COMM8102 Econometric Analysis
Assignment 2

Yuhao Liu
z5097536

July 5, 2021

Excercise 4.1

(a)

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$$

(b)

$$\begin{aligned} E\left[\hat{\mu}_k\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^k\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[Y_i^k\right] \\ &= \frac{1}{n}\cdot n\cdot \mu_k \\ &= \mu_k \end{aligned}$$

(c)

$$\begin{aligned}
Var\left[\hat{\mu}_k\right] &= E\left[\hat{\mu}_k^2\right]-\left(E\left[\hat{\mu}_k\right]\right)^2\\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}Y_i^k\right)^2\right]-\mu_k^2\\
&= \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n}Y_i^k\right)^2\right]-\mu_k^2\\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n}Y_i^{2k}+\sum_{i=1,j=1,i\neq j}^{n}Y_i^kY_j^k\right]-\mu_k^2\\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}E\left[Y_i^{2k}\right]+\sum_{i=1,j=1,i\neq j}^{n}E\left[Y_i^k\right]E\left[Y_j^k\right]\right)-\mu_k^2\\
&= \frac{1}{n^2}\left(n\cdot\mu_{2k}+\left(n^2-n\right)\cdot\mu_k^2\right)-\mu_k^2\\
&= \frac{1}{n}\cdot\mu_{2k}+\mu_k^2-\frac{1}{n}\cdot\mu_k^2-\mu_k^2\\
&= \frac{1}{n}\left(\mu_{2k}-\mu_k^2\right)
\end{aligned}$$

We need to assume $\mu_{2k}=E\left[Y^{2k}\right]<\infty$.

(d)

$$\hat{V}_{\hat{\mu}_k}=\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i^{2k}-\left(\frac{1}{n}\sum_{i=1}^{n}Y_i^k\right)^2\right)$$

Excercise 4.5

(4.15)

$$\begin{aligned}
E\left[\hat{\beta}\middle|X\right] &= E\left[\left(X'X\right)^{-1}X'Y\middle|X\right]\\
&= \left(X'X\right)^{-1}X'E\left[Y|X\right]\\
&= \left(X'X\right)^{-1}X'E\left[X\beta+e|X\right]\\
&= \beta+\left(X'X\right)^{-1}X'E\left[e|X\right]\\
&= \beta
\end{aligned}$$

(4.16)

$$\begin{aligned}
Var\left[\hat{\beta}\middle|X\right] &= Var\left[\hat{\beta}-\beta\middle|X\right] \\
&= Var\left[(X'X)^{-1}X'Y-\beta\middle|X\right] \\
&= Var\left[(X'X)^{-1}X'(X\beta+e)-\beta\middle|X\right] \\
&= Var\left[(X'X)^{-1}X'e\middle|X\right] \\
&= (X'X)^{-1}X'Var\left[e|X\right]X(X'X)^{-1} \\
&= (X'X)^{-1}X'\Omega X(X'X)^{-1} \\
&= (X'X)^{-1}(X'\Omega X)(X'X)^{-1}
\end{aligned}$$

Excercise 4.12

$$\begin{aligned}
E\left[\left(\bar{Y}-\mu\right)^3\right] &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}Y_i-\mu\right)^3\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu)\right)^3\right] \\
&= \frac{1}{n^3}E\left[\left(\sum_{i=1}^{n}(Y_i-\mu)\right)^3\right] \\
&= \frac{1}{n^3}E\left[\begin{array}{c}\sum_{i=1}^{n}(Y_i-\mu)^3+3\cdot\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(Y_i-\mu)^2(Y_j-\mu) \\ +\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i,j}^{n}(Y_i-\mu)(Y_j-\mu)(Y_k-\mu)\end{array}\right] \\
&= \frac{1}{n^3}\cdot\left(\begin{array}{c}\sum_{i=1}^{n}E\left[(Y_i-\mu)^3\right]+3\cdot\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}E\left[(Y_i-\mu)^2\right]E\left[Y_j-\mu\right] \\ +\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i,j}^{n}E\left[Y_i-\mu\right]E\left[Y_j-\mu\right]E\left[Y_k-\mu\right]\end{array}\right) \\
&= \frac{1}{n^3}\cdot\sum_{i=1}^{n}E\left[(Y_i-\mu)^3\right] \\
&= \frac{\mu_3}{n^2}
\end{aligned}$$

Excercise 4.23

$$\begin{aligned}
E\left[\hat{\beta}\middle|X\right] &= E\left[(X'X+I_k\lambda)^{-1}X'Y\middle|X\right] \\
&= E\left[(X'X+I_k\lambda)^{-1}X'(X\beta+e)\middle|X\right] \\
&= E\left[(X'X+I_k\lambda)^{-1}X'X\beta+(X'X+I_k\lambda)^{-1}X'e\middle|X\right] \\
&= (X'X+I_k\lambda)^{-1}X'X\beta+(X'X+I_k\lambda)^{-1}X'E\left[e|X\right] \\
&= (X'X+I_k\lambda)^{-1}X'X\beta
\end{aligned}$$

$\hat{\beta}$ is biased for β, since $\forall\lambda>0, E\left[\hat{\beta}\middle|X\right]\neq\beta$.

