Final Exam (January 2023)

Problem 1. (8 points) 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 $\{(Y_i, X_i) : i = 1, ..., n\}$ are i.i.d. and the following assumptions hold: $\mathrm{E}[X_i U_i] = 0$; $0 < \mathrm{E}[X_i^2] < \infty$; $\mathrm{E}[U_i^2 \mid X_i] = \sigma^2$. 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 (that is, the coverage probability converges to $1-\alpha$). $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution.

(i) $\left[\hat{\beta}_n - z_{1-\alpha/2}\left(\hat{\sigma}_n/\sqrt{n}\right), \hat{\beta}_n + z_{1-\alpha/2}\left(\hat{\sigma}_n/\sqrt{n}\right)\right]$.

(ii) $\left(-\infty, \hat{\beta}_n - z_\alpha \sqrt{\hat{\sigma}_n^2 / \left(\sum_{i=1}^n X_i^2\right)}\right]$.

Solution.

(i) No. Write

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

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 $\mathbb{E}X_i^2 = 1$.

(ii) Yes. We have

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

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

Problem 2. (12 Points) The family of Pareto distributions has been used as a model for a density function with a slowly decaying tail:

$$f(x \mid x_0, \theta) = \theta x_0^\theta x^{-\theta-1}$$

for $x \ge x_0$, $\theta > 1$. Assume that x_0 is given and that $X_1, ..., X_n$ is an i.i.d. sample with density $f(\cdot \mid x_0, \theta_*)$.

(i) We can find that $\mathrm{E}[X_1] = (x_0\theta_*)/(\theta_* - 1)$ (you are not required to prove this). Find an estimator of θ_* based on the method of moments.

(ii) Find the maximum likelihood estimator of θ_*, $\hat{\theta}_n^{MLE}$.

(iii) Find the asymptotic variance of the maximum likelihood estimator. i.e. Find σ^2 such that

$$\sqrt{n}\left(\hat{\theta}_n^{MLE} - \theta_*\right) \to_d \mathrm{N}\left(0, \sigma^2\right).$$

Solution.

(i) Let us denote $\mu = \mathrm{E}[X_1]$. Then we have $\theta_* = \frac{\mu}{\mu - x_0}$. A method of moment estimator of θ_* is $\hat{\theta}^{MM} = \frac{\overline{X}}{\overline{X} - x_0}$ where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean.

(ii) The log-likelihood function is

$$\ell(\theta) = \log\left(\prod_{i=1}^{n} \theta x_0^{\theta} X_i^{-\theta-1}\right) = n\log(\theta) + n\theta\log(x_0) - (\theta+1)\sum_{i=1}^{n}\log(X_i).$$

The derivative is

$$\ell'(\theta) = \frac{n}{\theta} + n\log(x_0) - \sum_{i=1}^{n}\log(X_i).$$

Solving for $\ell'(\theta) = 0$, the maximum likelihood estimator is given by

$$\hat{\theta}^{MLE} = \frac{n}{\sum_{i=1}^{n}\log(X_i) - n\log(x_0)}.$$

(iii) We have

$$\ell''(\theta) = -\frac{n}{\theta^2}.$$

The asymptotic variance is

$$\sigma^2 = -n\left(\ell''(\theta_*)\right)^{-1} = \theta_*^2.$$

Problem 3. (12 Points) Consider the following linear model:

$$Y = \beta_0 + \beta_1 X + U.$$

Suppose $\mathrm{E}[U] = 0$ and $\mathrm{E}[XU] \neq 0$, but you have two valid instruments, Z_1 and Z_2 ($\mathrm{E}[Z_1 U] = \mathrm{E}[Z_2 U] = 0$). Write the first-stage regression as

$$X = \pi_0^X + \pi_1^X Z_1 + \pi_2^X Z_2 + \epsilon, \tag{1}$$

where $\mathrm{E}[\epsilon] = \mathrm{E}[\epsilon Z_1] = \mathrm{E}[\epsilon Z_2] = 0$.

