \text{Funds}] = \mathrm{E}[U \cdot \text{Risk}] = 0$. PS is the percentage of a person's savings invested in the stock market, Funds is the number of mutual funds that the person can choose from, and Risk is some measure of risk tolerance (larger Risk means the person has a higher tolerance for risk).

1. If Funds and Risk are positively correlated, does the slope coefficient in the simple regression of PS on Funds overestimate or underestimate β_1, in large samples?

2. We are unable to observe Risk directly, but we have data on the amount of life insurance a worker has, Insurance. Assume that Insurance is noisy measure of Risk, Insurance = Risk + e, with $\mathrm{E}[e] = \mathrm{E}[\text{Risk} \cdot e] = \mathrm{E}[\text{Funds} \cdot e] = \mathrm{E}[eU] = 0$. Will the OLS estimate of the coefficient on Funds in a regression of PS on Funds and Insurance be a consistent estimate of β_1?

3. Suppose we also have data on how often a worker gambles, Gamble. Assume that Gamble is an independent noisy measure of Risk, Gamble = Risk + v, with $\mathrm{E}[v] = \mathrm{E}[vU] = \mathrm{E}[ve] = \mathrm{E}[\text{Risk} \cdot v] = \mathrm{E}[\text{Funds} \cdot v] = 0$. Explain how we can consistently estimate β_1 using our data on PS, Funds, Insurance, and Gamble.

Solution.

1. The slope coefficient in the simple regression of PS on Funds converges in probability to $\beta_1 + \beta_2 \frac{\text{Cov(Funds,Risk)}}{\text{Var(Funds)}}$. We expect $\beta_2 \geq 0$. Therefore, the slope coefficient in the simple regression overestimates β_1.

2. No. Insurance is endogenous, due to measurement error.

3. Run an IV regression, using Gamble as an instrument for Insurance.

Problem 7. Consider a simple model to estimate the effect of personal computer (PC) ownership on college grade point average for graduating seniors at a large public university:

$$GPA = \beta_0 + \beta_1 PC + u,$$

where PC is a binary variable indicating PC ownership.

1. Why might PC ownership be correlated with u?

2. Explain why PC is likely to be related to parents' annual income. Does this mean parental income is a good IV for PC? Why or why not?

3. Suppose that, four years ago, the university gave grants to buy computers to roughly one-half of the incoming students, and the students who received grants were randomly chosen. Carefully explain how you would use this information to construct an instrumental variable for PC.

Solution.

1. Personal income/wealth is an omitted variable in this regression, as wealth can affect GPA (the student can hire more tutors), however, wealth is correlated with PC.

2. While parents' income is correlated with PC ownership, it does not satisfy the exogeneity assumption as it is correlated with personal income of the student.

3. A dummy variable indicating whether the student received the grant can be used as an IV: it is correlated with PC and independent of students' income since it was given randomly. However, the econometrician will have to restrict his sample only to those who began there studies four years ago when the university gave the grant.

Problem 8. Suppose we observe the i.i.d. random sample $\{(Y_i, X_i)\}_{i=1}^n$. Denote $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\overline{Y}_n = n^{-1}\sum_{i=1}^n Y_i$, $\mu_X = \mathrm{E}[X_i]$ ($\mu_X \neq 0$) and $\mu_Y = \mathrm{E}[Y_i]$. We are interested in μ_Y/μ_X. Derive the asymptotic distribution of $\sqrt{n}\left(\overline{Y}_n/\overline{X}_n - \mu_Y/\mu_X\right)$.

Solution.

$$\begin{aligned}
\frac{\overline{Y}_n}{\overline{X}_n} - \frac{\mu_Y}{\mu_X} &= \left(\frac{\overline{Y}_n}{\mu_X} - \frac{\mu_Y}{\mu_X}\right) - \left(\frac{\overline{Y}_n}{\mu_X \overline{X}_n} - \frac{\mu_Y}{\mu_X^2} + \frac{\mu_Y}{\mu_X^2}\right) \cdot \left(\overline{X}_n - \mu_X\right) \\
&= \frac{1}{n}\sum_{i=1}^n \left\{\frac{1}{\mu_X}\left(Y_i - \mu_Y\right) - \frac{\mu_Y}{\mu_X^2}\left(X_i - \mu_X\right)\right\} - \left(\frac{\overline{Y}_n}{\mu_X \overline{X}_n} - \frac{\mu_Y}{\mu_X^2}\right)\left(\overline{X}_n - \mu_X\right).
\end{aligned}$$

By Slutsky lemma, $\frac{\overline{Y}_n}{\mu_X \overline{X}_n} - \frac{\mu_Y}{\mu_X^2} \to_p 0$. By CLT, $\sqrt{n}\left(\overline{X}_n - \mu_X\right) \to_d N\left(0, \sigma_X^2\right)$. Therefore,

$$-\left(\frac{\overline{Y}_n}{\mu_X \overline{X}_n} - \frac{\mu_Y}{\mu_X^2}\right)\sqrt{n}\left(\overline{X}_n - \mu_X\right) \to_p 0.$$

By CLT,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\frac{1}{\mu_X}\left(Y_i - \mu_Y\right) - \frac{\mu_Y}{\mu_X^2}\left(X_i - \mu_X\right)\right\} \to_d N\left(0, \mathrm{E}\left[\left(\frac{1}{\mu_X}\left(Y_i - \mu_Y\right) - \frac{\mu_Y}{\mu_X^2}\left(X_i - \mu_X\right)\right)^2\right]\right).$$

We have

$$\sqrt{n}\left(\frac{\overline{Y}_n}{\overline{X}_n} - \frac{\mu_Y}{\mu_X}\right) \to_d N\left(0, \mathrm{E}\left[\left(\frac{1}{\mu_X}\left(Y_i - \mu_Y\right) - \frac{\mu_Y}{\mu_X^2}\left(X_i - \mu_X\right)\right)^2\right]\right).$$