Excercise 4.26

(a)

For the results of standard errors, see Table 1 below. In terms of absolute value, *Tracking* changes the most, and *Percentile* changes the least. In terms of proportion, *Tracking* still changes the most, but *Gender* changes the least.

Standard Errors	Tracking	Age	Gender	Assigned	Percentile	Intercept
Clustering based	0.07665	0.01339	0.02868	0.03771	0.00072	0.13319
Conventional robust	0.02417	0.00855	0.02424	0.02385	0.0004	0.08280

(b)

In comparison to the results from (4.55), the $\hat{\beta}$ on *Tracking* increases from 0.138 to 0.174 by inclusion of the individual controls, and the standard errors also decreases a bit in both clustering based on the school and conventional robust formula.

Q6

Paper: Miyazaki, T., 2019. Clarifying the response of gold return to financial indicators: An empirical comparative analysis using ordinary least squares, robust and quantile regressions. Journal of Risk and Financial Management, 12(1), p.33.
Main research questions: How gold return responds to the changes in various financial indicators, specifically stock return, stock return volatility, financial market stress, crude oil, and the value of the US dollar.
Main estimation results: See appendix.
How standard error calculated: From lecture slides, we know that for OLS, $Var\left[\hat{\beta}\middle| X\right] = \sigma^2 \left(X^\intercal X\right)^{-1}$. The unbiased estimator for σ^2 is $\frac{e^\intercal e}{n-k}$, with $e = Y - X\hat{\beta}$ and k is the number of regressors. The estimator of standard error is the sqaure root of the diagonal elements of

$$\frac{e^\intercal e}{n-k}\left(X^\intercal X\right)^{-1}$$

.

Table 5. OLS and robust regression results.

Dependent Variable: GOLD												
	A. Full Sample: 2/16/1990–4/27/2018				B. First Sample: 2/16/1990–1/26/1996				C. Second Sample: 2/02/1996–11/25/2005			
	Number of Observations: 1472				Number of Observations: 311				Number of Observations: 513			
	OLS		Robust regression		OLS		Robust regression		OLS		Robust regression	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
SPX	−0.099 **	0.036	−0.055 **	0.020	−0.228 **	0.045	−0.182 **	0.047	0.004	0.028	−0.011	0.026
SPVOL	−0.101	0.111	−0.099	0.063	−0.367 *	0.165	−0.672 **	0.146	−0.115	0.099	−0.183 *	0.093
FSI1	0.096	0.058	0.078	0.041	0.145	0.120	0.160	0.113	0.148	0.076	0.201 **	0.066
WTI	0.065 **	0.015	0.046 **	0.011	0.090 **	0.019	0.083 **	0.018	0.015	0.016	0.018	0.014
TWEX	−0.962 **	0.082	−0.892 **	0.048	−0.155	0.105	−0.151	0.079	−0.850 **	0.091	−0.870 **	0.070
Constant	0.312	0.231	0.302 *	0.136	0.744 *	0.324	1.240 **	0.265	0.344	0.227	0.460 *	0.226
Adj R^2		0.186		0.263		0.126		0.232		0.200		0.338
	Breusch–Pagan–Godfrey test				Breusch–Pagan–Godfrey test				Breusch–Pagan–Godfrey test			
	$\chi^2(10)$			0.000	$\chi^2(10)$			0.136	$\chi^2(10)$			0.001
	White test				White test				White test			
	$\chi^2(65)$			0.000	$\chi^2(65)$			0.000	$\chi^2(65)$			0.006
	D. Third sample: 12/02/2005–5/03/2013				E. Fourth sample: 5/10/2013–4/27/2018							
	Number of observations: 388				Number of observations: 260							
	OLS	Robust regression	OLS	Robust regression								
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.				
SPX	−0.267 **	0.078	−0.177 **	0.049	−0.126	0.069	−0.095	0.065				
SPVOL	−0.063	0.249	0.157	0.149	0.222	0.163	0.092	0.210				
FSI1	0.045	0.127	−0.109	0.092	−0.062	0.139	−0.049	0.164				
WTI	0.154 **	0.032	0.129 **	0.030	−0.052	0.027	−0.046	0.029				
TWEX	−1.573 **	0.182	−1.513 **	0.129	−1.131 **	0.130	−1.102 **	0.117				
Constant	0.437	0.513	0.060	0.322	−0.329	0.297	−0.116	0.361				
Adj R^2		0.316		0.401		0.299		0.381				
	Breusch–Pagan–Godfrey test				Breusch–Pagan–Godfrey test							
	$\chi^2(10)$			0.000	$\chi^2(10)$			0.007				
	White test				White test							
	$\chi^2(65)$			0.000	$\chi^2(65)$			0.181				

