Homework 2 Solutions

Hansen 3.2

Let's call $\tilde{\beta}$ and $\tilde{\mathbf{e}}$ the OLS estimates and residuals from the regression of $\mathbf{Y}$ on $\mathbf{Z}$. Notice that

$$\begin{aligned}
\tilde{\beta} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \\
&= ((\mathbf{X}\mathbf{C})'\mathbf{X}\mathbf{C})^{-1}(\mathbf{X}\mathbf{C})'\mathbf{Y} \\
&= (\mathbf{C}'\mathbf{X}'\mathbf{X}\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}'\mathbf{Y} \\
&= \mathbf{C}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'^{-1}\mathbf{C}'\mathbf{X}'\mathbf{Y} \\
&= \mathbf{C}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{C}^{-1}\hat{\beta}
\end{aligned}$$

where the second equality holds by plugging in $\mathbf{Z} = \mathbf{X}\mathbf{C}$, the third equality holds by taking the transpose of $\mathbf{X}\mathbf{C}$, the fourth equality holds because $\mathbf{C}$ and $\mathbf{X}'\mathbf{X}$ are nonsingular, the fifth equality holds by canceling the $\mathbf{C}'^{-1}\mathbf{C}'$, and the last equality holds by the definition of $\hat{\beta}$.

Now, for the residuals, notice that

$$\begin{aligned}
\tilde{\mathbf{e}} &= \mathbf{Y} - \mathbf{Z}\tilde{\beta} \\
&= \mathbf{Y} - \mathbf{X}\mathbf{C}\mathbf{C}^{-1}\hat{\beta} \\
&= \mathbf{Y} - \mathbf{X}\hat{\beta} \\
&= \hat{\mathbf{e}}
\end{aligned}$$

where this result holds just by plugging in and canceling terms. This says that the residuals from the regression of $\mathbf{Y}$ on $\mathbf{Z}$ are exactly the same as the residuals from the regression of $\mathbf{Y}$ on $\mathbf{X}$.

As a side-comment, a simple example of this problem would be something like scaling all the regressors by, say, 100. If you did this, it would change the value of the estimated coefficients (divide them by 100) but would fit the data equally well.

Hansen 3.3

$$\begin{aligned}
\mathbf{X}'\hat{\mathbf{e}} &= \mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\
&= \mathbf{X}'(\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\
&= \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\
&= \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{Y} \\
&= \mathbf{0}
\end{aligned}$$

which holds by plugging in for $\hat{\mathbf{e}}$ and canceling terms.

Hansen 3.5

The OLS coefficient from a regression of $\hat{\mathbf{e}}$ on $\mathbf{X}$ is given by

$$
\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{e}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} \\
&= \hat{\beta} - \hat{\beta} \\
&= 0
\end{aligned}
$$

where the first part of the third equality holds by the definition of $\hat{\beta}$ and the last part holds by canceling the terms involving $(\mathbf{X}'\mathbf{X})$.

Hansen 3.7

$$
\begin{aligned}
\mathbf{PX} &= \mathbf{P}\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{PX}_1 & \mathbf{PX}_2\end{bmatrix}
\end{aligned}
$$

Further, since $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$ and $\mathbf{PX} = \mathbf{X}$ (from the properties of the projection matrix $\mathbf{P}$), this implies that $\mathbf{PX}_1 = \mathbf{X}_1$.

Similarly,

$$
\begin{aligned}
\mathbf{MX} &= \mathbf{M}\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{MX}_1 & \mathbf{MX}_2\end{bmatrix}
\end{aligned}
$$

but we also know that $\mathbf{MX} = \mathbf{0}_{n\times k} = \begin{bmatrix}\mathbf{0}_{n\times k_1} & \mathbf{0}_{n\times k_2}\end{bmatrix}$ where, for example, $\mathbf{0}_{n\times k_1}$ is an $n \times k_1$ matrix of zeroes. This implies that $\mathbf{MX}_1 = \mathbf{0}_{n\times k_1}$.

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

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

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
# 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
```

```r
# 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.

Extra Question

a)

$$
\begin{aligned}
ATE &= \mathrm{E}[Y(1) - Y(0)] \\
&= \mathrm{E}[Y(1)] - \mathrm{E}[Y(0)] \\
&= \mathrm{E}[\mathrm{E}[Y(1)|X]] - \mathrm{E}[\mathrm{E}[Y(0)|X]] \\
&= \mathrm{E}[\mathrm{E}[Y(1)|X, D = 1]] - \mathrm{E}[\mathrm{E}[Y(0)|X, D = 0]] \\
&= \mathrm{E}[\mathrm{E}[Y|X, D = 1]] - \mathrm{E}[\mathrm{E}[Y|X, D = 0]]
\end{aligned}
$$

where the first equality is the definition of ATE, the second equality pushes the expectation through the difference, the third equality holds by the law iterated expectations, the fourth equality holds by unconfoundedness, and the last equality holds because $Y = Y(1)$ among the treated group and $Y = Y(0)$ among the untreated group. This shows that ATE is identified.

b) In class, we showed that $ATT = \mathrm{E}[Y|D = 1] - \mathrm{E}[\mathrm{E}[Y|X, D = 0]|D = 1]$. These are notably different. The expression for ATE takes the $\mathrm{E}[Y|X, D = 1]$ (the mean of Y conditional on X among the treated group) and averages it over the distribution of X for the entire population and then subtracts $\mathrm{E}[Y|X, D = 0]$ (the mean of Y conditional on X among the untreated group) averaged over the population distribution of X.

In contrast, ATT compares the mean of Y among the treated group to $\mathrm{E}[Y|X, D = 0]$ where this is averaged over the distribution of X *among the treated group*.

An intuition for why ATE involves averaging over the distribution of X for the entire population is that ATE is the average treatment effect for the entire population.

c) We have that

$$\begin{aligned} Y_i &= Y_i(0) + D_i(Y_i(1) - Y_i(0)) \\ &= X_i'\beta + e_i + \alpha D_i \\ &= \alpha D_i + X_i'\beta + e_i \end{aligned}$$

where the first equality holds by writing observed outcomes in terms of potential outcomes, the second equality uses the model for untreated potential outcomes and treatment effect homogeneity, and the last equality rearranges terms.

Furthermore, unconfoundedness implies that $\mathrm{E}[e|X, D] = 0$ which implies that α can be estimated from the regression of Y on D and X.

d) This is exactly the same regressions as we talked about in class after we had identified the ATT. This should not be surprising because, if we restrict the effect of participating in the treatment to be the same across all units, then $ATT = ATE = \alpha$.