

ECON 709B - Problem Set 3

Alex von Hafften*

12/1/2020

1. 3.24 Use the data set from Section 3.22 and the sub-sample used for equation (3.50) (see Section 3.25 for data construction).

(a) Estimate equation (3.50) and compute the equation R^2 and sum of squared errors.

```
library(tidyverse)

cps09mar <- read_delim("cps09mar.txt",
                      delim = "\t",
                      col_names = c("age", "female", "hisp", "education", "earnings",
                                     "hours", "week", "union", "uncov", "region", "race",
                                     "marital"),
                      col_types = "dddddddddddd") %>%
  mutate(experience = age - education - 6,
         experience_2 = (experience^2)/100,
         wage = earnings / (hours*week),
         l_wage = log(wage),
         constant = 1) %>%
  filter(race == 4,
         marital == 7,
         female == 0,
         experience < 45)

y <- cps09mar$l_wage

x <- cps09mar %>%
  select(education, experience, experience_2, constant) %>%
  as.matrix() %>%
  unname()

n <- dim(x)[1]
i <- diag(nrow = n, ncol = n)

beta <- solve(t(x) %*% x) %*% t(x) %*% y
print(beta)

##           [,1]
## [1,]  0.14430729
## [2,]  0.04263326
## [3,] -0.09505636
## [4,]  0.53089068
```

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

```
p_x <- x %*% solve(t(x) %*% x) %*% t(x)
m_x <- i - p_x
e_hat <- m_x %*% y
sse_a <- sum(e_hat^2)
print(sse_a)
```

```
## [1] 82.505
```

```
y_bar <- mean(y)
r_squared_a <- 1 - sse_a/sum((y-y_bar)^2)
print(r_squared_a)
```

```
## [1] 0.3893207
```

- (b) Re-estimate the slope on education using the residual regression approach. Regress $\log(\text{wage})$ on experience and its square, regress education on experience and its square, and the residuals on the residuals. Report the estimates from this final regression, along with the equation R^2 and sum of squared errors. Does the slope coefficient equal the value in (3.50)? Explain.

```
x_1 <- cps09mar %>%
  select(experience, experience_2, constant) %>%
  as.matrix() %>%
  unname()

m_x_1 <- i - x_1 %*% solve(t(x_1) %*% x_1) %*% t(x_1)

z <- cps09mar$education

u_hat <- m_x_1 %*% z
v_hat <- m_x_1 %*% y

beta_1 <- (t(u_hat) %*% u_hat)^(-1) %*% t(u_hat) %*% v_hat
print(beta_1)
```

```
## [1]
```

```
## [1,] 0.1443073
```

```
m_u_hat <- i - u_hat %*% (t(u_hat) %*% u_hat)^(-1) %*% t(u_hat)
d_hat <- m_u_hat %*% v_hat
sse_b <- sum(d_hat^2)
print(sse_b)
```

```
## [1] 82.505
```

```
v_hat_bar <- mean(v_hat)
r_squared_b <- 1 - sse_b/sum((v_hat - v_hat_bar)^2)
print(r_squared_b)
```

```
## [1] 0.368738
```

- (c) Are the R^2 and sum of squared errors from parts (a) and (b) equal? Explain.

The SSE is the same between (a) and (b) because the residuals are the same no matter if you do a partition regression or a regression with all variables at once. The R^2 in part (b) is lower than in part (a). Since the SSE is the same, both parts have the same denominator. Because there is some explanatory power from the first regression, the denominator is larger in part (a) than in part (b).

3.25 Estimate equation (3.50) as in part (a) of the previous question. Let \hat{e}_i be the OLS residual, \hat{Y}_i the predicted value from the regression, X_{1i} be education and X_{2i} be experience.

```
x1 <- x[,1]
x2 <- x[,2]
p_x <- x %*% solve(t(x) %*% x) %*% t(x)
m_x <- i - p_x
y_hat <- p_x %*% y
e_hat <- m_x %*% y
```

Numerically calculate the following. Are these calculations consistent with the theoretical properties of OLS? Explain.

Yes. When working with a computer, there is always some rounding, so we should interpret very small numbers as zero in terms of consistency with the theoretical predictions.

