2022년 2학기
계량경제학연구
담당교수: 이서정

Problem Set #3

- Due date/time: 12월 6일 23:59까지 eTL로 pdf파일 제출

X: 여 - 1
남 - 0

	X	Y (yr)
1	1	13
2	1	12
3	0	10
4	1	11
⋮	0	12
⋮	1	13
n	1	14

$E[Y|X=1]$

교과서 연습문제

1. 12.9
2. 12.12
3. 12.22 (Note: (l) is optional)
4. 13.2
5. 13.19
6. 13.27
7. Suppose that Z is a binary variable with $P(Z = 1) = p$. Show that

$$\text{cov}(Y, Z) = \{E[Y|Z = 1] - E[Y|Z = 0]\}p(1 - p)$$

8. Suppose that Z is a binary variable and $\sum_{i=1}^{n} Z_i \neq 0$ for any n. Show that

$$\overline{Y}_1 = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i} \xrightarrow{p} E[Y|Z = 1]$$

as $n \to \infty$.

Exercise 12.9 Consider the model $Y = X'\beta + e$ with $\mathbb{E}[e \mid Z] = 0$ with Y scalar and X and Z each a k vector. You have a random sample $(Y_i, X_i, Z_i : i = 1, ..., n)$.

(a) Assume that X is exogenous in the sense that $\mathbb{E}[e \mid Z, X] = 0$. Is the IV estimator $\widehat{\beta}_{\mathrm{iv}}$ unbiased?

(b) Continuing to assume that X is exogenous, find the conditional covariance matrix $\operatorname{var}\left[\widehat{\beta}_{\mathrm{iv}} \mid \boldsymbol{X}, \boldsymbol{Z}\right]$.

Exercise 12.12 Consider the structural equation and reduced form

$$
\begin{aligned}
Y &= \beta X^2 + e \\
X &= \gamma Z + u \\
\mathbb{E}[Ze] &= 0 \\
\mathbb{E}[Zu] &= 0
\end{aligned}
$$

with X^2 treated as endogenous so that $\mathbb{E}\left[X^2 e\right] \neq 0$. For simplicity assume no intercepts. Y, Z, and X are scalar. Assume $\gamma \neq 0$. Consider the following estimator. First, estimate γ by OLS of X on Z and construct the fitted values $\widehat{X}_i = \widehat{\gamma} Z_i$. Second, estimate β by OLS of Y_i on $\left(\widehat{X}_i\right)^2$.

(a) Write out this estimator $\widehat{\beta}$ explicitly as a function of the sample.

(b) Find its probability limit as $n \to \infty$.

(c) In general, is $\widehat{\beta}$ consistent for β? Is there a reasonable condition under which $\widehat{\beta}$ is consistent?

Exercise 12.22 You will replicate and extend the work reported in Acemoglu, Johnson, and Robinson (2001). The authors provided an expanded set of controls when they published their 2012 extension and posted the data on the AER website. This dataset is `AJR2001` on the textbook website.

(a) Estimate the OLS regression (12.86), the reduced form regression (12.87), and the 2SLS regression (12.88). (Which point estimate is different by 0.01 from the reported values? This is a common phenomenon in empirical replication).

(b) For the above estimates calculate both homoskedastic and heteroskedastic-robust standard errors. Which were used by the authors (as reported in (12.86)-(12.87)-(12.88)?)

(c) Calculate the 2SLS estimates by the Indirect Least Squares formula. Are they the same?

(d) Calculate the 2SLS estimates by the two-stage approach. Are they the same?

(e) Calculate the 2SLS estimates by the control variable approach. Are they the same?

(f) Acemoglu, Johnson, and Robinson (2001) reported many specifications including alternative regressor controls, for example *latitude* and *africa*. Estimate by least squares the equation for log-GDP adding *latitude* and *africa* as regressors. Does this regression suggest that *latitude* and *africa* are predictive of the level of GDP?

(g) Now estimate the same equation as in (f) but by 2SLS using log(mortality) as an instrument for *risk*. How does the interpretation of the effect of *latitude* and *africa* change?

(h) Return to our baseline model (without including *latitude* and *africa*). The authors' reduced form equation uses log(*mortality*) as the instrument, rather than, say, the level of mortality. Estimate the reduced form for risk with *mortality* as the instrument. (This variable is not provided in the dataset so you need to take the exponential of log(mortality).) Can you explain why the authors preferred the equation with log(*mortality*)?

(i) Try an alternative reduced form including both log(*mortality*) and the square of log(*mortality*). Interpret the results. Re-estimate the structural equation by 2SLS using both log(*mortality*) and its square as instruments. How do the results change?

(j) For the estimates in (i) are the instruments strong or weak using the Stock-Yogo test?

(k) Calculate and interpret a test for exogeneity of the instruments.

(l) Estimate the equation by LIML using the instruments log(*mortality*) and the square of log(*mortality*).

Exercise 13.2 Take the model $Y = X'\beta + e$ with $\mathbb{E}[e \mid Z] = 0$. Let $\widehat{\beta}_{\text{gmm}}$ be the GMM estimator using the weight matrix $\boldsymbol{W}_n = \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$. Under the assumption $\mathbb{E}\left[e^2 \mid Z\right] = \sigma^2$ show that

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \underset{d}{\longrightarrow} \mathrm{N}\left(0, \sigma^2\left(\boldsymbol{Q}'\boldsymbol{M}^{-1}\boldsymbol{Q}\right)^{-1}\right)$$

where $\boldsymbol{Q} = \mathbb{E}\left[ZX'\right]$ and $\boldsymbol{M} = \mathbb{E}\left[ZZ'\right]$.

Exercise 13.19 You want to estimate $\mu = \mathbb{E}[Y]$ under the assumption that $\mathbb{E}[X] = 0$, where Y and X are scalar and observed from a random sample. Find an efficient GMM estimator for μ.

Exercise 13.27 Continuation of Exercise 12.22, based on the empirical work reported in Acemoglu, Johnson, and Robinson (2001).

data

(a) Re-estimate the model estimated in part (j) by efficient GMM. Use the 2SLS estimates as the first-step for the weight matrix and then calculate the GMM estimator using this weight matrix without further iteration. Report the estimates and standard errors.

(b) Calculate and report the J statistic for overidentification.

(c) Compare the GMM and 2SLS estimates. Discuss your findings.