COMM8102, T2 2021

Solutions to Problem Set #4

Q1. Exercise 12.9

The dimension of Z and X are the same $(k \times 1)$. The IV estimator is

$$\widehat{\beta}_{iv} = \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} Z_i Y_i\right).$$

(a) If $E[e|Z, X] = 0$, then

$$\begin{aligned} E[\widehat{\beta}_{iv}|Z_i, X_i] =& E\left[\left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} Z_i Y_i\right) |Z_i, X_i\right] \\ =& \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} Z_i E[Y_i|Z_i, X_i] \\ =& \beta + \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} Z_i E[e_i|Z_i, X_i] = \beta. \end{aligned}$$

By the iterated law of expectations,

$$E[E[\widehat{\beta}_{iv}|Z_i, X_i]] = E[\widehat{\beta}_{iv}] = \beta.$$

(b) Since

$$\widehat{\beta}_{iv} - \beta = \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} Z_i e_i,$$

using the assumption of random sampling, the conditional covariance matrix can be found as

$$Var[\widehat{\beta}_{iv}|Z_i, X_i] = Var[\widehat{\beta}_{iv} - \beta|Z_i, X_i] = \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1} \sum_{i=1}^{n} Z_i E[e_i^2|Z_i, X_i] Z_i' \left(\sum_{i=1}^{n} Z_i X_i'\right)^{-1}.$$

Q2. Exercise 12.12

(a) The first step is to estimate γ by OLS of X_i on Z_i: $\widehat{\gamma} = \frac{\sum_{i=1}^n X_i Z_i}{\sum_{i=1}^n Z_i^2}$ and construct $\widehat{X}_i = \widehat{\gamma} Z_i$. The second step is to estimate β by OLS of Y_i on $\left(\widehat{X}_i\right)^2$:

$$\widehat{\beta} = \frac{\sum_{i=1}^n \left(\widehat{X}_i\right)^2 Y_i}{\sum_{i=1}^n \left(\widehat{X}_i\right)^4} = \frac{\widehat{\gamma}^2 \sum_{i=1}^n Z_i^2 Y_i}{\widehat{\gamma}^4 \sum_{i=1}^n Z_i^4} = \left(\frac{\sum_{i=1}^n Z_i^2}{\sum_{i=1}^n X_i Z_i}\right)^2 \frac{\sum_{i=1}^n Z_i^2 Y_i}{\sum_{i=1}^n Z_i^4}.$$

(b) By replacing Y_i into $\widehat{\beta}$,

$$\widehat{\beta} = \left(\frac{\sum_{i=1}^n Z_i^2}{\sum_{i=1}^n X_i Z_i}\right)^2 \frac{\sum_{i=1}^n Z_i^2 Y_i}{\sum_{i=1}^n Z_i^4} = \left(\frac{\sum_{i=1}^n Z_i^2}{\sum_{i=1}^n X_i Z_i}\right)^2 \frac{\sum_{i=1}^n Z_i^2 X_i^2 \beta + \sum_{i=1}^n Z_i^2 e_i}{\sum_{i=1}^n Z_i^4}$$

By the WLLN and CMT,

$$\begin{aligned}
\widehat{\beta} &= \left(\frac{\frac{1}{n}\sum_{i=1}^n Z_i^2}{\frac{1}{n}\sum_{i=1}^n X_i Z_i}\right)^2 \frac{\frac{1}{n}\sum_{i=1}^n Z_i^2 X_i^2 \beta + \frac{1}{n}\sum_{i=1}^n Z_i^2 e_i}{\frac{1}{n}\sum_{i=1}^n Z_i^4} \\
&\xrightarrow{p} \frac{(E[Z_i^2])^2}{(E[X_i Z_i])^2} \frac{E[Z_i^2 X_i^2]\beta + E[Z_i^2 e_i]}{E[Z_i^4]} = \frac{(E[Z_i^2])^2}{E[Z_i^4]} \frac{E[Z_i^2 X_i^2]}{(E[Z_i X_i])^2}\beta + \frac{(E[Z_i^2])^2}{E[Z_i^4]} \frac{E[Z_i^2 e_i]}{(E[Z_i X_i])^2}.
\end{aligned}$$

(c) The above result shows that $\widehat{\beta}$ is not consistent for β in general. But when will it be consistent? First, we would like to have

$$\frac{(E[Z_i^2])^2}{E[Z_i^4]} \frac{E[Z_i^2 X_i^2]}{(E[Z_i X_i])^2} = 1. \tag{1}$$

