## Hansen 3.16

To start with, notice that

$$R_1^2 = 1 - \frac{\tilde{\mathbf{e}}'\tilde{\mathbf{e}}}{(\mathbf{Y} - \mathbf{1}_n\bar{Y})'(\mathbf{Y} - \mathbf{1}_n\bar{Y})}$$
$$R_2^2 = 1 - \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{(\mathbf{Y} - \mathbf{1}_n\bar{Y})'(\mathbf{Y} - \mathbf{1}_n\bar{Y})}$$

where these expressions come from Section 3.14 in Hansen (these are just saying that $R^2$ is 1 minus the ratio of the sum of squared residuals to the total sum of squares). We aim to show that $R_2^2 \geq R_1^2$. To do this, given that the only difference between the two expressions comes $\tilde{\mathbf{e}}'\tilde{\mathbf{e}}$ versus $\hat{\mathbf{e}}'\hat{\mathbf{e}}$, the result will hold if we can show that $\tilde{\mathbf{e}}'\tilde{\mathbf{e}} \geq \hat{\mathbf{e}}'\hat{\mathbf{e}}$.

One useful property of annihilator matrices that is useful below is that

$$\begin{aligned} \mathbf{M}\mathbf{M_1} &= \mathbf{M}(\mathbf{I} - \mathbf{P}_1) \\ &= \mathbf{M} \end{aligned} \tag{1}$$

where $\mathbf{P}_1$ and $\mathbf{M}_1$ are the projection and annihilator matrices for $\mathbf{X}_1$. The second equality holds because $\mathbf{M}\mathbf{P}_1 = \underbrace{\mathbf{M}\mathbf{X}_1}_{=\mathbf{0}}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' = \mathbf{0}$. An implication of Equation 1 is that $\mathbf{M}_1\mathbf{M} = \mathbf{M}$, which follows from the symmetry properties of annihilator matrices which have used many times before. This is useful below.

Next, notice that

$$\begin{aligned} \tilde{\mathbf{e}}'\tilde{\mathbf{e}} &= (\mathbf{Y} - \mathbf{X}_1\tilde{\beta}_1)'(\mathbf{Y} - \mathbf{X}_1\tilde{\beta}_1) \\ &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}_1\tilde{\beta}_1 + \tilde{\beta}_1'\mathbf{X}_1'\mathbf{X}_1\tilde{\beta}_1 \\ &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{P}_1\mathbf{Y} + \mathbf{Y}'\mathbf{P}_1'\mathbf{P}_1\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{P}_1\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{M}_1\mathbf{Y} \end{aligned}$$

where the first equality holds be the definition of $\tilde{\mathbf{e}}$, the second equality expands the previous line, the third equality holds because $\mathbf{X}_1\tilde{\beta}_1 = \mathbf{P}_1\mathbf{Y}$, the fourth equality holds because $\mathbf{P}_1$ is symmetric and idempotent and by cancelling terms, and the last equality holds because $\mathbf{M}_1 = \mathbf{I} - \mathbf{P}_1$.

Second, notice that

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_2\hat{\beta}_2 + \hat{\mathbf{e}} \\ \implies \mathbf{M}_1\mathbf{Y} &= \underbrace{\mathbf{M}_1\mathbf{X}_1}_{=0}\hat{\beta}_1 + \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2 + \mathbf{M}_1\hat{\mathbf{e}} \\ &= \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2 + \mathbf{M}_1\mathbf{M}\mathbf{e} \\ &= \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2 + \mathbf{M}\mathbf{e} \\ &= \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2 + \hat{\mathbf{e}} \\ \implies \hat{\mathbf{e}} &= \mathbf{M}_1\mathbf{Y} - \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2 \end{aligned}$$

where the first line is given in the problem, the second line comes from pre-multiplying by $\mathbf{M}_1$, the third equality cancels the first term and holds because $\hat{\mathbf{e}} = \mathbf{M}\mathbf{e}$, the fourth equality holds because (as discussed above) $\mathbf{M}_1\mathbf{M} = \mathbf{M}$, the fifth equality again uses that $\mathbf{M}\mathbf{e} = \hat{\mathbf{e}}$, and the last line holds by rearranging terms. From this expression, we have that

$$\begin{aligned}
\hat{\mathbf{e}}'\hat{\mathbf{e}} &= (\mathbf{M}_1\mathbf{Y} - \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2)'(\mathbf{M}_1\mathbf{Y} - \mathbf{M}_1\mathbf{X}_2\hat{\beta}_2) \\
&= \mathbf{Y}'\mathbf{M}_1\mathbf{Y} - 2\hat{\beta}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} + \hat{\beta}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\hat{\beta}_2 \\
&= \mathbf{Y}'\mathbf{M}_1\mathbf{Y} - 2\hat{\beta}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} + \hat{\beta}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{M}_1\mathbf{Y} - \hat{\beta}_2'\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{M}_1\mathbf{Y} - \underbrace{\mathbf{Y}'\mathbf{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y}}_{\geq 0}
\end{aligned}$$

where the first equality holds from the previous expression for $\hat{\mathbf{e}}$, the second equality expands the previous line, the third equality holds by plugging in $\hat{\beta}_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y}$ (which holds by FWL), the fourth equality holds by cancelling and combining terms from the previous line, and the last equality holds by plugging in for $\hat{\beta}_2$ again. The underlined term is non-negative because it is a quadratic form.