(i) Explain how you compute the 2SLS estimator of β_1.

(ii) Also write the linear regression of Y on Z_1 and Z_2 as

$$Y = \pi_0^Y + \pi_1^Y Z_1 + \pi_2^Y Z_2 + \nu, \tag{2}$$

where $\mathrm{E}[\nu] = \mathrm{E}[\nu Z_1] = \mathrm{E}[\nu Z_2] = 0$. Let $\hat{\pi}_1^X$ and $\hat{\pi}_1^Y$ denote the OLS estimators of the regressions (1) and (2). Show that

$$\hat{\beta}_1^{ILS,1} = \frac{\hat{\pi}_1^Y}{\hat{\pi}_1^X}$$

is a consistent estimator of β_1. You may take as given that $\hat{\pi}_1^X \to_p \pi_1^X$ and $\hat{\pi}_1^Y \to_p \pi_1^Y$.

(iii) Using the coefficients on Z_2 instead of Z_1, we define

$$\hat{\beta}_1^{ILS,2} = \frac{\hat{\pi}_2^Y}{\hat{\pi}_2^X}.$$

Like $\hat{\beta}_1^{ILS,1}$, $\hat{\beta}_1^{ILS,2}$ is a consistent estimator of β_1. In some data, suppose we find $\hat{\beta}_1^{ILS,1} - \hat{\beta}_1^{ILS,2}$ is large. What might this indicate about our instruments?

Solution.

(i) First stage: regress X on Z_1, Z_2 and a constant and get fitted value $\widehat{X}$. Second stage: regress Y on $\widehat{X}$ and a constant. The OLS coefficient for $\widehat{X}$ is the 2SLS estimate for β_1.

(ii) Plug $X = \pi_0^X + \pi_1^X Z_1 + \pi_2^X Z_2 + \epsilon$ into $Y = \beta_0 + \beta_1 X + U$:

$$\begin{aligned} Y &= \beta_0 + \beta_1\left(\pi_0^X + \pi_1^X Z_1 + \pi_2^X Z_2 + \epsilon\right) + U \\ &= \left(\beta_0 + \beta_1 \pi_0^X\right) + \beta_1 \pi_1^X Z_1 + \beta_1 \pi_2^X Z_2 + \left(\beta_1 \epsilon + U\right). \end{aligned}$$

$V = \beta_1 \epsilon + U$ satisfies $\mathrm{E}[VZ_1] = \mathrm{E}[VZ_2] = \mathrm{E}[V] = 0$. Therefore, $\pi_1^Y = \beta_1 \pi_1^X$, $\pi_0^Y = \beta_0 + \beta_1 \pi_0^X$ and $\pi_2^Y = \beta_1 \pi_2^X$. $\hat{\pi}_1^Y$ consistently estimates $\pi_1^Y = \beta_1 \pi_1^X$. $\hat{\pi}_1^X$ consistently estimates π_1^X. Therefore, $\hat{\beta}_1^{ILS,1} = \hat{\pi}_1^Y / \hat{\pi}_1^X$ consistently estimates $\pi_1^Y / \pi_1^X = \beta_1$.

(iii) If both instruments are valid, $\hat{\beta}_1^{ILS,1} - \hat{\beta}_1^{ILS,2}$ should converge in probability to zero and we should find that $\hat{\beta}_1^{ILS,1} - \hat{\beta}_1^{ILS,2}$ is small. If we find $\hat{\beta}_1^{ILS,1} - \hat{\beta}_1^{ILS,2}$ is large, this indicates that one of the instruments could be invalid: $\mathrm{E}[Z_1 U] \neq 0$ or $\mathrm{E}[Z_2 U] \neq 0$.