(a) $\sum_{i=1}^n \hat{e}_i$

```
sum(e_hat)
```

```
## [1] 1.479396e-12
```

The sum of residuals is zero.

(b) $\sum_{i=1}^n X_{1i} \hat{e}_i$

```
sum(x1 * e_hat)
```

```
## [1] 2.403766e-11
```

Residuals are orthogonal to any partition of X .

(c) $\sum_{i=1}^n X_{2i} \hat{e}_i$

```
sum(x2 * e_hat)
```

```
## [1] 4.451689e-12
```

Residuals are orthogonal to any partition of X .

(d) $\sum_{i=1}^n X_{1i}^2 \hat{e}_i$

```
sum(x1^2 * e_hat)
```

```
## [1] 133.1331
```

[No particular theoretical property.]

(e) $\sum_{i=1}^n X_{2i}^2 \hat{e}_i$

```
sum(x2^2 * e_hat)
```

```
## [1] -8.045165e-11
```

Note that squared experience is a column in X , so it is a partition of X . Residuals are orthogonal to any partition of X .

(f) $\sum_{i=1}^n \hat{Y}_i \hat{e}_i$

```
sum(y_hat * e_hat)
```

```
## [1] 4.701786e-12
```

Residuals are orthogonal to fitted values.

(g) $\sum_{i=1}^n X_{1i} \hat{e}_i^2$

```
sum(x1 * e_hat^2)
```

[1] 1240.049

[No particular theoretical property.]

2. 7.2 Take the model $y = X'\beta + e$ with $E[Xe] = 0$. Define the ridge regression estimator: $\hat{\beta} = (\sum_{i=1}^n X_i X_i' + \lambda I_k)^{-1} (\sum_{i=1}^n X_i Y_i)$ here $\lambda > 0$ is a fixed constant. Find the probability limit of $\hat{\beta}$ as $n \rightarrow \infty$. Is $\hat{\beta}$ consistent for β ?

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

As $n \rightarrow \infty$,

$$\frac{1}{n} \lambda I_k \rightarrow 0$$

Furthermore, by the weak law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n X_i X_i' \rightarrow_p E[X_i X_i']$$

$$\frac{1}{n} \sum_{i=1}^n X_i Y_i \rightarrow_p E[X_i Y_i]$$

Thus, $\hat{\beta}$ is consistent: $\hat{\beta} \rightarrow_p (E[X_i X_i'])^{-1} E[X_i Y_i] = \beta$.

7.3 For the ridge regression estimator, set $\lambda = cn$ where $c > 0$ is fixed as $n \rightarrow \infty$. Find the probability limit of $\hat{\beta}$ as $n \rightarrow \infty$.

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

Thus, $\hat{\beta}$ is not consistent: $\hat{\beta} \rightarrow_p (E[X_i X_i'] + c I_k)^{-1} E[X_i Y_i] \neq \beta$

7.4 Verify some of the calculations reported in Section 7.4. Specifically, suppose that X_1 and X_2 only take the values $\{-1, +1\}$, symmetrically, with $P[X_1 = X_2 = 1] = P[X_1 = X_2 = -1] = 3/8$, $P[X_1 = 1, X_2 = -1] = P[X_1 = -1, X_2 = 1] = 1/8$, $E[e_i^2 | X_1 = X_2] = 5/4$, and $E[e_i^2 | X_1 \neq X_2] = 1/4$. Verify the following:

- (a) $E[X_1] = 0$

$$\begin{aligned} E[X_1] &= P[X_1 = 1] - P[X_1 = -1] \\ &= P[X_1 = X_2 = 1] + P[X_1 = 1, X_2 = -1] - P[X_1 = X_2 = -1] - P[X_1 = -1, X_2 = 1] \\ &= 3/8 + 1/8 - 3/8 - 1/8 \\ &= 0 \end{aligned}$$

(b) $E[X_1^2] = 1$

If $X_1 = 1$, then $X_1^2 = 1$. If $X_1 = -1$, then $X_1^2 = 1$. Thus $X_1^2 = 1$ for all X_1 . Thus, trivially, $E[X_1^2] = 1$.

