SDPE Econometrics I, Spring 2024

Problem Set 4: Marek Chadim[1]

---

1. Rewrite the covariance matrix using Law of iterated expectations and $\mathbb{E}(ee'|X) = \Sigma$

$$\Omega = \mathbb{E}\left(\bar{X}_i'\Sigma_i\bar{X}_i\right) = \mathbb{E}\left(\mathbb{E}\left(\bar{X}_i'\Sigma_i\bar{X}_i|X\right)\right) = \mathbb{E}\left(\bar{X}_i'\mathbb{E}[\Sigma_i|X]\bar{X}_i\right) = \mathbb{E}\left(\bar{X}_i'\Sigma\bar{X}_i\right)$$

2. (a) The two equations can be estimated by least squares. For $j - 1, 2$:

$$\hat{\beta}_j = \left(\frac{1}{n}\sum_{i=1}^{n} X_{ji}X_{ji}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} X_{ji}y_{ji}.$$

We can alternatively write this estimator using the systems notation. $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \left(\bar{X}'\bar{X}\right)^{-1}\left(\bar{X}'y\right) = \left(\sum_{i=1}^{n}\bar{X}_i'\bar{X}_i\right)^{-1}\left(\sum_{i=1}^{n}\bar{X}_i'y_i\right).$$

(b) Center the estimator as

$$\hat{\beta} - \beta = \left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'e.$$

Since the vector $\bar{X}_i'e_i$. is i.i.d. across i and has mean zero under $E\left[X_je_j\right] = 0$, the central limit theorem implies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{X}_i'e_i \xrightarrow{d} N(0, \Omega)$$

Applied to the centered and normalized estimator, we obtain the asymptotic distribution:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, V_\beta).$$

with

$$V_\beta = Q^{-1}\Omega Q^{-1}$$

where

$$Q = I_m \otimes \mathbb{E}[XX'],\ \Omega = E[ee' \otimes XX']$$

since the regressors are common.

(c) Construct the common regressors estimator for the variance of $\hat{\beta}$

$$\hat{V}_{\hat{\beta}} = (I_m \otimes (X_i'X_i)^{-1})(\sum_{i=1}^{n}(e_i'e_i \otimes X_iX_i'))(I_m \otimes (X_i'X_i)^{-1}),$$

Perform a Wald test by comparing χ_k^2 critical with the test statistic

$$W = (R'\hat{\beta})'(R'\hat{V}_{\hat{\beta}}R)^{-1}(R'\hat{\beta})$$

where

$$R = \frac{\partial}{\partial\beta}r(\beta)' = \begin{bmatrix} I_k \\ -I_k \end{bmatrix}.$$

[1] 42624@student.hhs.se

3. Nonlinear regression model written as $y = m(X) + e$ is a parametric regression function which is nonlinear in the parameters . The given regression function is linear in (β_0, β_1) but nonlinear in λ. Yet for $\lambda \neq 0$

$$y = (\lambda(\beta_0 + \beta_1 X + e) + 1)^{\frac{1}{\lambda}}$$

is not a nonlinear regression model because $y \neq m(X) + e$.

4.
```
// Stata (a) (b) (d)
use Nerlove1963.dta,clear, gen lC = log(cost),
gen lQ = log(output), gen lPL = log(Plabor),
gen lPF = log(Pfuel), gen lPK = log(Pcapital)
// 10 observations of log Q both below and above
gsort +lQ, list lQ in 11                        3.218876
gsort -lQ. list lQ in 11                        8.880863
nl(lC={b1}+{b2}*lQ +{b3}*(lPL+lPK+lPF)
+{b4}*(lQ/(1+exp(-(lQ-{gamma}))))),
initial(b1 1 b2 1 b3 1 b4 1 gamma 7) r
------------+-------------------------
         /b1 |   -5.32109   .5025214
         /b2 |   .4377937   .1013331
         /b3 |   .3707369   .0449295
         /b4 |   .2240117   .0555175
      /gamma |   6.875142   .3649995

// R c)
data <- read.table("Nerlove1963.txt",header=TRUE)
y <- matrix(log(data$Cost),ncol=1)
x <- matrix(log(data$output),ncol=1)
lPL <- matrix(log(data$Plabor),ncol=1)
lPK <- matrix(log(data$Pcapital),ncol=1)
lPF <- matrix(log(data$Pfuel),ncol=1)
z <- lPL+lPK+lPF
x <- as.matrix(cbind(matrix(1,nrow(lC),1),x,z))
transf <- function(gamma) G <- (lQ/(1+exp(gamma - lQ)))
SSE <- function(gamma) {G <- transf(gamma)
  X <- cbind(x,G)
  b <- solve(crossprod(X,X),crossprod(X,y))
  e <- y - X%*%b
  sse <- mean(e^2)
  return(sse)}
BC <- optimize(SSE,c(3,9))
gamma <- BC$minimum
G <- transf(gamma)
X <- cbind(x,G)
beta <- solve(crossprod(X,X),crossprod(X,y))
print(rbind(beta,gamma))
                                  -5.3210907
                                   0.4377923
                                   0.3707373
                                   0.2240123
                            gamma  6.8751278
```