Problem 4. (8 Points) Suppose we observe the i.i.d. random sample $\{(Y_i, X_i) : i = 1, ..., n\}$. Denote $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$, $\overline{Y}_n = n^{-1} \sum_{i=1}^n Y_i$, $\mu_X = \mathrm{E}[X_i]$ and $\mu_Y = \mathrm{E}[Y_i]$. We are interested in $\mu_Y \cdot \mu_X$. Derive the asymptotic distribution of $\sqrt{n}\left(\overline{Y}_n \cdot \overline{X}_n - \mu_Y \cdot \mu_X\right)$. Hint: Write

$$\begin{aligned} \overline{Y}_n \cdot \overline{X}_n &= \left(\overline{Y}_n - \mu_Y + \mu_Y\right)\left(\overline{X}_n - \mu_X + \mu_X\right) \\ &= \left(\overline{Y}_n - \mu_Y\right)\left(\overline{X}_n - \mu_X\right) + \mu_Y\left(\overline{X}_n - \mu_X\right) + \left(\overline{Y}_n - \mu_Y\right)\mu_X + \mu_Y \cdot \mu_X \end{aligned}$$

You may use the following result: $W_n \rightarrow_d \mathrm{N}\left(0, \sigma^2\right)$ and $\theta_n \rightarrow_p 0$, then $\theta_n W_n \rightarrow_p 0$.

Solution. Write

$$\overline{Y}_n \cdot \overline{X}_n - \mu_Y \cdot \mu_X = \left(\overline{Y}_n - \mu_Y\right)\left(\overline{X}_n - \mu_X\right) + \mu_Y\left(\overline{X}_n - \mu_X\right) + \left(\overline{Y}_n - \mu_Y\right)\mu_X$$

and

$$\sqrt{n}\left(\overline{Y}_n \cdot \overline{X}_n - \mu_Y \cdot \mu_X\right) = \sqrt{n}\left(\overline{Y}_n - \mu_Y\right)\left(\overline{X}_n - \mu_X\right) + \mu_Y\sqrt{n}\left(\overline{X}_n - \mu_X\right) + \sqrt{n}\left(\overline{Y}_n - \mu_Y\right)\mu_X. \tag{3}$$

By using $\sqrt{n}\left(\overline{Y}_n - \mu_Y\right) \rightarrow_d N\left(0, \sigma_Y^2\right)$ (σ_Y^2 denotes the variance of Y_i) and $\overline{X}_n - \mu_X \rightarrow_p 0$, we have

$$\sqrt{n}\left(\overline{Y}_n - \mu_Y\right)\left(\overline{X}_n - \mu_X\right) \rightarrow_p 0. \tag{4}$$

Write

$$\mu_Y\sqrt{n}\left(\overline{X}_n - \mu_X\right) + \sqrt{n}\left(\overline{Y}_n - \mu_Y\right)\mu_X = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mu_Y X_i + \mu_X Y_i - 2\mu_X\mu_Y\right). \tag{5}$$

Note that $\mathrm{E}[\mu_Y X_i + \mu_X Y_i] = 2\mu_X\mu_Y$. By central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mu_Y X_i + \mu_X Y_i - 2\mu_X\mu_Y\right) \rightarrow_d N\left(0, \mathrm{E}\left[\left(\mu_Y X_i + \mu_X Y_i - 2\mu_X\mu_Y\right)^2\right]\right). \tag{6}$$

By (3), (4), (5) and (6),

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

Problem 5. (8 Points) A researcher has data on the following variables for 5,061 respondents in the US National Longitudinal Survey of Youth:

- MARRIED, marital status in 1994, defined to be 1 if the respondent was married with his/her spouse present and 0 otherwise (a man/woman may not be in marriage legally with his/her spouse);
- MALE, defined to be 1 if the respondent was male and 0 if female;
- AGE in 1994 (the range being 29-37);
- S, years of schooling, defined as highest grade completed, and
- ASVABC, score on a test of cognitive ability, scaled so as to have mean 50 and standard deviation 10.

