# Homework 2 Solutions

## 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 $\mathbb{E}[\tilde{\beta}|\mathbf{X}] = \beta$. Then, notice that under linearity, we have that

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

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

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

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

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

where the above result repeatedly uses $\mathbf{\Sigma}$ 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) $\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{I}_n$, and $\mathbf{A}'\mathbf{X} = \mathbf{I}_k$ (due to linearity and unbiasedness as discussed above); the third equality holds by factoring out $\mathbf{A}'\mathbf{\Sigma}^{1/2}$ and from a slight manipulation of the inside term; the fourth equality holds by the definition of $\mathbf{M}_{\Sigma^{-1/2}X}$ (which is an annihilator matrix); the fifth equality holds because $\mathbf{M}_{\Sigma^{-1/2}X}$ is idempotent and symmetric; and the last equality holds because the previous expression is a quadratic form.

## Hansen 7.7

### (a)

$\beta$ is defined as the coefficient of the linear projection of $Y^*$ on $X$. Thus,

$$\beta = \mathbb{E}[XX']^{-1}\mathbb{E}[XY^*] = \mathbb{E}[XX']^{-1}\mathbb{E}[X(Y-u)] = \mathbb{E}[XX']^{-1}\mathbb{E}[XY]$$

where the second equality holds by plugging in $Y = Y^* + u$, and the third equality holds because $\mathbb{E}[Xu] = 0$. Thus, $\beta$ is the coefficient from the linear projection of $Y$ on $X$.

I think the above is the correct answer to the question, but there is one more thing that is worth pointing out. As in the problem, let's define $\hat{\beta}$ as the estimate that comes from running a regression of $Y$ on $X$, and additionally define $\hat{\beta}^*$ as the (infeasible) regression coefficient that you would get if you could run the regression of $Y^*$ on $X$. Note that

$$\hat{\beta}^* = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i^*$$

and

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i (Y_i^* + u_i) \qquad (1)\end{aligned}$$

so, in general, $\hat{\beta} \neq \hat{\beta}^*$; that is, if we were to observe $Y_i^*$, we would not get numerically estimates from the regression of $Y$ on $X$ as from the regression of $Y^*$ on $X$.

### (b)

From Equation 1, we can write

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i^* + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i u_i \\ &\xrightarrow{p} \mathbb{E}[XX']^{-1}\mathbb{E}[XY^*] + 0 \\ &= \beta\end{aligned}$$

where the second equality holds by the law of large numbers and the continuous mapping theorem. This implies that, despite the measurement error, $\hat{\beta}$ is consistent for $\beta$.

### (c)

Plugging in $Y_i^* = X_i'\beta + e_i$ into Equation 1 and multiplying by $\sqrt{n}$, we have that

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i (e_i + u_i) \\ &= \mathbb{E}[XX']^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i (e_i + u_i) + o_p(1) \\ &\xrightarrow{d} \mathcal{N}(0, \mathbb{E}[XX']^{-1}\boldsymbol{\Omega}\mathbb{E}[XX']^{-1})\end{aligned}$$

where

$$\boldsymbol{\Omega} = \mathbb{E}[XX'(e+u)^2]$$

This is related, but different, from the case without measurement error; recall that, in that case $\boldsymbol{\Omega} = \mathbb{E}[XX'e^2]$.

Altogether, this suggests that, when there is this relatively simple kind of measurement error in the outcome, using the measured-with-error outcome still delivers consistent estimates of $\beta$, but the asymptotic variance changes; it is likely to be bigger.

### Hansen 7.28 (part a only)

### (a)

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

# construct subset of white, male, Hispanic
data <- subset(cps, race == 1 & female == 0 & hisp == 1)

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

# run regression
Y <- log(data$wage)
X <- cbind(data$education, data$exp, data$exp^2 / 100, 1)
bet <- solve(t(X) %*% X) %*% t(X) %*% Y
round(bet, 5)
```

```
         [,1]
[1,]  0.09045
[2,]  0.03538
[3,] -0.04651
[4,]  1.18521
```

```
# construct standard errors
ehat <- as.numeric(Y - X %*% bet)
Xe <- X * ehat
n <- nrow(data)
Omeg <- t(Xe) %*% Xe / n
XX <- t(X) %*% X / n
V <- solve(XX) %*% Omeg %*% solve(XX)
se <- sqrt(diag(V)) / sqrt(n)
round(data.frame(beta = bet, se = se), 5)
```

```
      beta      se
1  0.09045 0.00292
2  0.03538 0.00258
3 -0.04651 0.00530
4  1.18521 0.04608
```

In case you are interested, I have provided some code to compare the above results to the standard errors that you get with several different regression implementations.