5. Take expectations of the structural equation given $D = 1$ and $D = 0$, respectively

$$E[Y|D = 1] = E[Z|D = 1]\beta \tag{1}$$
$$E[Y|D = 0] = E[Z|D = 0]\beta. \tag{2}$$

Subtract and divide to obtain an expression for the slope coefficient

$$\beta = \frac{E[Y|D = 1] - E[Y|D = 0]}{E[Z|D = 1] - E[Z|D = 0]}.$$

Define the group means

$$\bar{Y}_1 = \frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i}, \bar{Y}_0 = \frac{\sum_{i=1}^n (1 - D_i) Y_i}{\sum_{i=1}^n (1 - D_i)},$$
$$\bar{Z}_1 = \frac{\sum_{i=1}^n D_i Z_i}{\sum_{i=1}^n D_i}, \bar{Z}_0 = \frac{\sum_{i=1}^n (1 - D_i) Z_i}{\sum_{i=1}^n (1 - D_i)},$$

and the moment estimator

$$\hat{\beta} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{Z}_1 - \bar{Z}_0}.$$

note $\hat{\beta}$ equals the standard IV estimator

$$\hat{\beta}_{\text{IV}} = \frac{\sum_{i=1}^n D_i (Y_i - \bar{Y})}{\sum_{i=1}^n D_i (Z_i - \bar{Z})} = \frac{\bar{Y}_1 - \bar{Y}}{\bar{Z}_1 - \bar{Z}}$$

since

$$\bar{Y}_1 - \bar{Y} = \bar{Y}_1 - (\frac{1}{n}\sum_{i=1}^n D_i \bar{Y}_1 + \frac{1}{n}\sum_{i=1}^n (1 - D_i)\bar{Y}_O) = (1 - \bar{D})(\bar{Y}_1 - \bar{Y}_0)$$

similarly

$$\bar{Z}_1 - \bar{Z} = (1 - \bar{D})(\bar{Z}_1 - \bar{Z}_0)$$

thus

$$\hat{\beta}_{\text{IV}} = \frac{(1 - \bar{D})(\bar{Y}_1 - \bar{Y}_0)}{(1 - \bar{D})(\bar{Z}_1 - \bar{Z}_0)} = \hat{\beta}.$$

A model without an intercept is overidentified. From (1), (2) additional estimators are

$$\hat{\beta}_{IV1} = \frac{\bar{Y}_1}{\bar{Z}_1} \quad \text{and} \quad \hat{\beta}_{IV2} = \frac{\bar{Y}_0}{\bar{Z}_0}.$$

6. Define

$$M = I_n - P = I_n - X(X'X)^{-1}X'$$

where, I_n is the $n \times n$ identity matrix. Note that

$$MX = (I_n - P)X = X - PX = X - X = 0.$$

Substituting $Y = X\beta + e$ into $\hat{e} = MY$ and using $MX = 0$

$$\hat{e} = MY = M(X\beta + e) = Me.$$

Since $MM = M$

$$\hat{e}'\hat{e} = e'MMe = e'Me = e'e - e'X(X'X)^{-1}X'e$$

Write the IV residual as

$$\begin{aligned}\tilde{e} &= y - X\tilde{\beta} \\ &= y - X(Z'X)^{-1}Z'y \\ &= X\beta + e - X(Z'X)^{-1}Z'(X'\beta + e) \\ &= e - X(Z'X)^{-1}Z'e \\ &= (I_n - X(Z'X)^{-1}Z')e.\end{aligned}$$

Then

$$\begin{aligned}\tilde{e}'\tilde{e} &= e'(I_n - X(Z'X)^{-1}Z')(I_n - X(Z'X)^{-1}Z')e \\ &= e'e - e'Z(X'Z)^{-1}X'e - e'X(Z'X)^{-1}Z'e + e'Z(X'Z)^{-1}X'X(Z'X)^{-1}Z'e.\end{aligned}$$

Since $E[e_iX_i'] \neq 0$ $\text{plim}\,(\hat{e}'\hat{e}) < e'e$ while $E[e_iZ_i'] = 0$ implies $\text{plim}\,(\tilde{e}'\tilde{e}) = e'e$ therefore OLS fits better than IV in the sense that $\sum_i \tilde{e}_i^2 \geq \sum_i \hat{e}_i^2$ even in large samples.