She uses Probit analysis to regress MARRIED on the other variables. The sample means of the explanatory variables and their (average) marginal effects evaluated at the sample means are shown in the table.

```
. probit MARRIED MALE AGE S ASVABC

Probit estimates                                  Number of obs   =       5061
                                                  LR chi2(4)      =     229.78
                                                  Prob > chi2     =     0.0000
Log likelihood = -3286.1289                       Pseudo R2       =     0.0338

------------------------------------------------------------------------------
     MARRIED |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        MALE |  -.1215281    .036332    -3.34    0.001    -.1927375   -.0503188
         AGE |    .028571   .0081632     3.50    0.000     .0125715    .0445705
           S |  -.0017465     .00919    -0.19    0.849    -.0197587    .0162656
      ASVABC |   .0252911   .0022895    11.05    0.000     .0208038    .0297784
       _cons |  -1.816455   .2798724    -6.49    0.000    -2.364995   -1.267916
------------------------------------------------------------------------------
```

Variable	Mean	Marginal effect
$MALE$	0.4841	-0.0467
AGE	32.52	0.0110
S	13.31	-0.0007
$ASVABC$	48.94	0.0097

(i) Discuss the conclusions one may reach, given the Probit output and the table, comment on whether you think they are reasonable.

(ii) The researcher considers including CHILD, a dummy variable defined to be 1 if the respondent had children and 0 otherwise, as an explanatory variable. When she does this, its t-statistic is found to be 33.65 and its average marginal effect is found to be 0.5685. Discuss these findings.

Solution.

(i) Being male has a small but highly significant negative effect. This is plausible because males tend to marry later than females and the cohort is still relatively young. Age has a highly significant positive effect, again plausible because older people are more likely to have married than younger people. Schooling has no apparent effect at all. It is not obvious whether this is plausible. Cognitive ability has a highly significant positive effect. Again, it is not obvious whether this is plausible.

(ii) Obviously one would expect a high positive correlation between being married and having children and this would account for the huge and highly significant coefficient. However getting married and having children are often a joint decision, and accordingly it is simplistic to suppose that one characteristic is a determinant of the other. The finding should not be taken at face value.

Problem 6. (16 Points) Which of the following can and which cannot cause the usual OLS-based t-statistic to be invalid (that is, not to have the t distributions under H_0, even in large samples)? Explain briefly.

(i) Heteroskedasticity.

(ii) A sample correlation coefficient of 0.95 between two regressors.

(iii) Omitting an important explanatory variable.

(iv) Using the bootstrap critical value.

Solution.

(i) The t-statistic is invalid if the conventional standard error is used. It is valid if the heteroskedasticity-robust (White) standard error is used.

(ii) Valid. The t-statistic is asymptotically standard normal even if two regressors are highly correlated. However, the "no multicollinearity" assumption is violated if the correlation is 1 or -1. In this case, the t-statistic is invalid.

(iii) Invalid. The OLS estimator is inconsistent in this case.

(iv) Valid. The bootstrap critical value is just another way to approximate the exact distribution of the t-statistic.

Problem 7. (12 points) Consider the regression model

$$\begin{aligned} Y_i &= \beta X_i + U_i, \\ \mathrm{E}\left[U_i \mid X_i\right] &= 0, \\ \mathrm{E}\left[U_i^2 \mid X_i\right] &= \sigma^2, \end{aligned}$$

where $\beta \in \mathbb{R}$ is an unknown scalar parameter. Assume that $\{(Y_i, X_i) : i = 1, \ldots, n\}$ are i.i.d. Consider the following estimator of β:

$$\widetilde{\beta}_n = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}$$

and the OLS estimator:

$$\widehat{\beta}_n = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}.$$

Let $\overline{\beta}_n = \gamma \widetilde{\beta}_n + (1-\gamma)\widehat{\beta}_n$ for some $\gamma \in [0,1]$.

(i) Show that $\overline{\beta}_n \to_p \beta$ as $n \to \infty$.

(ii) Show that $\sqrt{n}\left(\overline{\beta}_n - \beta\right)$ is asymptotically normal, and find the asymptotic variance.

