SDPE Econometrics I, Spring 2024

Problem Set 4: Marek Chadim[1]

---

1. (a) Lindeberg-Levy CLT applies since the statistics are mutually independent. The asymptotic distribution is
$$N(0, V_{\beta_1} + V_{\beta_2})$$
where $V_{\beta_i} = (Q_{XX})^{-1}\Omega(Q_{XX})^{-1}$, $Q_{XX} = E[X_i X_i']$, $\Omega = E[X_i X_i' e_i^2]$, $i = 1, 2$.

   (b) $\hat{\beta}_2/\text{se}(\hat{\beta}_2) - \hat{\beta}_1/\text{se}(\hat{\beta}_1)$ for $k = 1$. The general version involves a Wald statistic, for a two-sided test: $W = (\hat{\beta}_2 - \hat{\beta}_1)'\hat{V}^{-1}(\hat{\beta}_2 - \hat{\beta}_1)$ where $\hat{V}^{-1}(\hat{\beta}_2 - \hat{\beta}_1) = \left(\frac{\hat{V}\hat{\beta}_2 + \hat{V}\hat{\beta}_1}{2}\right)^{-1}$

   (c) Under the null $\frac{\sqrt{n}(\hat{\beta}_2 - \hat{\beta}_1)}{\text{se}(\hat{\beta}_2) + \text{se}(\hat{\beta}_1)} \xrightarrow{d} N(0, 1)$.
   In general $W \xrightarrow{d} \chi_k^2$ as $n \to \infty$, where k is the dimension of β_1 and β_2.

2.
   - H_0: $40\beta_1 + 40^2\beta_2 = 20$
   - H_1: $40\beta_1 + 40^2\beta_2 \neq 20$

   A Wald test with H_0 rewriten as $1 = 2\beta_1 + 80\beta_2$ utilizing a χ^2 distribution with $q = 1$ degree of freedom. The linear restriction is $R'\hat{\beta} = 1$, where $\hat{\beta}$ is the OLS estimate and $R' = (2, 80)$. The statistic is
$$W_n = \frac{(2\hat{\beta}_1 + 80\hat{\beta}_2 - 1)^2}{4\hat{\sigma}_1^2 + 320\hat{\sigma}_{12} + 6400\hat{\sigma}_2^2}.$$
   Critical value c is determined by setting test size $\alpha = 1 - G_{\chi_1^2}(c)$. If $W_n > c$ reject H_0 and fail to reject H_0 otherwise.

3. (a) $T = (\hat{\sigma}^2 - 1)(\frac{\hat{V}}{n})^{-\frac{1}{2}}$

   (b) $g(\sigma^2) = \sqrt{\sigma^2} = \sigma$ continuous and differentiable with $g'(\sigma^2) = \frac{1}{2\sqrt{\sigma^2}} = \frac{1}{2\sigma}$
   since $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, V)$ Delta method gives $\sqrt{n}(g(\hat{\sigma}^2) - g(\sigma^2)) \xrightarrow{d} \mathcal{N}(0, \frac{V}{4\sigma^2})$

   (c) $T = 2\hat{\sigma}(\hat{\sigma} - 1)(\frac{\hat{V}}{n})^{-\frac{1}{2}}$

   (d) the null hypothesis are the same as $\sigma = 1 \iff \sigma^2 = 1$ yet the tests are not since contradictory results can arise due to the additional factor $2\hat{\sigma}$ in c)

4. (a)
```
# Load the data and create variables
x <- as.matrix(cbind(matrix(1,nrow(lC),1),lQ,lPL,lPK,lPF))
y <- lC
n <- nrow(x)
k <- ncol(x)

# Unrestriced regression
invx <-solve(t(x)%*%x)
b_ols <- solve((t(x)%*%x),(t(x)%*%y))
e_ols <- rep((y-x%*%b_ols),times=k)
xe_ols <- x*e_ols
V_ols <- (n/(n-k))*invx%*%(t(xe_ols)%*%xe_ols)%*%invx
se_ols <- sqrt(diag(V_ols))
```

[1] 42624@student.hhs.se

```
ols_mat = cbind(b_ols,matrix(se_ols))
print(ols_mat)
[1,] -3.5265028 1.71860065
[2,]  0.7203941 0.03259753
[3,]  0.4363412 0.24563580
[4,] -0.2198884 0.32381213
[5,]  0.4265170 0.07548271
```

The restriction is ensuring constant returns to scale.

