Problem Set 3

Tate Mason

2025-03-09

Question 1 - Hansen 3.16

For the first regression, let's first show that residuals such that $\tilde{e} = Y - X_1\tilde{\beta}_1$. Further, for the second, $\hat{e} = Y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2$. Since the total sum of squares does not depend on the beta's, we will focus on the squared sum of residuals. $SSR_1 = \tilde{e}^T\tilde{e}$ and $SSR_2 = \hat{e}^T\hat{e}$. Further, $R_i^2 = 1 - (\frac{SSR_i}{SST})$. Because the second model includes the explanatory power of the first in the case of $\hat{\beta}_2 = 0$, as well as more when $\hat{\beta}_2 \neq 0$, $SSR_2 \leq SSR_1$. Because of this, $R_2^2 = 1 - (\frac{SSR_2}{SST}) \geq 1 - (\frac{SSR_1}{SST}) = R_1^2$. The two are equal in the case when $\hat{\beta}_2 = 0$, or there is no omitted variable in the first regression.

Question 2 - Hansen 3.24

```r
library(haven)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
dat <- read_dta('~/SchoolWork/Sem2/Metrics/PSets/PS3/cps09mar.dta')
sample <- (dat[,11]==4)&(dat[,12]==7)&(dat[,2]==0)
df <- dat[sample,]
y <- as.matrix(log(df[,5]/(df[,6]*df[,7])))
exp <- df[,1]-df[,4]-6
exp2 <- (exp^2)/100
x_df <- data.frame(
  education = df[,4],
  experience = exp,
  exp_squared = exp2,
  intercept = 1
)
x <- as.matrix(x_df)
xx <- t(x)%*%x
xy <- t(x)%*%y
beta <- solve(xx,xy)
```

Part A

```r
fit <- x %*% beta
resid <- y - fit
SST <- sum((y - mean(y))^2)
SSE <- sum(resid^2)
R2 <- 1 - (SSE/SST)
cat("Part (a):\n")
```

```
Part (a):
```

```r
cat("R^2:", R2, "\n")
```

```
R^2: 0.3875804
```

```r
cat("Sum of squared errors:", SSE, "\n\n")
```

```
Sum of squared errors: 83.01622
```

Part B

```
x1_df <- data.frame(
  experience = exp,
  exp_squared = exp2,
  intercept = 1
)
x1 <- as.matrix(x1_df)
xx1 <- t(x1)%*%x1
xy1 <- t(x1)%*%y
beta1 <- solve(xx1,xy1)
fitted1 <- x1%*%beta1
resid_lwage <- y - fitted1

edu <- as.matrix(df[,4])
xx2 <- t(x1) %*% x1
xy2 <- t(x1) %*% edu
beta2 <- solve(xx2,xy2)
fitted2 <- x1 %*% beta2
resid_edu <- edu - fitted2

x3 <- cbind(resid_edu, rep(1,nrow(df)))
xx3 <- t(x3)%*%x3
xy3 <- t(x3)%*%resid_lwage
betab <- solve(xx3, xy3)

fittedb <- x3 %*% betab
residb <- resid_lwage - fittedb

SSTb <- sum((resid_lwage - mean(resid_lwage))^2)
SSEb <- sum(residb^2)
R2b <- 1 - (SSEb/SSTb)
cat("Part (b):\n")
```

```
Part (b):
```

```
cat("R^2:", R2b, "\n")
```

```
R^2: 0.3651369
```

```
cat("Sum of squared errors:", SSEb, "\n\n")
```

```
Sum of squared errors: 83.01622
```

Part C

While the SSE is equal in both approaches, the R^2 differs slightly. This is because in the first part, we are using R^2 to measure how much education, experience, and experience squared explain total variation in log wage. In the second, we are using residuals which will have smaller variation than the raw variable.