(iii) What would the optimal γ be? Explain.

Solution.

(i) Write

$$\widetilde{\beta}_n = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} = \beta + \frac{n^{-1}\sum_{i=1}^n U_i}{n^{-1}\sum_{i=1}^n X_i}$$

and

$$\widehat{\beta}_n = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \beta + \frac{n^{-1}\sum_{i=1}^n U_i X_i}{n^{-1}\sum_{i=1}^n X_i^2}.$$

By using $n^{-1}\sum_{i=1}^n U_i \to_p 0$, $n^{-1}\sum_{i=1}^n U_i X_i \to_p 0$, $n^{-1}\sum_{i=1}^n X_i^2 \to_p \mathrm{E}\left[X_i^2\right] \neq 0$ and $n^{-1}\sum_{i=1}^n X_i \to_p \mathrm{E}\left[X_i\right] \neq 0$, we have $\widetilde{\beta}_n \to_p \beta$, $\widehat{\beta}_n \to_p \beta$ and $\overline{\beta}_n = \gamma\widetilde{\beta}_n + (1-\gamma)\widehat{\beta}_n \to_p \gamma\beta + (1-\gamma)\beta = \beta$.

(ii) Write

$$\overline{\beta}_n = \gamma\widetilde{\beta}_n + (1-\gamma)\widehat{\beta}_n = \beta + \gamma \cdot \frac{n^{-1}\sum_{i=1}^n U_i}{n^{-1}\sum_{i=1}^n X_i} + (1-\gamma)\frac{n^{-1}\sum_{i=1}^n U_i X_i}{n^{-1}\sum_{i=1}^n X_i^2}.$$

Then, we can write

$$\begin{aligned} \sqrt{n}\left(\overline{\beta}_n - \beta\right) &= \gamma \cdot \frac{n^{-1/2}\sum_{i=1}^n U_i}{n^{-1}\sum_{i=1}^n X_i} + (1-\gamma)\frac{n^{-1/2}\sum_{i=1}^n U_i X_i}{n^{-1}\sum_{i=1}^n X_i^2} \\ &= \gamma \cdot \frac{n^{-1/2}\sum_{i=1}^n U_i}{\mathrm{E}\left[X_i\right]} + (1-\gamma)\frac{n^{-1/2}\sum_{i=1}^n U_i X_i}{\mathrm{E}\left[X_i^2\right]} \\ &\quad + \gamma \cdot \left(\frac{1}{n^{-1}\sum_{i=1}^n X_i} - \frac{1}{\mathrm{E}\left[X_i\right]}\right) n^{-1/2}\sum_{i=1}^n U_i \\ &\quad + (1-\gamma)\left(\frac{1}{n^{-1}\sum_{i=1}^n X_i^2} - \frac{1}{\mathrm{E}\left[X_i^2\right]}\right) n^{-1/2}\sum_{i=1}^n U_i X_i. \end{aligned} \tag{7}$$

By using $n^{-1/2}\sum_{i=1}^n U_i \to_d N\left(0, \mathrm{E}\left[U_i^2\right]\right)$, $n^{-1/2}\sum_{i=1}^n U_i X_i \to_d N\left(0, \mathrm{E}\left[X_i^2\right]\sigma^2\right)$,

$$\frac{1}{n^{-1}\sum_{i=1}^n X_i} - \frac{1}{\mathrm{E}\left[X_i\right]} \to_p 0$$

and

$$\frac{1}{n^{-1}\sum_{i=1}^n X_i^2} - \frac{1}{\mathrm{E}\left[X_i^2\right]} \to_p 0,$$

we have

$$\gamma \cdot \left(\frac{1}{n^{-1}\sum_{i=1}^n X_i} - \frac{1}{\mathrm{E}\left[X_i\right]}\right) n^{-1/2}\sum_{i=1}^n U_i \to_p 0 \tag{8}$$

and

$$(1-\gamma)\left(\frac{1}{n^{-1}\sum_{i=1}^{n}X_i^2}-\frac{1}{\mathrm{E}\left[X_i^2\right]}\right)n^{-1/2}\sum_{i=1}^{n}U_iX_i\to_p 0. \tag{9}$$