```r
# compare to R's lm function
reg <- lm(log(wage) ~ education + exp + I(exp^2 / 100), data = data)
summary(reg)
```

```
Call:
lm(formula = log(wage) ~ education + exp + I(exp^2/100), data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.0275 -0.3135  0.0063  0.3411  2.8603 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.185209   0.044745  26.488   <2e-16 ***
education     0.090449   0.002737  33.051   <2e-16 ***
exp           0.035380   0.002512  14.083   <2e-16 ***
I(exp^2/100) -0.046506   0.005027  -9.251   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5739 on 4226 degrees of freedom
Multiple R-squared:  0.2334,	Adjusted R-squared:  0.2328 
F-statistic: 428.8 on 3 and 4226 DF,  p-value: < 2.2e-16
```

```r
# Notice that estimates of beta are the same but
# standard errors are different

library(estimatr)
reg2 <- lm_robust(log(wage) ~ education + exp + I(exp^2 / 100), data = data, se_type = "HC0")
summary(reg2)
```

```
Call:
lm_robust(formula = log(wage) ~ education + exp + I(exp^2/100), 
    data = data, se_type = "HC0")

Standard error type:  HC0 

Coefficients:
             Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper   DF
(Intercept)   1.18521   0.046078  25.722 1.129e-135  1.09487  1.27555 4226
education     0.09045   0.002915  31.028 1.312e-190  0.08473  0.09616 4226
exp           0.03538   0.002584  13.691  8.859e-42  0.03031  0.04045 4226
I(exp^2/100) -0.04651   0.005304  -8.767  2.606e-18 -0.05691 -0.03611 4226

Multiple R-squared:  0.2334 ,	Adjusted R-squared:  0.2328 
F-statistic: 372.7 on 3 and 4226 DF,  p-value: < 2.2e-16
```

```
# these are exactly the same now

# Homoskedasticity standard errors
sigma2 <- mean(ehat^2)
V0 <- sigma2 * solve(XX)
se0 <- sqrt(diag(V0)) / sqrt(n)
se0
```

```
[1] 0.002735371 0.002511055 0.005024746 0.044723948
```

```
# these are very, very close to R's lm standard errors
# but not exactly the same

# Homoskedasticity w/ degree of freedom adjustment
k <- 4 # number of regressors (including intercept)
s2 <- sum(ehat^2) / (n - k)
Vs <- s2 * solve(XX)
ses <- sqrt(diag(Vs)) / sqrt(n)
ses
```

```
[1] 0.002736665 0.002512243 0.005027124 0.044745109
```

```
# these are exactly the same now
```

## Hansen 9.29

### (a)

We will run the following regresion:

$$\begin{aligned}\log(\text{wage}) = &\beta_0 + \beta_1\text{black} + \beta_2\text{female} + \beta_3\text{education} + \beta_4(\text{black} \times \text{female}) + \beta_5(\text{black} \times \text{education}) \\ &+ \beta_6(\text{female} \times \text{education}) + \beta_7(\text{black} \times \text{female} \times \text{education}) + e\end{aligned}$$

$\beta_3$ captures the return to education for white males, $\beta_3 + \beta_5$ captures the return to education for black males, $\beta_3 + \beta_6$ captures the return to education for white females, and $\beta_3 + \beta_5 + \beta_6 + \beta_7$ captures the return to education for black females.

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

# construct subset of white or black (drop Hispanic)
data <- subset(cps, race %in% c(1, 2))

# construct wage
data$wage <- data$earnings / (data$hours * data$week)

# construct demographic factors
```

```
data$black <- as.factor(ifelse(data$race == 2, 1, 0))
data$female <- as.factor(data$female)

# run regression
Y <- log(data$wage)

X <- model.matrix(~ black * female * education, data = data)
bet <- solve(t(X) %*% X) %*% t(X) %*% Y

# estimate asymptotic variance matrix and standard errors
ehat <- as.numeric(Y - X %*% bet)
Xe <- X * ehat
n <- nrow(data)
Omeg <- t(Xe) %*% Xe / n
XX <- t(X) %*% X / n
V <- solve(XX) %*% Omeg %*% solve(XX)
se <- sqrt(diag(V)) / sqrt(n)
round(data.frame(beta = bet, se = se), 5)
```

```
                            beta      se
(Intercept)              1.52448 0.02117
black1                  -0.15689 0.08208
female1                 -0.21352 0.03206
education                0.11171 0.00153
black1:female1          -0.07257 0.10972
black1:education        -0.00065 0.00596
female1:education       -0.00458 0.00229
black1:female1:education 0.01313 0.00789
```