(b)
```
# Constrained regression
R <- c(0,0,1,1,1)
iR = invx%*%R%*%solve(t(R)%*%invx%*%R)%*%t(R)
b_cls <- b_ols -
invx%*%R%*%solve(t(R)%*%invx%*%R)%*%(t(R)%*%b_ols - 1)
e_cls <- rep((y-x%*%b_cls),times=k)
xe_cls <- x*e_cls
V_tilde <- (n/(n-k+1))*invx%*%(t(xe_cls)%*%xe_cls)%*%invx
V_cls <- V_tilde - iR%*%V_tilde - V_tilde%*%t(iR) +
  iR%*%V_tilde%*%t(iR)
se_cls <- sqrt(diag(V_cls))
cls_mat = cbind(b_cls,matrix(se_cls))
print(cls_mat)
[1,] -4.690789123 0.81485793
[2,]  0.720687524 0.03245926
[3,]  0.592909608 0.16906852
[4,] -0.007381064 0.15579133
[5,]  0.414471455 0.07286728
```

(c)
```
# Efficient minimum distance regression
b_emd <- b_ols -
V_ols%*%R%*%solve(t(R)%*%V_ols%*%R)%*%(t(R)%*%b_ols-1)
e_emd <- rep((y-x%*%b_emd),times=k)
xe_emd <- x*e_emd
V2 <- (n/(n-k+1))*invx%*%(t(xe_emd)%*%xe_emd)%*%invx
V_emd <- V2 - V2%*%R%*%solve(t(R)%*%V2%*%R)%*%t(R)%*%V2
se_emd <- sqrt(diag(V_emd))
emd_mat = cbind(b_emd,matrix(se_emd))
print(emd_mat)
[1,] -4.744646018 0.81541660
[2,]  0.720190849 0.03230573
[3,]  0.580519645 0.16946463
[4,]  0.009219041 0.15524763
[5,]  0.410261314 0.07244074
```

(d)
```
# Wald statistic
c <- qchisq(.95, df=1) # chi^2(1) critical value
W <- t(t(R)%*%b_ols - 1) %*% solve(t(R)%*%V_ols%*%R)
%*% (t(R)%*%b_ols - 1)
print(c(W,c))
[1] 0.6454737 3.8414588
```

We cannot reject H_0 since $W < c$.

(e)
```
# Minimum distance statistic
J <- t(b_ols-b_emd) %*% solve(V_ols) %*% (b_ols-b_emd)
print(c(J,c))
[1] 0.6454737 3.8414588
```

We cannot reject H_0 since $J < c$.

5. (a) Omit variables Hispanic, Black, Americian Indian, Asian, Mixed Race.

   (b) There are 4 restrictions: all coefficients related to marital should equal 0.

   (c)
```
#Load the data and create subsamples

dat <- read.table("cps09mar.txt")
edu12 <- (dat[,4]>11)
dat <- dat[edu12,]
black <- (dat[,11]==2)
dat <- dat[black,]
marriedF <- (dat[,12]<=3)&(dat[,2]==1)
marriedM <- (dat[,12]<=3)&(dat[,2]==0)
unionF <- (dat[,8]==1)&(dat[,2]==1)
unionM <- (dat[,8]==1)&(dat[,2]==0)
fmarriedF <- (dat[,12]<=6)&(dat[,12]>3)&(dat[,2]==1)
fmarriedM <- (dat[,12]<=6)&(dat[,12]>3)&(dat[,2]==0)
exp <- dat[,1]-dat[,4]-6
exp2 <- (exp^2)/100

# Unrestriced regression

y <- as.matrix(log(dat[,5]/(dat[,6]*dat[,7])))
x <- cbind(rep(1,nrow(y)),dat[,2],dat[,4],exp,exp2,
           unionF,unionM,marriedF,marriedM,fmarriedF,fmarriedM)
n <- nrow(x)
k <- ncol(x)
invx <-solve(t(x)%*%x)
b_ols <- solve((t(x)%*%x),(t(x)%*%y))
e_ols <- rep((y-x%*%b_ols),times=k)
xe_ols <- x*e_ols
V_ols <- (n/(n-k))*invx%*%(t(xe_ols)%*%xe_ols)%*%invx

# Wald test

R <- cbind(c(0,0,0,0,0,0,0,1,0,0,0),
           c(0,0,0,0,0,0,0,0,1,0,0),
           c(0,0,0,0,0,0,0,0,0,1,0),
           c(0,0,0,0,0,0,0,0,0,0,1))
W = t(t(R)%*%b_ols)%*%solve(t(R)%*%V_ols%*%R)%*%(t(R)%*%b_ols)
1-pchisq(W,df=4) # chi^2(4) pvalue
4.272095e-08
```

   (d) The P-Value is essentially 0, we can reject the null hypothesis.