Write

$$\gamma\cdot\frac{n^{-1/2}\sum_{i=1}^{n}U_i}{\mathrm{E}\left[X_i\right]}+(1-\gamma)\frac{n^{-1/2}\sum_{i=1}^{n}U_iX_i}{\mathrm{E}\left[X_i^2\right]}=n^{-1/2}\sum_{i=1}^{n}\left(\frac{\gamma U_i}{\mathrm{E}\left[X_i\right]}+\frac{(1-\gamma)U_iX_i}{\mathrm{E}\left[X_i^2\right]}\right).$$

We have

$$\mathrm{E}\left[\frac{\gamma U_i}{\mathrm{E}\left[X_i\right]}+\frac{(1-\gamma)U_iX_i}{\mathrm{E}\left[X_i^2\right]}\right]=0.$$

By central limit theorem,

$$n^{-1/2}\sum_{i=1}^{n}\left(\frac{\gamma U_i}{\mathrm{E}\left[X_i\right]}+\frac{(1-\gamma)U_iX_i}{\mathrm{E}\left[X_i^2\right]}\right)\to_d N\left(0,\mathrm{E}\left[\left(\frac{\gamma U_i}{\mathrm{E}\left[X_i\right]}+\frac{(1-\gamma)U_iX_i}{\mathrm{E}\left[X_i^2\right]}\right)^2\right]\right). \tag{10}$$

By (7), (8), (9) and (10),

$$\sqrt{n}\left(\beta_n-\beta\right)\to_d N\left(0,\mathrm{E}\left[\left(\frac{\gamma U_i}{\mathrm{E}\left[X_i\right]}+\frac{(1-\gamma)U_iX_i}{\mathrm{E}\left[X_i^2\right]}\right)^2\right]\right).$$

The asymptotic variance is

$$\mathrm{E}\left[\left(\frac{\gamma U_i}{\mathrm{E}\left[X_i\right]}+\frac{(1-\gamma)U_iX_i}{\mathrm{E}\left[X_i^2\right]}\right)^2\right]=\gamma^2\frac{\sigma^2}{\left(\mathrm{E}\left[X_i\right]\right)^2}+2\gamma(1-\gamma)\frac{\sigma^2}{\mathrm{E}\left[X_i^2\right]}+(1-\gamma)^2\frac{\sigma^2}{\mathrm{E}\left[X_i^2\right]},$$

where the equality follows from the law of iterated expectations.

(iii) The optimal γ minimizes the asymptotic variance. Note that $\left(\mathrm{E}\left[X_i\right]\right)^2\leq\mathrm{E}\left[X_i^2\right]$. Therefore, for all $\gamma\in[0,1]$,

$$\begin{aligned}\gamma^2\frac{\sigma^2}{\left(\mathrm{E}\left[X_i\right]\right)^2}+2\gamma(1-\gamma)\frac{\sigma^2}{\mathrm{E}\left[X_i^2\right]}+(1-\gamma)^2\frac{\sigma^2}{\mathrm{E}\left[X_i^2\right]} &\geq \left(\gamma^2+2\gamma(1-\gamma)+(1-\gamma)^2\right)\frac{\sigma^2}{\mathrm{E}\left[X_i^2\right]}\\ &= \frac{\sigma^2}{\mathrm{E}\left[X_i^2\right]}.\end{aligned}$$

The right hand side is attained by setting $\gamma=0$.

Problem 8. (16 Points) Suppose we observe an i.i.d. sample $\{(Y_i,D_i):i=1,...,n\}$ where the explanatory variable D_i is binary: $D_i\in\{0,1\}$ so that $D_i=D_i^2$. Suppose that Y_i is generated by $Y_i=\alpha+\beta_iD_i+U_i$, where the marginal effect β_i is random.