(c) $E[X_1 X_2] = \frac{1}{2}$

$$\begin{aligned} E[X_1 X_2] &= P[X_1 = X_2 = 1] - P[X_1 = 1, X_2 = -1] + P[X_1 = X_2 = -1] - P[X_1 = -1, X_2 = 1] \\ &= 3/8 - 1/8 + 3/8 - 1/8 \\ &= 1/2 \end{aligned}$$

(d) $E[e^2] = 1$

$$\begin{aligned} E[e^2] &= P[X_1 = X_2]E[e_i^2|X_1 = X_2] + P[X_1 \neq X_2]E[e_i^2|X_1 \neq X_2] \\ &= (P[X_1 = X_2 = 1] + P[X_1 = X_2 = -1])E[e_i^2|X_1 = X_2] \\ &\quad + (P[X_1 = 1, X_2 = -1] + P[X_1 = -1, X_2 = 1])E[e_i^2|X_1 \neq X_2] \\ &= (3/8 + 3/8)(5/4) + (1/8 + 1/8)(1/4) \\ &= (3/4)(5/4) + (1/4)(1/4) \\ &= (15/16) + (1/16) \\ &= 1 \end{aligned}$$

(e) $E[X_1^2 e^2] = 1$

Notice that $X_1^2 = 1$ for $X_1 = -1$ or $X_1 = 1$, so

$$\begin{aligned} E[X_1^2 e^2] &= E[e^2] \\ &= 1 \end{aligned}$$

(f) $E[X_1 X_2 e^2] = \frac{7}{8}$

$$\begin{aligned} E[X_1 X_2 e^2] &= P[X_1 = X_2]E[e^2|X_1 = X_2] - P[X_1 \neq X_2]E[e^2|X_1 \neq X_2] \\ &= (3/4)E[e^2|X_1 = X_2] - (1/4)E[e^2|X_1 \neq X_2] \\ &= (3/4)(5/4) - (1/4)(1/4) \\ &= 7/8 \end{aligned}$$

3. 7.8 Find the asymptotic distribution of $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ as $n \rightarrow \infty$.

Let us assume OLS 1, 2, 3, and 4. Let us further assume that $E[e_i^4]$ exists. From Theorem 7.4, we know that $\hat{\sigma}^2 \rightarrow_p \sigma^2$. Using equation 7.18:

$$\begin{aligned} \sqrt{n}(\hat{\sigma}^2 - \sigma^2) &= \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n e_i^2 - 2 \left(\frac{1}{n} \sum_{i=1}^n e_i X_i' \right) (\hat{\beta} - \beta) + (\hat{\beta} - \beta)' \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right) (\hat{\beta} - \beta) - \sigma^2 \right] \\ &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n e_i^2 - \sigma^2 \right) - 2 \left(\frac{1}{n} \sum_{i=1}^n e_i X_i' \right) \sqrt{n}(\hat{\beta} - \beta) + (\hat{\beta} - \beta)' \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right) \sqrt{n}(\hat{\beta} - \beta) \end{aligned}$$

We can consider each term individually. By the central limit theorem, the first term is

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n e_i^2 - \sigma^2 \right) \rightarrow_d N(0, V)$$

where $V = \text{Var}(e_i^2) = E[e_i^4] - (E[e_i^2])^2$.

The other terms converge based on results we discussed in lecture that depend on the weak law of large numbers or the central limit theorem:

$$\frac{1}{n} \sum_{i=1}^n e_i X_i' \rightarrow_p E[e_i X_i'] = 0$$

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N(0, \sigma^2 E(X_i X_i')^{-1})$$

$$\hat{\beta} - \beta \rightarrow_p 0$$

$$\frac{1}{n} \sum_{i=1}^n X_i X_i' \rightarrow_p E[X_i X_i'] < \infty$$

Therefore, $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \rightarrow_d N(0, E[e_i^4] - (E[e_i^2])^2)$.

4. 7.9a The model is $Y = X\beta + e$ with $E[e|X] = 0$ and $X \in \mathbb{R}$. Consider the two estimators: $\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$ and $\tilde{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{X_i}$. Under the stated assumptions are both estimators consistent for β ?

