ECON 710A - Problem Set 3

Alex von Hafften*

2/16/2020

1. Let $(Y, X', Z')'$ be a random vector such that $Y = X'\beta + U$, $E[U|Z] = 0$ where $E[ZX']$ is invertible and $E[Y^4 + ||X||^4 + ||Z||^4] < \infty$. Also, let $\{(Y_i, X_i', Z_i')'\}_{i=1}^n$ be a random sample from the distribution of $(Y, X', Z')'$. We showed in lecture 4 that $\sqrt{n}(\hat{\beta}^{IV} - \beta) \to_d N(0, \Omega)$, $\Omega = E[ZX']^{-1}E[ZZ'U^2]E[XZ']^{-1}$. Now suppose that $X = (X_1, X_2')'$, $Z = (Z_1, X_2')'$, and $E[U^2|Z] = \sigma_U^2$ and let Ω_{11} be the upper left entry in Ω.

(i) Show that $E[ZX']$ is invertible iff $E[ZZ']$ is invertible and $\pi_1 \neq 0$ where $(\pi_1, \pi_2')' = E[ZZ']^{-1}E[ZX_1]$.

Observe that

$$E[ZX'] = E[\begin{pmatrix} Z_1 \\ X_2 \end{pmatrix} \begin{pmatrix} X_1 & X_2' \end{pmatrix}] = \begin{pmatrix} E[Z_1X_1] & E[Z_1X_2'] \\ E[X_2X_1] & E[X_2X_2'] \end{pmatrix}$$

$$E[XX'] = E[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{pmatrix} X_1 & X_2' \end{pmatrix}] = \begin{pmatrix} E[X_1X_1] & E[X_1X_2'] \\ E[X_2X_1] & E[X_2X_2'] \end{pmatrix}$$

($\Rightarrow$)

$E[Z_1X_1]$ is scalar, so it is trivially invertible if $E[Z_1X_1] \neq 0$. By the block inversion formula, $E[X_2X_2'] - E[X_2X_1]E[Z_1Z_1]^{-1}E[Z_1X_2']$ must also be invertible.

....

($\Leftarrow$)

...

(ii) Show that $\Omega_{11} = \frac{\sigma_U^2}{E[\tilde{Z}_1^2]\pi_1^2}$ where $\tilde{Z}_1 = Z_1 - X_2'E[X_2X_2']^{-1}E[X_2Z_1]$.

...

$$\Omega = E[ZX']^{-1}E[ZZ'U^2]E[XZ']^{-1} = E[ZX']^{-1}E[ZZ'U^2]E[XZ']^{-1} = \begin{pmatrix} E[Z_1X_1] & E[Z_1X_2'] \\ E[X_2X_1] & E[X_2X_2'] \end{pmatrix}^{-1}$$

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Sarah Bass, Emily Case, Danny Edgel, and Katherine Kwok.

2.

3. Consider the data from Angrist and Krueger (1991) provided on the course website and the following linear model for $log(wage)$ as a function of educationand additional control variables: $log(wage) = \beta_0 + educ \cdot \beta_1 + \sum_{t=31}^{39} 1\{yob = t\}\beta_t + \sum_{s=1}^{50} 1\{sob = s\}\gamma_s + U$, where yob is year of birth and sob is state of birth. As instruments for $educ$ consider three instruments: $1\{qob = 2\}, 1\{qob = 3\}, 1\{qob = 4\}$ where qob is quarter of birth. Using matrix algebra and your preferred statistical software, write code that loads the data and computes the 2SLS estimate of β_1 and the heteroskedasticity robust standard error stemming from the variance estimator formula (12.40) on page 354 of Bruce Hansen's textbook. The solution to this exercise should include code and the two numbers produced by it.

```r
data <- read_csv("AK91.csv", col_types = "ddddd")

n <- nrow(data)

# prep variables
y <- data$lwage
x_1 <- data$educ
controls <- cbind(rep(1, n),
                  to.dummy(data$yob, prefix = "yob")[,1:9],
                  to.dummy(data$sob, prefix = "sob")[,1:50])
instruments <- to.dummy(data$qob, prefix = "qob")[,2:4]

z <- cbind(instruments, controls)
x <- cbind(x_1, controls)

k <- ncol(x)
l <- ncol(z)

# Estimating beta_2sls using 12.29 in Hansen
beta_2sls <- inv(t(x) %*% z %*% inv(t(z) %*% z) %*% t(z) %*% x) %*%
  t(x) %*% z %*% inv(t(z) %*% z) %*% t(z) %*% y

# Estimating heteroskedastic robust standard errors using 12.40 in Hansen
q_zz <- t(z) %*% z / n
q_xz <- t(x) %*% z / n
q_zx <- t(q_xz)

e_hat <- y - x %*% beta_2sls

omega <- matrix(rep(0, times = l*l), nrow = l, ncol = l)
for (i in 1:n) omega <- omega + z[i, ] %*% t(z[i, ]) * e_hat[i]^2 / n

varcov <- inv(q_xz %*% inv(q_zz) %*% q_zx) %*% (q_xz %*% inv(q_zz) %*% omega %*%
            inv(q_zz) %*% q_zx) %*% inv(q_xz %*% inv(q_zz) %*% q_zx)

print(beta_2sls[1])
print(sqrt(varcov[1,1]))
```