(i) Suppose that U_i is the error term independent from D_i and β_i. Also assume that β_i is independent from D_i. Show that the OLS estimator consistently estimates $\mathrm{E}\left[\beta_i\right]$.

(ii) Assume that $\mathrm{Cov}\left[U_i,D_i\right]\neq 0$ but a binary IV $Z_i\in\{0,1\}$ that is independent of (U_i,β_i) is available. Assume that D_i and Z_i are positively correlated. Give a real-life example to show that the model is plausible.

(iii) By using the law of iterated expectations, show that $\mathrm{E}\left[Y_iZ_i\right]=\mathrm{E}\left[Y_i\mid Z_i=1\right]\mathrm{E}\left[Z_i\right]$. Similarly,

$$\mathrm{E}\left[Y_i\mid Z_i=0\right]=\frac{\mathrm{E}\left[Y_i\left(1-Z_i\right)\right]}{\mathrm{E}\left[1-Z_i\right]}.$$

(iv) Show that the instrumental variable estimator $\widehat{\beta}$ converges in probability to

$$\beta=\frac{\mathrm{E}\left[Y_i\mid Z_i=1\right]-\mathrm{E}\left[Y_i\mid Z_i=0\right]}{\mathrm{E}\left[D_i\mid Z_i=1\right]-\mathrm{E}\left[D_i\mid Z_i=0\right]}.$$

Solution.

(i) Denote $\bar{\beta} = \mathrm{E}[\beta_i]$. Write

$$Y_i = \alpha + \beta_i D_i + U_i \iff Y_i = \alpha + \bar{\beta} D_i + U_i + \left(\beta_i - \bar{\beta}\right) D_i$$

and

$$Y_i = \alpha + \bar{\beta} D_i + V_i,$$

where in this model, the error term is $V_i = U_i + \left(\beta_i - \bar{\beta}\right) D_i$. We have

$$\mathrm{E}[V_i] = \mathrm{E}[U_i] + \mathrm{E}\left[\left(\beta_i - \bar{\beta}\right) D_i\right] = 0$$

and

$$\mathrm{E}[V_i D_i] = \mathrm{E}[U_i D_i] + \mathrm{E}\left[\left(\beta_i - \bar{\beta}\right) D_i\right] = 0.$$

Then OLS of Y_i against D_i consistently estimates $\bar{\beta}$.

(ii) Any example with a binary endogenous explanatory variable and a binary IV is fine.

(iii) By the law of iterated expectations and noticing that

$$Z_i \cdot \mathrm{E}[Y_i \mid Z_i] = \begin{cases} \mathrm{E}[Y_i \mid Z_i = 1] & \text{if } Z_i = 1 \\ 0 & \text{if } Z_i = 0, \end{cases}$$

we have

$$\begin{aligned} \mathrm{E}[Y_i Z_i] &= \mathrm{E}[\mathrm{E}[Y_i Z_i \mid Z_i]] = \mathrm{E}[Z_i \cdot \mathrm{E}[Y_i \mid Z_i]] \\ &= \Pr[Z_i = 1] \cdot \mathrm{E}[Y_i \mid Z_i = 1] + \Pr[Z_i = 1] \cdot 0 = \Pr[Z_i = 1] \cdot \mathrm{E}[Y_i \mid Z_i = 1] = \mathrm{E}[Z_i] \cdot \mathrm{E}[Y_i \mid Z_i = 1]. \end{aligned}$$

(iv) The IV estimator:

$$\widehat{\beta} = \frac{n^{-1} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)\left(Z_i - \overline{Z}\right)}{n^{-1} \sum_{i=1}^{n} \left(D_i - \overline{D}\right)\left(Z_i - \overline{Z}\right)} \to_p \frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(D_i, Z_i)}.$$