Note that $E[Z_i X_i] = \gamma E[Z_i^2]$ from the reduced form equation. Thus, the above equality becomes

$$\gamma^2 = \frac{E[Z_i^2 X_i^2]}{E[Z_i^4]} = \delta$$

where δ is the projection coefficient of the reduced form equation

$$X_i^2 = \delta Z_i^2 + v_i, \; E[Z_i^2 v_i] = 0.$$

In words, (1) would hold if the covariance between X_i and Z_i is the same as the covariance between X_i^2 and Z_i^2. This can happen if both X_i and Z_i are binary, but otherwise would not hold in general. Secondly, we would like to have

$$\frac{(E[Z_i^2])^2}{E[Z_i^4]} \frac{E[Z_i^2 e_i]}{(E[Z_i X_i])^2} = 0. \tag{2}$$

Since $\gamma \neq 0$ and $E[Z_i^4] \neq 0$ (otherwise $Z_i = 0$ and the questions becomes trivial), this is equivalent to

$$E[Z_i^2 e_i] = 0. \tag{3}$$

This would hold if $E[e_i|Z_i] = 0$ because

$$E[Z_i^2 e_i] = E[E[Z_i^2 e_i|Z_i]] = E[Z_i^2 E[e_i|Z_i]] = 0.$$

In sum, $\widehat{\beta}$ is consistent for β if (i) $\gamma^2 = \delta$ and (ii) $E[e_i|Z_i] = 0$.

Q3.

First consider $(D_i, Z_i) = (1, 1)$. The person cannot be a never-taker ($D_i(1) = D_i(0) = 0$) so she is either always-taker or complier. We don't know if she takes the treatment because of the instrument so her exact type cannot be determined. Likewise, $(D_i, Z_i) = (0, 0)$ is either never-taker or complier but we cannot further pin down. On the other hand, if $(D_i, Z_i) = (0, 1)$, then the person is a never-taker (because $D_i = 0$ means she is not an always-taker, and plus $Z_i = 1$ excludes a possibility of being a complier too). Similarly, $(D_i, Z_i) = (1, 0)$ is an always-taker.

Q4. Exercise 13.10

According to the model, $e = Y - m(X, \beta)$. Assume that the observations, $i = 1, 2, ..., n$, are i.i.d. The moment condition is

$$E[Z_i(Y_i - m(X_i, \beta_0))] = 0$$

for a unique β_0 (I use β_0 to emphasize that this is a unique value, rather than a generic one). The moment function is

$$g_i(\beta) = Z_i(Y_i - m(X_i, \beta))$$

and the sample mean of the moment function is

$$\overline{g}(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta).$$

Note that $g_i(\beta)$ and $\overline{g}(\beta)$ are $l \times 1$ vectors. The one-step GMM estimator using the identity matrix as the weight matrix is

$$\widehat{\beta}_1 = \arg\min_{\beta \in \mathcal{B}} \overline{g}(\beta)'\overline{g}(\beta),$$

where β is a given parameter space (you can think of this as a set of possible values of β). The one-step GMM estimator $\widehat{\beta}_1$ is consistent for β_0. Using $\widehat{\beta}_1$, we construct the efficient weight matrix as

$$\widehat{W}(\widehat{\beta}_1) = \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\widehat{\beta}_1) g_i(\widehat{\beta}_1)'\right)^{-1}$$

which is the $l \times l$ matrix. This is an efficient weight matrix because its probability limit is the inverse of the asymptotic covariance matrix of the moment function $g_i(\beta_0)$: $\Omega = E[g_i(\beta_0)g_i(\beta_0)']$. Finally the efficient two-step GMM can be obtained by

$$\widehat{\beta}_2 = \arg\min_{\beta \in \mathcal{B}} \overline{g}(\beta)'\widehat{W}(\widehat{\beta}_1)\overline{g}(\beta).$$

This two-step GMM estimator can be further iterated until convergence, and asymptotically all the GMM estimators beyond the initial step are efficient under the correct specification of the moment condition.

Q5. Exercise 13.24

(a) The model is $Y = \theta + e$, where both Y and θ are scalars. The moment function can be formed using the information $E[Xe] = 0$ as

$$g_i(\theta) = X_i(Y_i - \theta).$$

The dimension of $g_i(\theta)$ is k. Since this is linear in θ, we can find a closed-form solution for the GMM estimator. The first derivative of $g_i(\theta)$ w.r.t. θ is

$$G_i = -X_i.$$

Let $\overline{G} = \frac{1}{n}\sum_{i=1}^n G_i = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$ and $\overline{g}(\theta) = \frac{1}{n}\sum_{i=1}^n g_i(\theta) = \frac{1}{n}\sum_{i=1}^n X_iY_i - \overline{X}\theta$. Let $\overline{XY} = \frac{1}{n}\sum_{i=1}^n X_iY_i$. The one-step GMM criterion using the identity matrix is the solution to the first-order condition:

$$0 = \overline{G}'\overline{g}(\widehat{\theta}_1) = \overline{X}'(\overline{XY} - \overline{X}\widehat{\theta}_1).$$