Notes: S.E. stands for standard error. For the OLS regression, the standard errors are adjusted by using the Newey–West (1987) method. Adj R^2 for robust regression shows adjusted $R^2{}_W$ proposed by Renaud and Victoria-Feser (2010). * and ** denote statistical significance at the 5% and 1% levels, respectively.

PS2 code

R Markdown

This is the code part of PS2

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.2
```

```
data <- read_excel("ddk2011test.xlsx")
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV2296 / R2296C48: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV2480 / R2480C48: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV2492 / R2492C48: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV3592 / R3592C48: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AV3901 / R3901C48: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA4904 / R4904C27: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5107 / R5107C27: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5116 / R5116C27: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5121 / R5121C27: got '.'

## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting numeric in AA5122 / R5122C27: got '.'
```

```
#data clean
data<-data.frame(data$schoolid,data$totalscore,data$tracking,data$etpteacher,data$girl,data$agetest,data
data<-data[data$data.girl!='.',]
data<-data[data$data.agetest!='.',]
data$data.girl<-sapply(data$data.girl,as.numeric)-2
data$data.agetest<-sapply(data$data.agetest,as.numeric)
data$data.percentile<-sapply(data$data.percentile,as.numeric)
#cluster based
y <- scale(as.matrix(data$data.totalscore))
n <- nrow(y)
x<-cbind(data$data.tracking,data$data.agetest,data$data.girl,data$data.etpteacher,data$data.percentile,r
schoolid <- as.matrix(data$data.schoolid)
k <- ncol(x)
```

```r
xx <- t(x)%*%x
invx <- solve(xx)
beta <- solve(xx,t(x)%*%y)
xe <- x*rep(y-x%*%beta,times=k)
# Clustered robust standard error
xe_sum <- rowsum(xe,schoolid)
G <- nrow(xe_sum)
omega <- t(xe_sum)%*%xe_sum
scalee <- G/(G-1)*(n-1)/(n-k)
V_clustered <- scalee*invx%*%omega%*%invx
se_clustered <- sqrt(diag(V_clustered))
print(beta,digits = 5)
```

```
##            [,1]
## [1,]  0.173580
## [2,] -0.041056
## [3,]  0.081706
## [4,]  0.180989
## [5,]  0.017424
## [6,] -0.576944
```

```r
print(se_clustered,digits = 5)
```

```
## [1] 0.07665351 0.01339398 0.02867527 0.03770844 0.00072473 0.13318931
```

```r
#conventional robust
e <- y-x%*%beta
leverage <- rowSums(x*(x%*%invx))
a <- n/(n-k)
sig2 <- (t(e) %*% e)/(n-k)
u1 <- x*(e%*%matrix(1,1,k))
u2 <- x*((e/sqrt(1-leverage))%*%matrix(1,1,k))
u3 <- x*((e/(1-leverage))%*%matrix(1,1,k))

v0 <- invx*as.numeric(sig2)
v1 <- invx %*% (t(u1)%*%u1) %*% invx
v1a <- a * invx %*% (t(u1)%*%u1) %*% invx
v2 <- invx %*% (t(u2)%*%u2) %*% invx
v3 <- invx %*% (t(u3)%*%u3) %*% invx
s0 <- sqrt(diag(v0)) # Homoskedastic formula
s1 <- sqrt(diag(v1)) # HC0
s1a <- sqrt(diag(v1a)) # HC1
s2 <- sqrt(diag(v2)) # HC2
s3 <- sqrt(diag(v3)) # HC3
print(s0)
```

```
## [1] 0.0241654219 0.0084575472 0.0241484133 0.0239139818 0.0004292976
## [6] 0.0834363938
```

```r
print(s1)
```

```
## [1] 0.0241571564 0.0085405393 0.0242239304 0.0238385277 0.0004269618
## [6] 0.0827288533
```

```r
print(s1a)
```

```
## [1] 0.0241709225 0.0085454062 0.0242377345 0.0238521122 0.0004272051
```

```
## [6] 0.0827759967
```

```r
print(s2)
```

```
## [1] 0.0241710799 0.0085487064 0.0242377785 0.0238521380 0.0004272418
## [6] 0.0827994159
```

```r
print(s3)
```

```
## [1] 0.0241850131 0.0085568932 0.0242516361 0.0238657582 0.0004275219
## [6] 0.0828701442
```