Then,

$$\begin{aligned} \mathrm{E}[Y_i \mid Z_i = 1] - \mathrm{E}[Y_i \mid Z_i = 0] &= \frac{\mathrm{E}[Y_i Z_i]}{\mathrm{E}[Z_i]} - \frac{\mathrm{E}[Y_i (1 - Z_i)]}{\mathrm{E}[1 - Z_i]} \\ &= \frac{\mathrm{E}[Y_i Z_i](1 - \mathrm{E}[Z_i]) - \mathrm{E}[Z_i](\mathrm{E}[Y_i] - \mathrm{E}[Y_i Z_i])}{\mathrm{E}[Z_i](1 - \mathrm{E}[Z_i])} = \frac{\mathrm{E}[Y_i Z_i] - \mathrm{E}[Z_i]\mathrm{E}[Y_i]}{\mathrm{E}[Z_i](1 - \mathrm{E}[Z_i])}. \end{aligned}$$

Note that since Z_i is binary,

$$\mathrm{E}[Z_i](1 - \mathrm{E}[Z_i]) = \mathrm{E}[Z_i] - (\mathrm{E}[Z_i])^2 = \mathrm{E}\left[Z_i^2\right] - (\mathrm{E}[Z_i])^2 = \mathrm{Var}[Z_i].$$

Then, it follows that

$$\mathrm{E}[Y_i \mid Z_i = 1] - \mathrm{E}[Y_i \mid Z_i = 0] = \frac{\mathrm{E}[Y_i Z_i] - \mathrm{E}[Z_i]\mathrm{E}[Y_i]}{\mathrm{E}[Z_i](1 - \mathrm{E}[Z_i])} = \frac{\mathrm{Cov}[Z_i, Y_i]}{\mathrm{Var}[Z_i]}.$$

Similarly,

$$\mathrm{E}[D_i \mid Z_i = 1] - \mathrm{E}[D_i \mid Z_i = 0] = \frac{\mathrm{Cov}[Z_i, D_i]}{\mathrm{Var}[Z_i]}.$$

Therefore,

$$\frac{\mathrm{Cov}(Y_i, Z_i)}{\mathrm{Cov}(D_i, Z_i)} = \frac{\mathrm{E}[Y_i \mid Z_i = 1] - \mathrm{E}[Y_i \mid Z_i = 0]}{\mathrm{E}[D_i \mid Z_i = 1] - \mathrm{E}[D_i \mid Z_i = 0]}.$$

Problem 9. (4 Points) The "no multicollinearity" condition is not satisfied by the difference-in-difference model $Y_i = \alpha + \gamma T_i + \lambda G_i + \delta D_i + U_i$ since $D_i = T_i \times G_i$. Is this statement true or false? Explain.

Solution. False. D_i is not a linear function of (T_i, G_i). Therefore, the "no multicollinearity" condition is not violated.

Problem 10. (4 points) Consider the simple regression model (with i.i.d. observations):

$$Y_i^* = \beta_0 + \beta_1 X_i + U_i.$$

Assume that $\mathrm{E}[U_i] = \mathrm{E}[X_i U_i] = 0$. However, instead of observing Y_i^*, we only observed $Y_i = Y_i^* + e_i$. We think of Y_i as some measurement of Y_i^* that is subject to error. Assume

$$\mathrm{E}[e_i] = \mathrm{E}[e_i U_i] = \mathrm{E}[Y_i^* e_i] = \mathrm{E}[X_i e_i] = 0.$$

Suppose we estimate the model using OLS with the observed Y_i in place of Y_i^*. Does the OLS consistently estimate β_1? Explain.

Solution. Yes. The OLS consistently estimate β_1. Write

$$Y_i - e_i = \beta_0 + \beta_1 X_i + U_i \iff Y_i = \beta_0 + \beta_1 X_i + U_i + e_i$$

and

$$Y_i = \beta_0 + \beta_1 X_i + V_i,$$

where in this model, the error term is $V_i = U_i + e_i$. Since $\mathrm{E}[V_i] = \mathrm{E}[V_i X_i] = 0$, OLS using Y_i in place of Y_i^* consistently estimates β_1.