### (b)

The return to education will be the same for all demographic groups if $\beta_5 = \beta_6 = \beta_7 = 0$, i.e., $H_0 : \mathbf{R}'\beta = 0$ where (don't forget the intercept is included in beta)

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

```
# construct R matrix
R <- matrix(0, nrow = length(bet), ncol = 3)
R[6, 1] <- 1
R[7, 2] <- 1
R[8, 3] <- 1
```

```
# calculate Wald statistic
Lambda <- t(R) %*% V %*% R
W <- n * t(t(R) %*% bet) %*% solve(Lambda) %*% (t(R) %*% bet)
round(W, 5)
```

```
        [,1]
[1,] 7.94746
```

```
# calculate p-value
1 - pchisq(W, df = 3)
```

```
           [,1]
[1,] 0.04711018
```

### (c)

The previous results indicate that we (barely) reject at the 5% level the null hypothesis that the return to education is the same for all demographic groups.

## Extra Question 1

### (a)

```
fibonacci <- function(n) {
    # handle n = 1 and n = 2 separately
    if (n == 1) {
        return(0)
    } else if (n == 2) {
        return(1)
    }

    # initialize the sequence
    fib_seq <- c(0, 1)

    # loop to compute the sequence
    for (i in 3:n) {
        fib_seq[i] <- fib_seq[i - 1] + fib_seq[i - 2]
    }

    # return the nth element
    return(fib_seq[n])
}

fibonacci(5)
```

```
[1] 3
```

```r
fibonacci(8)
```

```
[1] 13
```

### (b)

```r
alt_seq <- function(a, b, n) {
    # handle n = 1 and n = 2 separately
    if (n == 1) {
        return(a)
    } else if (n == 2) {
        return(b)
    }

    # initialize the sequence
    alt_seq <- c(a, b)

    # loop to compute the sequence
    for (i in 3:n) {
        alt_seq[i] <- alt_seq[i - 1] + alt_seq[i - 2]
    }

    # return the nth element
    return(alt_seq[n])
}
alt_seq(3, 7, 4)
```

```
[1] 17
```

## Extra Question 2

Recall that FWL says that

$$\beta_1 = \mathbb{E}[vv']^{-1}\mathbb{E}[vu]$$

### (a)

When $k_1 = 1$, FWL immediately simplifies to

$$\begin{aligned}
\beta_1 &= \frac{\mathbb{E}[vu]}{\mathbb{E}[v^2]} \\
&= \frac{\mathbb{E}[v(Y - X_2'\gamma_2)]}{\mathbb{E}[v^2]} \\
&= \frac{\mathbb{E}[vY]}{\mathbb{E}[v^2]} - \frac{\gamma_2'\mathbb{E}[vX_2]}{\mathbb{E}[v^2]} \\
&= \frac{\mathbb{E}[vY]}{\mathbb{E}[v^2]}
\end{aligned}$$

where the last line holds because $\mathbb{E}[vX_2] = 0$. This is what we wanted to show.

### (b)

When $X_2 = 1$, we have that

$$Y = \delta_2 + u \implies \delta_2 = \mathbb{E}[Y] \implies u = Y - \mathbb{E}[Y]$$

and

$$X_1 = \lambda_{12} + v \implies \lambda_{12} = \mathbb{E}[X_1] \implies v = X_1 - \mathbb{E}[X_1]$$

Then, from FWL, we have that

$$\begin{aligned}
\beta_1 &= \frac{\mathbb{E}[vu]}{\mathbb{E}[v^2]} \\
&= \frac{\mathbb{E}[(X_1 - \mathbb{E}[X_1])(Y - \mathbb{E}[Y])]}{\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2]} \\
&= \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)}
\end{aligned}$$

where the second equality holds by plugging in for $u$ and $v$, and the third equality holds by the definitions of covariance and variance. This is what we wanted to show.