Question 3 - Hansen 3.25

```
rm(list=ls())
library(haven)
dat <- read_dta('~/SchoolWork/Sem2/Metrics/PSets/PS3/cps09mar.dta')
sample <- (dat[,11]==4)&(dat[,12]==7)&(dat[,2]==0)
df <- dat[sample,]

y <- as.matrix(log(df[,5]/(df[,6]*df[,7])))
exp <- df[,1]-df[,4]-6
exp2 <- (exp^2)/100

x_df <- data.frame(
  education = df[,4],
  experience = exp,
  exp_squared = exp2,
  intercept = 1
)
x <- as.matrix(x_df)

xx <- t(x)%*%x
xy <- t(x)%*%y
beta <- solve(xx,xy)

fitted <- x %*% beta
resid <- y - fitted
```

```r
x1 <- as.matrix(df[,4])
x2 <- as.matrix(exp)
x1_sq <- x1^2
x2_sq <- x2^2
```

Part A

```r
sum_resid <- sum(resid)
cat("(a) Sum of residuals:", sum_resid, "\n")
```

```
(a) Sum of residuals: -2.207123e-13
```

Part B

```r
sum_x1_resid <- sum(x1*resid)
cat("(b) Sum of X1*residuals:", sum_x1_resid, "\n")
```

```
(b) Sum of X1*residuals: -6.975753e-12
```

Part C

```r
sum_x2_resid <- sum(x2*resid)
cat("(c) Sum of X2*residuals:", sum_x2_resid, "\n")
```

```
(c) Sum of X2*residuals: -6.394885e-13
```

Part D

```r
sum_x1sq_resid <- sum(x1_sq*resid)
cat("(d) Sum of X1^2*residuals:", sum_x1sq_resid, "\n")
```

```
(d) Sum of X1^2*residuals: 142.528
```

Part E

```r
sum_x2sq_resid <- sum(x2_sq*resid)
cat("(e) Sum of X2^2*residuals:", sum_x2sq_resid, "\n")
```

```
(e) Sum of X2^2*residuals: 7.560175e-12
```

Part F

```r
sum_fit_resid <- sum(fitted*resid)
cat("(f) Sum of fitted values*residuals:", sum_fit_resid, "\n")
```

```
(f) Sum of fitted values*residuals: -1.167733e-12
```

Part G

```r
sum_resid_sq <- sum(resid^2)
cat("(f) Sum of squared residuals:", sum_resid_sq, "\n")
```

```
(f) Sum of squared residuals: 83.01622
```

Question 4 - Hansen 4.6

By the given constraint of linear estimators, $\tilde{\beta} = CY$ such that C is a txp matrix of all constants. Variance can be computed via the following $var[\tilde{\beta}|X] = var[CY|X] = Cvar[Y|X]C^T = C\sigma^2CXC^T = \sigma^2CC^T$, following that $CX = I$ in the case of unbiasedness (Hansen 105). Now, using the Gauss-Markov theorem, $\hat{\beta} = (X^TX)^{-1}X^TY$ and $D = C - (X^TX)^{-1}X^t$. Via the definition $CX = I$, multiplying $D \times X$ yields $I = (X^TX)^{-1}X^TX$ and $DX = 0$. Now, we can infer $\tilde{\beta} = CY = [(X^TX)^{-1}X^T + D]Y = \hat{\beta} + DY$. Let's go back to variance: $var[\tilde{\beta}|X] = var[\hat{\beta}|X] + [DY|X] + 2cov[\hat{\beta}, DY|X]$. This can be expanded as follows $cov[\hat{\beta}, DY|X]cov[(X^TX)^{-1}X^TY, DY|X] = (X^TX)^{-1}X^Tcov[Y, DY|X]$. Expanding further, $cov[Y, DY|X] = cov[X\beta + \epsilon, D(X\beta + \epsilon)|X] = Dcov[\epsilon, \epsilon|X] = D\sigma^2$. Thus, $cov[\hat{\beta}, DY|X] = \sigma^2(X^TX)^{-1}X^TD$. Since, as found before, $DX = 0$ it follows that $X^TD^T = 0$. Thus, $cov[\hat{\beta}, DY|X] = 0$. Therefore, $var[\tilde{\beta}|X] = var[\hat{\beta}|X] + var[DY|X] = \sigma^2(X^TX)^{-1} + \sigma^2DD^T$. DD^T is positive semi-definite, therefore $var[\tilde{\beta}|X] \geq \sigma^2(X^TX)^{-1} = \sigma^2(X^T\Sigma^{-1}X)^{-1}$. In the text, in the case of linear estimation, $\Sigma = I$, thus our findings prove the given inequality.

