Homework 3 Solutions

Hansen 3.16

To start with, notice that

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

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

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

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 3.24

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)
```

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

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

Part b

```r
# 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
```

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

```
[1] 133.1331
```

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

```
[1] 0
```

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

```
[1] 0
```

```r
# 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 4.6

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

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

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

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

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.

7.7

(a)

β is defined as the coefficient of the linear projection of Y^* on X. Thus, $\beta = \mathbb{E}[XX']^{-1}\mathbb{E}[XY^*]$. Now, let's define

$$\tilde{\beta} = \underset{b}{\operatorname{argmin}}\ \mathbb{E}[(Y - X'b)^2]$$

so that $\tilde{\beta}$ is the coefficient from the linear projection of Y on X. Solving this, we get that

$$\begin{aligned}
\tilde{\beta} &= \mathbb{E}[XX']^{-1}\mathbb{E}[XY] \\
&= \mathbb{E}[XX']^{-1}\mathbb{E}[X(Y^* + u)] \\
&= \mathbb{E}[XX']^{-1}\mathbb{E}[XY^*] + \mathbb{E}[XX']^{-1}\underbrace{\mathbb{E}[Xu]}_{=0} \\
&= \beta
\end{aligned}$$

Thus, $\tilde{\beta} = \beta$.

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

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i (Y_i^* + u_i) \qquad (2)
\end{aligned}$$

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 2, 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 β.

(c)

Plugging in $Y_i^* = X_i'\beta + e_i$ into Equation 2 and multiplying by $\sqrt{n}$, we have that

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

where

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

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

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 β, but the asymptotic variance changes; it is likely to be bigger.

7.14

(a)

$$\hat{\theta} = \hat{\beta}_1\hat{\beta}_2$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ come from the regression of Y on X_1 and X_2.

(b)

First, notice that our usual arguments imply that

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2\end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \mathbf{V}_\beta)$$

where $\mathbf{V}_\beta = \mathbb{E}[XX']^{-1}\boldsymbol{\Omega}\mathbb{E}[XX']^{-1}$, where we take $X = (X_1, X_2)'$ and where $\boldsymbol{\Omega} = \mathbb{E}[XX'e^2]$. Note that $\mathbf{V}_\beta$ is a 2×2 variance matrix.

Next, notice that we can write $\theta = r(\beta_1, \beta_2)$ and $\hat{\theta} = r(\hat{\beta}_1, \hat{\beta}_2)$ where $r(b_1, b_2) = b_1 b_2$. Moreover, using a mean value theorem argument, we have that

$$r(\hat{\beta}_1, \hat{\beta}_2) = r(\beta_1, \beta_2) + \nabla r(\bar{\beta}_1, \bar{\beta}_2)' \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix}$$

where

$$\nabla r(\bar{\beta}_1, \bar{\beta}_2) := \begin{bmatrix} \frac{\partial r(b_1,b_2)}{\partial b_1} \\ \\ \frac{\partial r(b_1,b_2)}{\partial b_2} \end{bmatrix} \Bigg|_{b_1=\bar{\beta}_1, b_2=\bar{\beta}_2} = \begin{bmatrix} b_2 \\ \\ b_1 \end{bmatrix} \Bigg|_{b_1=\bar{\beta}_1, b_2=\bar{\beta}_2} = \begin{bmatrix} \bar{\beta}_2 \\ \\ \bar{\beta}_1 \end{bmatrix}$$

This implies that

$$\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta) &= \begin{bmatrix} \bar{\beta}_2 \\ \\ \bar{\beta}_1 \end{bmatrix}' \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} \\
&= \begin{bmatrix} \beta_2 \\ \\ \beta_1 \end{bmatrix}' \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} + \underbrace{\left(\begin{bmatrix} \bar{\beta}_2 \\ \\ \bar{\beta}_1 \end{bmatrix} - \begin{bmatrix} \beta_2 \\ \\ \beta_1 \end{bmatrix} \right)'}_{=o_p(1)} \underbrace{\sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix}}_{=O_p(1)} \\
&= \begin{bmatrix} \beta_2 \\ \\ \beta_1 \end{bmatrix}' \sqrt{n} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0, V)
\end{aligned}$$

where the second equality holds by adding and subtracting, the third equality holds because $\begin{pmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \end{pmatrix}$ is between $\hat{\beta}$ and β (and because $\hat{\beta}$ is consistent for β), and where

$$V = \begin{bmatrix} \beta_2 \\ \\ \beta_1 \end{bmatrix}' \mathbf{V}_\beta \begin{bmatrix} \beta_2 \\ \\ \beta_1 \end{bmatrix}$$

(c)

To calculate a 95% confidence interval, the main step is to estimate V. The natural estimate is given by

$$\begin{bmatrix} \hat{\beta}_2 \\ \\ \hat{\beta}_1 \end{bmatrix}' \hat{\mathbf{V}}_\beta \begin{bmatrix} \hat{\beta}_2 \\ \\ \hat{\beta}_1 \end{bmatrix}$$

and where we use the usual estimate of $\mathbf{V}_\beta$ that is given by

$$\hat{\mathbf{V}}_\beta = \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n X_i X_i' \hat{e}_i^2 \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1}$$

and then we can construct a 95% confidence interval by

$$\hat{C} = \left[\hat{\theta} \pm 1.96\frac{\sqrt{\widehat{V}}}{\sqrt{n}}\right]$$

7.28

(a)

I am going to include a little bit of extra detail about comparing "manually" calculated standard errors with those coming directly from R as I think this is interesting. For part of the problem, I'll compare to results from the R package `estimatr` which is popular among economists for computing heteroskedasticity robust standard errors.

```r
# 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
```

```r
# 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
```

```r
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
```

```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
```

```r
# 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
```

```r
# 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
```

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