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i (X_i \beta + e_i)}{\sum_{i=1}^n X_i^2} = \frac{\sum_{i=1}^n X_i^2 \beta}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i e_i}{\sum_{i=1}^n X_i^2} = \beta + \frac{\frac{1}{n} \sum_{i=1}^n X_i e_i}{\frac{1}{n} \sum_{i=1}^n X_i^2}$$

By the weak law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n X_i X_i \rightarrow_p E[XX]$$

$$\frac{1}{n} \sum_{i=1}^n X_i e_i \rightarrow_p E[Xe] = E[E[e|X]X] = 0$$

Thus, $\hat{\beta} \rightarrow_p \beta$.

$$\tilde{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{X_i \beta + e_i}{X_i} = \beta + \frac{1}{n} \sum_{i=1}^n \frac{e_i}{X_i}$$

By the weak law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n \frac{e_i}{X_i} \rightarrow_p E\left[\frac{e_i}{X_i}\right] = E\left[\frac{1}{X_i} E[e_i|X_i]\right] = 0$$

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

5. 7.10 In the homoskedastic regression model $y = X'\beta + e$ with $E[e|x] = 0$ and $E[e^2|X] = \sigma^2$ suppose $\hat{\beta}$ is the OLS estimator of β with covariance matrix estimator $\hat{V}_{\hat{\beta}}$ based on a sample of size n . Let $\hat{\sigma}^2$ be the estimator of σ^2 . You wish to forecast an out-of-sample value of Y_{n+1} given that $X_{n+1} = x$. Thus the available information is the sample, the estimates $(\hat{\beta}, \hat{V}_{\hat{\beta}}, \hat{\sigma}^2)$, the residuals \hat{e}_i , and the out-of-sample value of the regressors X_{n+1} .

- (a) Find a point forecast of Y_{n+1} .

$$\hat{Y}_{n+1} = x'\hat{\beta}$$

- (b) Find an estimator of the variance of this forecast.

$$\text{var}(\hat{Y}_{n+1}) = \text{var}(X'_{n+1}\hat{\beta})$$

6. 7.13 Consider an iid sample $\{Y_i, X_i\}, i = 1, \dots, n$ where Y and X are scalar. Consider the reverse projection model $X = Y\gamma + u$ with $E[Yu] = 0$ and define the parameter of interest as $\theta = 1/\gamma$.

- (a) Propose an estimator $\hat{\gamma}$ of γ .

$$\hat{\gamma} = \left(\sum_{i=1}^n Y_i^2\right)^{-1} \sum_{i=1}^n Y_i X_i$$

- (b) Propose an estimator $\hat{\theta}$ of θ .

$$\hat{\theta} = \left(\sum_{i=1}^n Y_i X_i\right)^{-1} \sum_{i=1}^n Y_i^2$$

- (c) Find the asymptotic distribution of $\hat{\theta}$.

$$\hat{\theta} = \left(\sum_{i=1}^n Y_i(Y_i\gamma + u_i)\right)^{-1} \sum_{i=1}^n Y_i^2$$

$$\sqrt{n}(\hat{\theta} - \theta) =$$

- (d) Find an asymptotic standard error for $\hat{\theta}$.

...

- 7.14 Take the model $Y = X_1\beta_1 + X_2\beta_2 + e_i$ and $E[X_i e_i] = 0$ with both $\beta_1 \in \mathbb{R}$ and $\beta_2 \in \mathbb{R}$, and define the parameter $\theta = \beta_1\beta_2$.

- (a) What is the appropriate estimator $\hat{\theta}$ for θ ?

...

- (b) Find the asymptotic distribution of $\hat{\theta}$ under standard regularity conditions.

...

- (c) Show how to calculate an asymptotic 95% confidence interval for θ .

...

- 7.15 Take the linear model $Y = X\beta + e$ with $E[e|X] = 0$ and $X_i \in \mathbb{R}$. Consider the estimator $\hat{\beta} = \frac{\sum_{i=1}^n X_i^3 Y_i}{\sum_{i=1}^n X_i^4}$. Find the asymptotic distribution of $\hat{n}(\hat{\beta} - \beta)$ as $n \rightarrow \infty$.

...