Question 5 - Hansen 7.7

Part A

$\beta = [\mathbb{E}(XX^T)]^{-1}\mathbb{E}(XY)$. Substituting in the given Y, $\beta_{LP} = [\mathbb{E}(XX^T)]^{-1}\mathbb{E}(X(X^T\beta + e + u))$. Distributing the expectation, $\beta_{LP} = [\mathbb{E}(XX^T)]^{-1}(E(XX^T)\beta + \mathbb{E}(Xe) + \mathbb{E}(Xu))$. Given $\mathbb{E}(Xe),\ \mathbb{E}(Xu) = 0$, $\beta_{LP} = [\mathbb{E}(XX^T)]^{-1}\mathbb{E}(XX^T)\beta = \beta$ so, we can conclude that it is the true coefficient from the linear projection.

Part B

$plim(\hat{\beta}) = plim((X^TX)^{-1}X^TY)$. As before, substituting for Y, $plim(\hat{\beta}) = plim((X^TX)^{-1}X^T(X^T\beta + e + u))$. Again, we can distribute and simplify such that $plim(\hat{\beta}) = \beta + plim((X^TX)^{-1}X^Te) + plim((X^TX)^{-1}X^Tu)$. By law of large numbers, we can use the assumption in expectation the same way in this case, $plim((X^TX)^{-1}X^Te),\ plim((X^TX)^{-1}X^Tu) = 0$. Thus, $plim(\hat{\beta}) = \beta$. This states that as n gets sufficiently large, $\hat{\beta}$ is consistent for β.

Part C

We will go about this much the same way as above. $\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}((X^TX)^{-1}X^T(X^T\beta + e + u) - \beta)$. This can be expanded as such $\sqrt{n}((X^TX)^{-1}X^Te + (X^TX)^{-1}X^Tu)$. As in the book, let $Q = plim(\frac{X^TX}{n})$. By central limit theorem, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1})$ s.t. $\Omega = plim(\frac{1}{n}X^T(e+u)(e+u)^TX)$. Since e and u are i.i.d error terms, we can say that $var(u) = \sigma_u^2$ and $var(e) = \sigma_e^2$. Simplifying our Ω term, we have that $\Omega = plim(\frac{1}{n}X^t(ee^T + uu^T)X) = Q(\sigma_e^2 + \sigma_u^2)$. So, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}(\sigma_e^2 + \sigma_u^2))$. Thus, the distribution of $\sqrt{n}(\hat{\beta} - \beta)$ as n approaches ∞ is normal with mean 0 and variance $Q^{-1}(\sigma_e^2 + \sigma_u^2)$.

Question 6 - Hansen 7.14

Part A

$\hat{\theta} = \hat{\beta}_1\hat{\beta}_2$ given $\theta = \beta_1\beta_2$. This is appropriate as $\hat{\beta}_1, \hat{\beta}_2$ are the OLS estimators of Y on X_1, X_2, respectively.

Part B

Under standard conditions, $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0,\Sigma)$ s.t. $\hat{\beta}=(\hat{\beta}_1,\hat{\beta}_2)^T, \beta=(\beta_1,\beta_2)^T$ and Σ is the asymptotic covariance matrix of the OLS estimators. Using the delta method s.t. $g(\beta_1,\beta_2)=\beta_1\beta_2$, we have that $\nabla g(\beta)=(\frac{\partial g}{\partial \beta_1},\frac{\partial g}{\partial \beta_2})=(\beta_2,\beta_1)$. Further, the asymptotic distribution of estimator $\hat{\theta}$ is $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0,\nabla g(\beta)^T\Sigma\nabla g(\beta))$. Now, substituting in, $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0,\beta_2^2\sigma_{11}+2\beta_1\beta_2\sigma_{12}+\beta_1^2\sigma_{22})$. Such that $\sigma_{11},\sigma_{22},\sigma_{12}$ are the asymptotic variance of $\hat{\beta}_1$, $\hat{\beta}_2$, and covariance between $\hat{\beta}_1,\hat{\beta}_2$. Therefore, $\hat{\theta}$ is asymptotically normal with mean 0 and variance $\frac{1}{n}(\beta_2^2\sigma_{11}+2\beta_1\beta_2\sigma_{12}+\beta_1^2\sigma_{22})$.