Thus,

$$\widehat{\theta}_1 = \frac{\overline{X}'\overline{XY}}{\overline{X}'\overline{X}}.$$

The efficient weight matrix is

$$\widehat{W}(\widehat{\theta}_1) = \left(\frac{1}{n}\sum_{i=1}^n g_i(\widehat{\beta}_1)g_i(\widehat{\beta}_1)'\right)^{-1} = \left(\frac{1}{n}\sum_{i=1}^n X_iX_i'(Y_i - \widehat{\theta}_1)^2\right)^{-1}.$$

Or you can use the centered form

$$\widehat{W}(\widehat{\theta}_1) = \left(\frac{1}{n}\sum_{i=1}^n (g_i(\widehat{\beta}_1) - \overline{g}(\widehat{\theta}_1))(g_i(\widehat{\beta}_1) - \overline{g}(\widehat{\theta}_1))'\right)^{-1}.$$

In some cases, the centered form can eliminate θ, and this can simplify the calculation of the efficient GMM. The efficient two-step GMM solves the first-order condition

$$0 = \overline{G}'\widehat{W}(\widehat{\theta}_1)\overline{g}(\widehat{\theta}_2) = \overline{X}'\widehat{W}(\widehat{\theta}_1)(\overline{XY} - \overline{X}\widehat{\theta}_2)$$

and thus

$$\widehat{\theta}_2 = \frac{\overline{X}'\widehat{W}(\widehat{\theta}_1)\overline{XY}}{\overline{X}'\widehat{W}(\widehat{\theta}_1)\overline{X}}.$$

(b) Since the moment function is a $k \times 1$ vector (assume $k > 1$) and the parameter θ is a scalar, the model is over-identified.

(c) The over-identification test statistic is the efficient GMM criterion (multiply n to have the correct rate of convergence):

$$J(\widehat{\theta}_2) = n\overline{g}(\widehat{\theta}_2)'\widehat{W}(\widehat{\theta}_1)\overline{g}(\widehat{\theta}_2).$$

This is what most statistical packages report. But it uses the weight matrix evaluated at the one-step GMM. Instead, you might use

$$J^*(\widehat{\theta}_2) = n\overline{g}(\widehat{\theta}_2)'\widehat{W}(\widehat{\theta}_2)\overline{g}(\widehat{\theta}_2).$$

Q6. Exercise 12.25 (a)-(d)

(a) We can exactly replicate the results in Table 12.2 and the 2SLS regression result (2SLS(a)) in Table 12.1. We use the heteroskedasticity-robust standard errors.

(b) By including *near2c* in the reduced form as an additional IV. It does not change the results much. The first stage R^2 remains the same at 0.476. The point estimate is 0.068 (0.074) so it is statistically insignificant at the 5% level. The returns to schooling coefficient (second stage) changed from 0.160 (0.040) to 0.171 (0.041).

(c) The coefficient estimates of the two added interactions are 1.483 (0.200) and -1.554 (0.350), respectively. They are statistically significant at the 5% level. Interpretation: the age of an individual who grew up near 4-yr public college ($near4a = 1$) is a strong predictor of the individual's education level. By including both interaction terms we can capture the nonlinear effect of age of those who grew up near 4-yr public college on education.

(d) The return to schooling estimate is now 0.083 (0.006), almost halved compared with the previous point estimates. In addition, the standard error is much smaller, meaning that the estimate is very accurate. The reduced s.e. is due to the better fit of the first stage ($R^2 = 0.65$).

Q7. Exercise 13.28. In addition, provide the iterated GMM results.

(a) The efficient two-step and iterated GMM estimates are almost identical to the 2SLS. The estimated returns to schooling is 0.162 (0.041).

(b) Again, the efficient two-step and iterated GMM estimates are similar to the 2SLS. The estimated returns to schooling is 0.084 (0.006).

(c) For (a), the overidentification test statistic is 0.88 and the p-value is 0.35. Since $l - k = 1$, we use $\chi^2(1)$. There is no significant evidence of model misspecification. For (b), the test statistic is 10.60 and the p-value is 0.014 based on $\chi^2(3)$. We reject the null hypothesis that the moment condition is correctly specified at the 5% level. Meaning: We use four IVs, $nearc4a$, $nearc4b$, $nearc4a \times age$, and $nearc4a \times age^2/100$. At least one of them may not be valid (i.e., violates the exogeneity condition). Since the test for (a) is not significant, this suggests that the two added IVs might not be valid.