

Homework 3 Solutions

Hansen 3.23

I'll use the notation \mathbf{P}_X for the projection matrix of \mathbf{X} and \mathbf{P}_Z for the the projection matrix of \mathbf{Z} (and use similar notation for each annihilator matrix).

Notice that we can write

$$\mathbf{Z} = \mathbf{X}\Gamma$$

where

$$\Gamma = \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

which implies that

$$\mathbf{P}_X \mathbf{Z} = \mathbf{P}_X \mathbf{X}\Gamma = \mathbf{X}\Gamma = \mathbf{Z}$$

$$\mathbf{M}_X \mathbf{Z} = (\mathbf{I} - \mathbf{P}_X) \mathbf{Z} = \mathbf{0}$$

that is, if you project \mathbf{Z} on \mathbf{X} , you get \mathbf{Z} , and the residuals from this projection are equal to 0. This implies that $\mathbf{P}_X \mathbf{P}_Z = \mathbf{P}_Z$ which we use below (to see this, you can expand the definition of \mathbf{P}_Z and use the result in the previous display).

Similarly, notice that we can write $\mathbf{X} = \mathbf{Z}\Gamma^{-1}$ which, similar to above, implies that

$$\mathbf{P}_Z \mathbf{X} = \mathbf{P}_Z \mathbf{Z}\Gamma^{-1} = \mathbf{Z}\Gamma^{-1} = \mathbf{X}$$

$$\mathbf{M}_Z \mathbf{X} = (\mathbf{I} - \mathbf{P}_Z) \mathbf{X} = \mathbf{0}$$

so that if you project \mathbf{X} on \mathbf{Z} , then you get \mathbf{X} , and the residuals from this projection are equal to 0. Similar to above, this implies that $\mathbf{P}_Z \mathbf{P}_X = \mathbf{P}_X$ which we use below.

Next, notice that

$$\mathbf{M}_X \mathbf{M}_Z = (\mathbf{I} - \mathbf{P}_X)(\mathbf{I} - \mathbf{P}_Z) = \mathbf{I} - \mathbf{P}_X - \mathbf{P}_Z + \mathbf{P}_X \mathbf{P}_Z = \mathbf{I} - \mathbf{P}_X = \mathbf{M}_X \quad (1)$$

$$\mathbf{M}_Z \mathbf{M}_X = (\mathbf{I} - \mathbf{P}_Z)(\mathbf{I} - \mathbf{P}_X) = \mathbf{I} - \mathbf{P}_Z - \mathbf{P}_X + \mathbf{P}_Z \mathbf{P}_X = \mathbf{I} - \mathbf{P}_Z = \mathbf{M}_Z \quad (2)$$

Finally, notice that

$$\begin{aligned} n\hat{\sigma}^2 &= \mathbf{Y}'\mathbf{M}_X\mathbf{Y} \\ &= (\mathbf{M}_X\mathbf{Y})'\mathbf{M}_X\mathbf{Y} \\ &= (\mathbf{M}_X\mathbf{M}_Z\mathbf{Y})'\mathbf{M}_X\mathbf{M}_Z\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{M}_Z\mathbf{M}_X\mathbf{M}_X\mathbf{M}_Z\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{M}_Z\mathbf{M}_X\mathbf{M}_Z\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{M}_Z\mathbf{M}_Z\mathbf{Y} \\ &= n\tilde{\sigma}^2 \end{aligned}$$

where the first equality holds from the definition of $\hat{\sigma}^2$; the second equality holds because \mathbf{M}_X is idempotent and symmetric, the third equality uses Equation 1; the fourth equality applies the transpose to the term on the left and uses that \mathbf{M}_Z and \mathbf{M}_X are symmetric; the fifth equality holds because \mathbf{M}_X is idempotent; the sixth equality holds by Equation 2; and the last equality holds by the definition of $\tilde{\sigma}^2$. This implies that $\hat{\sigma}^2 = \tilde{\sigma}^2$.