7. (a) Since the instrumental variable is exogenous

$$\begin{aligned}E[X\nu] &= E[(\Gamma Z + u)\nu] = \Gamma E[Z\nu] + E[u\nu] = \Gamma E[Z\nu] \\ &= \Gamma E[Z(e - u'\gamma)] = \Gamma(E[Ze] - E[Zu']\gamma) = 0.\end{aligned}$$

(b) By the CLT the asymptotic distribution of the control function estimator is

$$\sqrt{n}\begin{bmatrix}\hat{\beta} - \beta \\ \hat{\gamma} - \gamma\end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V),$$

where

$$V = \begin{bmatrix}V_{\beta\beta} & V_{\beta\gamma} \\ V_{\gamma\beta} & V_{\gamma\gamma}\end{bmatrix}$$

and

$$\begin{aligned}V_{\beta\beta} &= (\Gamma' E[ZZ']\Gamma)^{-1}(\Gamma' E[ZZ'e^2]\Gamma)(\Gamma' E[ZZ']\Gamma)^{-1} \\ V_{\beta\gamma} &= E[uu']^{-1}(E[uZ'e\nu]\Gamma)(\Gamma' E[ZZ']\Gamma)^{-1} \\ V_{\gamma\gamma} &= E[uu']^{-1}E[uu'\nu^2]E[uu']^{-1}.\end{aligned}$$

.

8. (a) By the LIE $E[Xe] = E[XE[e|X]]$, thus

$$E[X^2e] = E[X^2E[e|X]] \neq 0$$

and X^2 should be treated as endogenous.

(b) Denote $(1, X, X^2)'$ the $k \times 1$ regressor vector and $(1, Z, Z^2)'$ the $l \times 1$ instrumental variable vector. The model is just-identified as $l = k = 3$.

(c)

$$\begin{bmatrix}1 \\ X_1 \\ X^2\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 \\ \gamma_0 & \gamma_1 & 0 \\ \delta_0 & \delta_1 & \delta_2\end{bmatrix}\begin{bmatrix}1 \\ Z_1 \\ Z^2\end{bmatrix} + \begin{bmatrix}u_0 \\ u_1 \\ u_2\end{bmatrix}$$

The reduced form equation for X^2 is

$$X^2 = (1, Z, Z^2)'(\delta_0, \delta_1, \delta_2) + u_2$$

δ_2, γ_1 must be distinct from zero for the parameters in (*) to be identified.

9. (a)

```
use AJR2001.dta                                   Coeff      Std. err.
reg loggdp risk                                   .516187    .0625186
reg risk logmort0                                −.6132892   .1269412
ivregress 2sls loggdp (risk=logmort0)             .9294897   .1536318
```

The 2SLS point estimate is different by 0.01 from the reported value random 0.94.

(b) Author's homoskedastic standard errors above, heteroskedastic-robust below

```
reg loggdp risk, r                                           .051101
reg risk logmort0, r                                         .1517849
ivregress 2sls loggdp (risk=logmort0), r                     .1700872
```

(c)

```
reg loggdp logmort0                          Coeff = −.570046
```

$$\frac{-.570046}{-.6132892} = 0.9294897089$$

(d)

```
reg risk logmort0, predict riskhat, xb             riskhat
reg loggdp riskhat                           Coeff = .9294897
```

(e)

```
reg risk logmort0. predict u, residual             risk
reg loggdp risk u                            Coeff = .9294897
```

(f)

```
reg loggdp risk latitude africa
    loggdp | Coefficient  Std. err.  P>|t|    [95% conf. interval]
  latitude |   1.382463   .6440401   0.036    .0941905    2.670735
    africa |  −.7232696   .1712967   0.000   −1.065914   −.38062
```

Both latitude and africa are predictive of log GDP at conventional significance levels.

(g)

```
ivregress 2sls loggdp latitude africa (risk=logmort0)
    loggdp |  Coefficient Std. err.  P>|t|    [95% conf. interval]
  latitude |  −.0553109  1.161701   0.962   −2.332203   2.221581
    africa |  −.3479258   .3062581  0.256   −.9481806    .252329
```

The IV regression coefficients on latitude and africa are statistically insignificant.

(h)

```
gen mort0 = exp(logmort0), reg risk mort0
                                loggdp|    Coeff      Std. err.
                                 risk |  −.0007862    .0003819
```

The authors preferred the equation with *logmort*0 since it has a stronger first stage.

(i)

```
gen logmort2 = logmort0*logmort0
ivregress 2sls loggdp (risk=logmort0 logmort2)
                                loggdp|    Coeff      Std. err.
                                 risk |   .7722554    .1130303
estat firststage
                       R-sq. 0.3766     F(2,61) 18.4227
```

Including both *logmort*0 and the square of *logmort*0 increases the fit of the first stage and precision of the second stage. The effect of risk on loggdp reduces to 0.77.

(j)

```
estat overid, forcenonrobust
Sargan chi2(1)           =    5.13535   (p = 0.0234)
```

Rejecting the null at $p = 0.02$ gives mild evidence against the model.