7.17 An economist reports a set of parameter estimates, including the coefficient estimates $\hat{\beta}_1 = 1.0$, $\hat{\beta}_2 = 0.8$, and standard errors $s(\hat{\beta}_1) = 0.07$ and $s(\hat{\beta}_2) = 0.07$. The author writes “The estimates show that β_1 is larger than β_2 .”

- (a) Write down the formula for an asymptotic 95% confidence interval for $\theta = \beta_1 - \beta_2$, expressed as a function of $\hat{\beta}_1$, $\hat{\beta}_2$, $s(\hat{\beta}_1)$, $s(\hat{\beta}_2)$, and $\hat{\rho}$, where $\hat{\rho}$ is the estimated correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$.

...

- (b) Can $\hat{\rho}$ be calculated from the reported information?

...

- (c) Is the author correct? Does the reported information support the author’s claim?

...

8. 7.19 Take the model $y = X'\beta + e$ with $E[Xe] = 0$ and suppose you have observations $i = 1, \dots, 2n$ (The number of observations is $2n$.) You randomly split the sample in half, (each has n observations), calculate $\hat{\beta}_1$ by least squares on the first sample, and $\hat{\beta}_2$ by least squares on the second sample. What is the asymptotic distribution of $\sqrt{n}(\hat{\beta}_1 - \hat{\beta}_2)$?

...

9. Suppose $y_i = 1 + x_i\gamma + \varepsilon_i$, where y_i, x_i, ε_i are scalar. Define $w_i = (1 \ x_i)'$ and $\beta = (1 \ \gamma)'$. Assume that x_i has a discrete distribution: $P(x_i = 1) = P(x_i = 4/3) = P(x_i = 5/3) = P(x_i = 1/4)$. We will use the following assumptions: (A0) (y_i, x_i) iid, (A1) $E(\varepsilon_i|w_i) = 0$, (A1') $E(w_i\varepsilon_i) = 0$, and (A2) $\text{var}(\varepsilon_i|w_i) = \sigma^2$. Assume that you will observe data $(y_1, x_1), \dots, (y_n, x_n)$ (a sample of size n). Below state any additional assumptions needed to obtain your answers. Consider the following OLS estimator from regressing y_i on w_i using only observations where $x_i = 1$ or $x_i = 2$: $\hat{\beta} = [\frac{1}{n} \sum_{i=1}^n w_i w_i' 1\{x_i \in \{1, 2\}\}]^{-1} \frac{1}{n} \sum_{i=1}^n w_i y_i 1\{x_i \in \{1, 2\}\}$ where $1\{A\}$ is an indicator function for the event A .

- (a) Under (A0) and (A1), does $\hat{\beta} \rightarrow^p \beta$?

...

- (b) Under (A0) and (A1'), does $\hat{\beta} \rightarrow^p \beta$?

...

- (c) Under (A0), (A1), and (A2), what is the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta)$? Simplify as much as possible.

...

- (d) Consider the following OLS estimator from regressing y_i on w_i using only observation where $x_i = \frac{4}{3}$ or $x_i = \frac{5}{3}$: $\hat{\beta} = [\frac{1}{n} \sum_{i=1}^n w_i w_i' 1\{x_i \in \{\frac{4}{3}, \frac{5}{3}\}\}]^{-1} \frac{1}{n} \sum_{i=1}^n w_i y_i 1\{x_i \in \{\frac{4}{3}, \frac{5}{3}\}\}$. Let $\hat{\beta}_2$ and $\hat{\hat{\beta}}_2$ denote the second elements of $\hat{\beta}$ and $\hat{\hat{\beta}}$. Note that $\hat{\beta}_2$ and $\hat{\hat{\beta}}_2$ are estimators for γ . Under (A0), (A1), and (A2), which estimator for γ do you prefer $\hat{\beta}_2$ and $\hat{\hat{\beta}}_2$? Explain.

...

- (e) Consider the OLS estimator, $\hat{\alpha}$, from regressing y_i on x_i (no constant term) using only the observation where $x_i = 1$ or $x_i = 2$. Under (A0), (A1), and (A2), what is the probability limit of $\hat{\alpha}$?

...

- (f) Let α denote your answer to part (e). Under (A0), (A1), and (A2), what is the asymptotic distribution of $\sqrt{n}(\hat{\alpha} - \alpha)$?

...