Hansen 3.24

Part a

```
# read data
library(haven)
cps <- read_dta("cps09mar.dta")

# construct subset of single, Asian men
data <- subset(cps, marital==7 & race==4 & female==0)

# ...not totally clear if this is exactly right subset
# confirm same number of rows as mentioned in textbook
nrow(data)
```

```
## [1] 268
```

```
# construct experience and wage
data$exp <- data$age - data$education - 6
data$wage <- data$earnings/(data$hours*data$week)

# also construct subset with < 45 years of experience
data <- subset(data, exp < 45)
```

```
# run regression
Y <- log(data$wage)
X <- cbind(1, data$education, data$exp, data$exp^2/100)
bet <- solve(t(X)%*%X)%*%t(X)%*%Y
round(bet,3)
```

```
##           [,1]
## [1,]  0.531
## [2,]  0.144
## [3,]  0.043
## [4,] -0.095
```

Part b

```
# residual regression
X1 <- data$education
X2 <- cbind(1, data$exp, data$exp^2/100)
ycoef <- solve(t(X2)%*%X2)%*%t(X2)%*%Y
yresid <- Y - X2%*%ycoef
x1coef <- solve(t(X2)%*%X2)%*%t(X2)%*%X1
x1resid <- X1 - X2%*%x1coef
fw_bet <- solve(t(x1resid)%*%x1resid)%*%t(x1resid)%*%yresid
round(fw_bet,3)
```

```
##           [,1]
## [1,]  0.144
```

This is the same as the estimate from part a. This is expected due to the Frisch-Waugh theorem.

Hansen 3.25

```
# a)
ehat <- Y - X%*%bet
round(sum(ehat),5)
```

```
## [1] 0
```

```
# b)
round(sum(data$education*ehat),5)
```

```
## [1] 0
```

```
# c)
round(sum(data$exp*ehat),5)
```

```
## [1] 0
```

```
# d)
round(sum(data$education^2 * ehat),5)
```

```
## [1] 133.1331
```

```
# e)
round(sum(data$exp^2 * ehat),5)
```

```
## [1] 0
```

```
# f)
Yhat <- X%*%bet
round(sum(Yhat*ehat),5)
```

```
## [1] 0
```

```
# g)
round(sum(ehat^2),5)
```

```
## [1] 82.505
```

Yes, these calculations are consistent with the theoretical properties of OLS. Parts a, b, c, e, and f all hold due to the property that $\sum_{i=1}^n X_i \hat{e}_i = 0$. Part d is not equal to 0 because X_1^2 is not an included regressor. Part g provides the sum of squared errors which is not generally equal to 0.

Hansen 4.1

Part (a):

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

Part (b):

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

where the third equality holds because the Y_i are identically distributed (implying the mean is the same across i). This result implies that $\hat{\mu}_k$ is unbiased for μ_k .

Part (c):

$$\begin{aligned} \text{var}(\hat{\mu}_k) &= \text{var}\left(\frac{1}{n} \sum_{i=1}^n Y_i^k\right) \\ &= \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n Y_i^k\right) \\ &= \frac{1}{n} \sum_{i=1}^n \text{var}(Y^k) \\ &= \frac{\text{var}(Y^k)}{n} \end{aligned}$$

For $\text{var}(\hat{\mu}_k)$ to exist, we need for $\text{var}(Y^k)$ to exist. Notice that,

$$\text{var}(Y^k) = E[(Y^k)^2] - E[Y^k]^2$$

Thus, the condition that we need is that $E[(Y^k)^2] = E[Y^{2k}] < \infty$.