Part C

$\hat{\theta} \pm 1.96 \times \sqrt{\hat{\beta}_2^2\hat{\sigma}_{11}+2\hat{\beta}_1\hat{\beta}_2\hat{\sigma}_{12}+\hat{\beta}_1\hat{\sigma}_{22}}$. This provides a range of plausible values for true parameter $\theta=\beta_1\beta_2$ based on the estimated variances of $\hat{\beta}_1,\hat{\beta}_2$, $\hat{\sigma}_{11},\hat{\sigma}_{22}$ and covariance $\hat{\sigma}_{12}$.

Question 7 - Hansen 7.28 (part (a) only)

```
rm(list = ls())

library(haven)
dat <- read_dta('~/SchoolWork/Sem2/Metrics/PSets/PS3/cps09mar.dta')
sample <- (dat[,11]==9)&(dat[,2]==0)
df <- dat[sample,]

y <- as.matrix(log(df[,5]/(df[,6]*df[,7])))
exp <- df[,1]-df[,4]-6
exp2 <- (exp^2)/100

x_df <- data.frame(
  education = df[,4],
  experience = exp,
  exp_squared = exp2,
  intercept = 1
)
x <- as.matrix(x_df)

xx <- t(x)%*%x
xy <- t(x)%*%y
```

```r
beta <- solve(xx,xy)

fitted <- x %*% beta
resid <- y - fitted

n <- nrow(x)
k <- ncol(x)
xx_inv <- solve(t(x)%*%x)

df <- n-k

hc0 <- matrix(0, nrow=k, ncol=k)
for (i in 1:n) {
  xi <- matrix(x[i,], nrow=k)
  hc0 <- hc0 + resid[i]^2 * (xi %*% t(xi))
}
cov_hc0 <- xx_inv %*% hc0 %*% xx_inv
se_hc0 <- sqrt(diag(cov_hc0))

# Calculate HC1 (small sample correction)
cov_hc1 <- cov_hc0 * (n/(n-k))
se_hc1 <- sqrt(diag(cov_hc1))

# Calculate HC2 (leverage adjustment)
# Compute the hat matrix diagonal (leverage values)
h <- rep(0, n)
for (i in 1:n) {
  h[i] <- x[i,] %*% xx_inv %*% x[i,]
}

# Calculate HC2
hc2 <- matrix(0, nrow=k, ncol=k)
for (i in 1:n) {
  xi <- matrix(x[i,], nrow=k)
  hc2 <- hc2 + (resid[i]^2/(1-h[i])) * (xi %*% t(xi))
}
cov_hc2 <- xx_inv %*% hc2 %*% xx_inv
se_hc2 <- sqrt(diag(cov_hc2))

# Calculate HC3 (more conservative leverage adjustment)
hc3 <- matrix(0, nrow=k, ncol=k)
for (i in 1:n) {
```

```r
  xi <- matrix(x[i,], nrow=k)
  hc3 <- hc3 + (resid[i]^2/((1-h[i])^2)) * (xi %*% t(xi))
}
cov_hc3 <- xx_inv %*% hc3 %*% xx_inv
se_hc3 <- sqrt(diag(cov_hc3))

var_labels <- c("Education", "Experience", "Experience Squared", "Intercept")
result <- function(coef, se) {
  formatted <- sprintf("%.4f (%.4f)", coef, se)
  return(formatted)
}

for (i in 1:k) {
  cat(var_labels[i], "\n")
  cat("  Coefficient: ", sprintf("%.6f", beta[i]), "\n")
  cat("  Std. Error:  ", sprintf("%.6f", se_hc3[i]), "\n")
}
```

```
Education
  Coefficient:  0.046560
  Std. Error:   0.066523
Experience
  Coefficient:  0.047109
  Std. Error:   0.036276
Experience Squared
  Coefficient:  -0.090884
  Std. Error:   0.084480
Intercept
  Coefficient:  2.031521
  Std. Error:   0.874656
```