6. The feasible moment estimator is $\hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{e}_i^2$. Obtain its bootstrap distribution by B estimations on independent samples $\{y_i^*, X_i^*\}$ created by i.i.d. sampling from the original dataset. For any $0 < \alpha < 1$, the empirical quantile $q^*\alpha$ is calculated such that $n\alpha$ bootstrap estimates are smaller than $q^*\alpha$. The percentile bootstrap $100(1-\alpha)\%$ confidence interval is then given by: $C_{pc} = \left(q^*_{\alpha/2}, q^*_{1-\alpha/2}\right)$ For instance, if $B = 1000$, $\alpha = 0.05$, and the empirical quantile estimator is used, then $C_{pc} = (\hat{\sigma}^{2*}_{25}, \hat{\sigma}^{2*}_{975})$.

7. The bootstrap algorithm generates B draws $T^*(b) = \frac{\hat{\beta}_2^* - \hat{\beta}_2}{s(\hat{\beta}_2)^*}$, where $b = 1, \ldots, B$, centered at the sample estimate $\hat{\beta}_2$ and calculated using the bootstrap standard error $s(\hat{\beta}_2{}^*)$. The bootstrap $100\alpha\%$ critical value is denoted as $q^*_{1-\alpha}$, where q^*_α represents the α-th quantile of the absolute values of the bootstrap t-ratios $|T^*(b)|$. For a $100\alpha\%$ test, we reject $H_0 : \beta_2 = 0$ in favor of $H_1 : \beta_2 \neq 0$ if $|T| > q^*_{1-\alpha}$ and fail to reject if $|T| \leq q^*_{1-\alpha}$.

8. (a)
```
// Load the data and create variables
use Nerlove1963.dta, clear
gen lC = log(cost)
gen lQ = log(output)
gen lPL = log(Plabor)
gen lPF = log(Pfuel)
gen lPK = log(Pcapital)

// Unrestricted regression
reg lC lQ lPL lPK lPF, r
est store asymp
reg lC lQ lPL lPK lPF, vce(jackknife)
est store jack
reg lC lQ lPL lPK lPF, vce(bootstrap, bca)
est store boot
esttab asymp jack boot, se
```

	coef	asymp	jack	boot
lQ	0.720	(0.0330)	(0.0339)	(0.0363)
lPL	0.436	(0.248)	(0.253)	(0.232)
lPK	−0.220	(0.328)	(0.336)	(0.406)
lPF	0.427	(0.0761)	(0.0778)	(0.0882)
_cons	−3.527	(1.740)	(1.788)	(2.067)

(b)
```
quietly: reg lC lQ lPL lPK lPF, vce(hc2)
nlcom _b[lPL] + _b[lPK] + _b[lPF]
jackknife (_b[lPL] + _b[lPK] + _b[lPF]): reg lC lQ lPL lPK lPF
bootstrap (_b[lPL] + _b[lPK] + _b[lPF]): reg lC lQ lPL lPK lPF
```

coef	asymp	jack	boot
0.643	(.4502086)	(.4626814)	(.4418348)

(c)
```
bootstrap (_b[lPL] + _b[lPK] + _b[lPF]): reg lC lQ lPL lPK lPF
bootstrap (_b[lPL] + _b[lPK] + _b[lPF] ), bca: reg lC lQ lPL lPK lPF
[95% conf. interval]            lower q      upper q
Bootstrap percentile          -.2237554     1.509695
Bootstrap BCa                 -.4276354     1.713575
```

9. (a)
```
* generate transformations
gen wage=ln(earnings/(hours*week))
gen experience = age - education - 6
gen exp2 = (experience^2)/100
* subset
keep if race == 1 & female == 0 & hisp == 1
& region == 2 & marital == 7
* estimate
reg wage education experience exp2, r
nlcom _b[education]/(_b[experience]+_b[exp2]/5)
jackknife (_b[education]/(_b[experience]+_b[exp2]/5)):
reg wage education experience exp2
bootstrap (_b[education]/(_b[experience]+_b[exp2]/5)):
reg wage education experience exp2
* report
Coefficient     asymp se     jack se     boot se
 2.899323       .7603923     .8229674    .8949533
```

(b) Small sample and nonlinearity of the parameter typically induces estimation bias. Discrepancy between the asymptotic and bootstrap standard error, and between bootstrap runs is a signal that there may be moment failure and consequently bootstrap standard errors are unreliable. Trimmed versions of bootstrap standard errors are preferred, especially for nonlinear functions of estimated coefficient.

(c)
```
bootstrap (_b[education]/(_b[experience]+_b[exp2]/5)), bca:
reg wage education experience exp2
[95% conf. interval]            lower q      upper q
Bootstrap BCa                  .8934213     4.905225
```