Part (d): We can estimate by

$$\widehat{\text{var}}(\hat{\mu}_k) = \frac{\widehat{\text{var}}(Y^k)}{n} = \frac{\frac{1}{n} \sum_{i=1}^n Y_i^{2k} - \left(\frac{1}{n} \sum_{i=1}^n Y_i^k\right)^2}{n}$$

Hansen 4.5

First (and notice that this is exactly the same as what we showed in class... because unbiasedness did not rely on homoskedasticity),

$$\begin{aligned}
E[\hat{\beta}|\mathbf{X}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}|\mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}|\mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\
&= \beta
\end{aligned}$$

Next,

$$\begin{aligned}
\text{var}(\hat{\beta}|\mathbf{X}) &= \text{var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}|\mathbf{X}) \\
&= \text{var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{e})|\mathbf{X}) \\
&= \text{var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}|\mathbf{X}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}(\mathbf{e}|\mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

where the first equality holds by the definition of $\hat{\beta}$, the second equality substitutes for \mathbf{Y} , the third equality holds because the first term is equal to β and therefore doesn't contribute to the variance, the fourth equality holds because the variance is conditional on \mathbf{X} (so the terms involving \mathbf{X} can come out but need to be “squared”), and the last equality holds by the definition of $\text{var}(\mathbf{e}|\mathbf{X})$ given in the problem. This is the result we were trying to show.

Hansen 4.6

Recall that the restriction to linear estimators implies that we can write any estimator in this class as $\tilde{\beta} = \mathbf{A}'\mathbf{Y}$ for an $n \times k$ matrix \mathbf{A} that is a function of \mathbf{X} . Unbiasedness implies that, it must be the case that $E[\tilde{\beta}|\mathbf{X}] = \beta$. Then, notice that under linearity, we have that

$$E[\tilde{\beta}|\mathbf{X}] = E[\mathbf{A}'\mathbf{Y}|\mathbf{X}] = \mathbf{A}'E[\mathbf{Y}|\mathbf{X}] = \mathbf{A}'\mathbf{X}\beta$$

where the second equality holds because \mathbf{A} is a function of \mathbf{X} . Therefore, together linearity and unbiasedness imply that $\mathbf{A}'\mathbf{X} = \mathbf{I}_k$. Next, notice that

$$\text{var}(\tilde{\beta}|\mathbf{X}) = \text{var}(\mathbf{A}'\mathbf{Y}|\mathbf{X}) = \mathbf{A}'\text{var}(\mathbf{Y}|\mathbf{X})\mathbf{A} = \mathbf{A}'\text{var}(\mathbf{X}\beta + \mathbf{e}|\mathbf{X})\mathbf{A} = \mathbf{A}'\text{var}(\mathbf{e}|\mathbf{X})\mathbf{A} = \sigma^2\mathbf{A}'\Sigma\mathbf{A}$$

We aim to show that $\text{var}(\tilde{\beta}|\mathbf{X}) - \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \geq \mathbf{0}$. Notice that

$$\begin{aligned}
\text{var}(\tilde{\beta}|\mathbf{X}) - \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} &= \sigma^2 \left(\mathbf{A}'\Sigma\mathbf{A} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \right) \\
&= \sigma^2 \left(\mathbf{A}'\Sigma\mathbf{A} - \mathbf{A}'\Sigma^{1/2}\Sigma^{-1/2}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1/2}\Sigma^{1/2}\mathbf{A} \right) \\
&= \sigma^2 \mathbf{A}'\Sigma^{1/2} \underbrace{\left(\mathbf{I} - \Sigma^{-1/2}\mathbf{X}((\Sigma^{-1/2}\mathbf{X})'\Sigma^{-1/2}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1/2} \right)}_{=\mathbf{M}_{\Sigma^{-1/2}\mathbf{X}}} \Sigma^{1/2}\mathbf{A} \\
&= \sigma^2 \mathbf{A}'\Sigma^{1/2}\mathbf{M}_{\Sigma^{-1/2}\mathbf{X}}\Sigma^{1/2}\mathbf{A} \\
&= \sigma^2 \left(\mathbf{M}_{\Sigma^{-1/2}\mathbf{X}}\Sigma^{1/2}\mathbf{A} \right)' \mathbf{M}_{\Sigma^{-1/2}\mathbf{X}}\Sigma^{1/2}\mathbf{A} \\
&\geq 0
\end{aligned}$$