Plugging in from the above expressions, we have that

$$\tilde{\mathbf{e}}'\tilde{\mathbf{e}} - \hat{\mathbf{e}}'\hat{\mathbf{e}} = \mathbf{Y}'\mathbf{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} \geq 0$$

which, as discussed above, implies that $R_2^2 \geq R_1^2$.

The case where $R_2^2 = R_1^2$ occurs when $\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} = 0$. This is equivalent to $\hat{\beta}_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y} = 0$; i.e., $R^2$ is the same for the two models if $\hat{\beta}_2 = 0$. This is the case where the second set of regressors does not help to explain the variation in $Y$ after accounting for the first set of regressors.

## Hansen 3.22

From the first regression, we immediately have that

$$\tilde{\mathbf{u}} = \mathbf{M}_1\mathbf{Y}$$

which holds because it is a regression of $Y$ on $X_1$ (and I use bold font above to indicate that, e.g., $\tilde{\mathbf{u}}$ is the $n \times 1$ vector of residuals from the first regression). Then, the coefficient from the second regression is given by

$$\begin{aligned}
\tilde{\beta}_2 &= (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\tilde{\mathbf{u}} \\
&= (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y}
\end{aligned}$$

We can compare this to $\hat{\beta}$ from the third regression given in the problem. We immediately know from FWL-type arguments that

$$\hat{\beta}_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{Y}$$

In general, these are not equal to each other.

## Hansen 3.24

### Part a

```r
# 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
```

```r
# 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
```

```r
ehat <- Y - X%*%bet

# sum of squared errors
ssr <- t(ehat)%*%ehat
round(ssr,3)
```

```
       [,1]
[1,] 82.505
```

```
# r-squared
tss <- t(Y-mean(Y)) %*% (Y-mean(Y))
r2 <- 1-ssr/tss
round(r2,3)
```

```
      [,1]
[1,] 0.389
```

### 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.

```
# calculate sum of squared errors
uhat <- yresid - x1resid%*%fw_bet
fw_ssr <- t(uhat)%*%uhat
round(fw_ssr,3)
```

```
       [,1]
[1,] 82.505
```

```
# calculate R2
fw_tss <- t(yresid-mean(yresid))%*%(yresid-mean(yresid))
fw_r2 <- 1-fw_ssr/fw_tss
round(fw_r2, 3)
```

```
      [,1]
[1,] 0.369
```

### Part c

The sum of squared errors is the same as in part (a). This is expected, e.g., Theorem 3.5 shows that the residuals from the FWL-type residual regression are the same as for the regression that includes both $X_1$ and $X_2$. This implies that the sum of squared errors will be the same too. On the other hand, $R^2$ is different because the total sum of squares is different between the case where it is calculated with $Y$ directly relative to using the residuals from $Y$ on $X_1$.

### 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

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

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 $\mu_k$.

### Part c

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

where the second equality holds because $1/n$ is a constant and it should be squared to come out of the variance, the third equality holds by passing the variance through the sum (in order for their not to be any covariance terms introduced here, it requires the "independence" part of iid; for this variance to be the same across all units requires the "identically distributed" part of iid), and the last equality holds because summing a constant $n$ times cancels one of the $n$'s from the denominator.

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

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

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

### 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),

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

For thinking about the sampling variance, first notice that

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

where the first equality holds by plugging in for $\mathbf{Y}$, the second equality holds because we are conditioning on $\mathbf{X}$, and the third equality holds from the way that $\mathbf{\Sigma}$ is defined in the textbook. Next,

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

where the first equality holds by plugging in for $\hat{\beta}$, the second 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 plugging in the expression for $\text{var}(\mathbf{Y}|\mathbf{X})$ that we derived above. 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 $\mathbb{E}[\tilde{\beta}|\mathbf{X}] = \beta$. 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

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

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

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

where the above result repeatedly uses $\boldsymbol{\Sigma}$ 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) $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2} = \mathbf{I}_n$, and $\mathbf{A}'\mathbf{X} = \mathbf{I}_k$ (due to linearity and unbiasedness as discussed above); the third equality holds by factoring out $\mathbf{A}'\boldsymbol{\Sigma}^{1/2}$ 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 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 $\beta$.