where the above result repeatedly uses Σ is positive definite and symmetric (which implies that it has a positive definite and symmetric inverse, and that it has a positive definite and symmetric square root matrix, and so does its inverse). In particular, the second equality holds because (i) $\Sigma^{-1/2}\Sigma^{1/2} = \mathbf{I}_n$, and $\mathbf{A}'\mathbf{X} = \mathbf{I}_k$ (due to linearity and unbiasedness as discussed above); the third equality holds by factoring out $\mathbf{A}'\Sigma^{1/2}$ and from a slight manipulation of the inside term; the fourth equality holds by the definition of $\mathbf{M}_{\Sigma^{-1/2}\mathbf{X}}$ (which is annihilator matrix); the fifth equality holds because $\mathbf{M}_{\Sigma^{-1/2}\mathbf{X}}$ is idempotent and symmetric; and the last equality holds because the previous expression is a quadratic form.

Hansen 4.23

Notice that

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_{ridge}|\mathbf{X}] &= \mathbb{E} \left[(\mathbf{X}'\mathbf{X} + \mathbf{I}_k\lambda)^{-1}\mathbf{X}'\mathbf{Y} \right] \\
&= (\mathbf{X}'\mathbf{X} + \mathbf{I}_k\lambda)^{-1}\mathbf{X}'\mathbb{E}[\mathbf{Y}|\mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X} + \mathbf{I}_k\lambda)^{-1}\mathbf{X}'\mathbf{X}\beta \\
&\neq \beta
\end{aligned}$$

This implies that $\hat{\beta}_{ridge}$ is not unbiased for β .

Extra Question

```

# function to run a single simulation
sim <- function() {
  # draw X1
  X1 <- rexp(n)

  # draw the error term

```

```

e <- mixtools::rnormmix(n, lambda=c(.5,.5), mu=c(-2,2), sigma=c(1,1))

## TODO: construct Y
Y <- b0 + b1*X1 + e

## TODO: use X1 and Y to estimate bet0 and bet1
X <- cbind(1,X1)
bet <- solve(t(X)%*%X)%*%t(X)%*%Y
bet1 <- bet[2,1]

# TODO: return estimated value of bet1
bet1
}

# function to run many simulations
# @param n_sims is the number of simulations to run
run_mc <- function(n_sims=1000) {

  # run n_sims simulations and store in a vector
  mc_res <- sapply(1:n_sims, function(s) {
    sim()
  })

  # print number of observations
  cat("n = ", n, "....\n")

  # print the mean of b1
  cat("mean b1 : ", mean(mc_res), "\n")

  # print the variance of b1
  cat("var b1 : ", var(mc_res), "\n")
}

# run the simulations
# set values of parameters and number of observations
set.seed(1234) # so can reproduce
b0 <- 0
b1 <- 1

```



```
n <- 2
run_mc()
```

```
## n = 2 ....
## mean b1 : 2.506428
## var b1 : 4841.861
```

```
n <- 10
run_mc()
```

```
## n = 10 ....
## mean b1 : 1.036217
## var b1 : 1.074335
```

```
n <- 50
run_mc()
```

```
## n = 50 ....
## mean b1 : 0.9884143
## var b1 : 0.122526
```

```
n <- 100
run_mc()
```

```
## n = 100 ....
## mean b1 : 1.004076
## var b1 : 0.05406661
```

```
n <- 500
run_mc()
```

```
## n = 500 ....
## mean b1 : 1.006865
## var b1 : 0.01053563
```

It looks like our theory is holding here. $\hat{\beta}_1$ appears to be unbiased — recall unbiasedness is a finite sample property — so this should hold for all values of n (the only case where there are issues is when $n = 2$; in this case, you can see that the variance is extremely high, and I think that we are not doing enough simulations to see that it is actually unbiased in this case). The other interesting thing to note is that, as expected, the variance of $\hat{\beta}_1$ is decreasing for larger sample sizes.