

STAT 666 HW #4

Maggie Buffum

May 30, 2019

Problem 1

Prove the expression (8.1.3):

$$\hat{\beta}_2 = \frac{\hat{\epsilon}'(\mathbf{x}_2|\mathbf{X}_1)\hat{\epsilon}(\mathbf{y}|\mathbf{X}_1)}{\hat{\epsilon}'(\mathbf{x}_2|\mathbf{X}_1)\hat{\epsilon}(\mathbf{x}_2|\mathbf{X}_1)}$$

given that $\hat{\epsilon}(Y|\mathbf{X}_1) = (\mathbf{I} - \mathbf{P}_1)\mathbf{y}$ are the ordinary residuals from a fit of Y on \mathbf{X}_1 and $\hat{\epsilon}(\mathbf{x}_2|\mathbf{X}_1) = (\mathbf{I} - \mathbf{P}_1)\mathbf{x}_2$ the ordinary residuals from a fit of X_2 on \mathbf{X}_1 . Also, $\mathbf{X} = (\mathbf{X}_1, \mathbf{x}_2)$, and \mathbf{P}_1 is the projection matrix corresponding to \mathbf{X}_1 .

Starting with expression 8.1.3, we will show that we can derive the OLS estimate for β_2

$$\begin{aligned} & \frac{\hat{\epsilon}'(\mathbf{x}_2|\mathbf{X}_1)\hat{\epsilon}(\mathbf{y}|\mathbf{X}_1)}{\hat{\epsilon}'(\mathbf{x}_2|\mathbf{X}_1)\hat{\epsilon}(\mathbf{x}_2|\mathbf{X}_1)} \\ &= \frac{[(I - P_1)\mathbf{x}_2]'[(I - P_1)\mathbf{y}]}{[(I - P_1)\mathbf{x}_2]'[(I - P_1)\mathbf{x}_2]} \\ &= \frac{\mathbf{x}_2'(I - P_1)'(I - P_1)\mathbf{y}}{\mathbf{x}_2'(I - P_1)'(I - P_1)\mathbf{x}_2} \\ &= \frac{\mathbf{x}_2'(I - P_1)(I - P_1)\mathbf{y}}{\mathbf{x}_2'(I - P_1)(I - P_1)\mathbf{x}_2} \\ &= \frac{\mathbf{x}_2'(I - P_1)\mathbf{y}}{\mathbf{x}_2'(I - P_1)\mathbf{x}_2} \end{aligned}$$

Now consider Example 2.1.2, in which the partitioned matrix $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$ has corresponding projection matrices $P_1 = A_1(A_1'A_1)^{-1}A_1'$ and $P_2 = B(B'B)^{-1}B'$, where $B = (I - P_1)A_2$. Translating to this situation, let $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$, $P_1 = X_1(X_1'X_1)^{-1}X_1'$, $P_2 = B(B'B)^{-1}B'$, where $B = (I - P_1)X_2$. Then the estimate of β_2 is

$$\begin{aligned} \hat{\beta}_2 &= (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{y} \\ &= ([(I - \mathbf{P}_1)\mathbf{x}_2]' [(I - \mathbf{P}_1)\mathbf{x}_2])^{-1} [(I - \mathbf{P}_1)\mathbf{x}_2]'\mathbf{y} \\ &= (\mathbf{x}_2'(I - \mathbf{P}_1)'(I - \mathbf{P}_1)\mathbf{x}_2)^{-1} \mathbf{x}_2'(I - \mathbf{P}_1)'\mathbf{y} \\ &= (\mathbf{x}_2'(I - \mathbf{P}_1)\mathbf{x}_2)^{-1} \mathbf{x}_2'(I - \mathbf{P}_1)\mathbf{y} \\ &= \frac{\mathbf{x}_2'(I - \mathbf{P}_1)\mathbf{y}}{\mathbf{x}_2'(I - \mathbf{P}_1)\mathbf{x}_2} \end{aligned}$$

Problem 2

Compute the OLS estimate in the Example 8.1.3. Show the unbiasedness of the estimate, and compute the variance as well. Compare the variance of the OLS estimate with that of the WLS estimate. Hint: Refer to Example 4.5.1.

In Example 8.1.1 we are asked to consider the simple regression model

$$Y_i = \beta_1 X_i + \epsilon_i, \quad i = 1, \dots, N,$$

where $\epsilon_i \sim N(0, \sigma^2 X_i^2)$, $X_i \neq 0$. Although uncorrelated, the variance of the errors increase with increases in the magnitudes of the X_i s and violate the assumption of constant variance in the errors.

Let's first find $\hat{\beta}_{OLS}$ and its variance. Knowing that $S(\beta) = \sum_{i=1}^N (Y_i - \beta_1 X_i)^2$, we evaluate the OLS estimate of β_1 as

$$\begin{aligned} \frac{dS(\beta)}{d\beta_1} &= 2 \sum_{i=1}^N (Y_i - \beta_1 X_i)(-X_i) = (\text{set}) 0 \\ \implies \sum_{i=1}^N Y_i X_i - \beta_1 \sum_{i=1}^N X_i^2 &= 0 \\ \implies \beta_1 \sum_{i=1}^N X_i^2 &= \sum_{i=1}^N Y_i X_i \\ \implies \hat{\beta}_{1,OLS} &= \frac{\sum_{i=1}^N Y_i X_i}{\sum_{i=1}^N X_i^2} \end{aligned}$$

Note that the OLS estimate is still unbiased for β_1 :

$$\begin{aligned} E[\hat{\beta}_{1,OLS}] &= E \left[\frac{\sum_{i=1}^N Y_i X_i}{\sum_{i=1}^N X_i^2} \right] = \frac{1}{\sum_{i=1}^N X_i^2} E \left[\sum_{i=1}^N Y_i X_i \right] = \frac{1}{\sum_{i=1}^N X_i^2} \sum_{i=1}^N E[Y_i] E[X_i] \\ &= \frac{1}{\sum_{i=1}^N X_i^2} \sum_{i=1}^N E[Y_i] X_i = \frac{1}{\sum_{i=1}^N X_i^2} \sum_{i=1}^N X_i (X_i \beta_1) = \beta_1 \end{aligned}$$

The estimated variance of $\hat{\beta}_{1,OLS}$ is

$$\begin{aligned} Var(\hat{\beta}_{1,OLS}) &= Var \left(\frac{\sum_{i=1}^N Y_i X_i}{\sum_{i=1}^N X_i^2} \right) = \frac{1}{\left(\sum_{i=1}^N X_i^2 \right)^2} Var \left(\sum_{i=1}^N Y_i X_i \right) = \frac{1}{\left(\sum_{i=1}^N X_i^2 \right)^2} \sum_{i=1}^N Var(Y_i X_i) \\ &= \frac{1}{\left(\sum_{i=1}^N X_i^2 \right)^2} \sum_{i=1}^N X_i^2 Var(Y_i) = \frac{\sum_{i=1}^N X_i^2 \sigma^2 X_i^2}{\left(\sum_{i=1}^N X_i^2 \right)^2} = \frac{\sigma^2 \sum_{i=1}^N X_i^4}{\left(\sum_{i=1}^N X_i^2 \right)^2} \end{aligned}$$

The WLS estimate of β_1 and the estimated variance of the Y_i s were provided in Example 8.1.1 as

$$\hat{\beta}_{1,WLS} = \frac{1}{N} \sum_{i=1}^N \frac{Y_i}{X_i}$$

Let's find the variance around $\hat{\beta}_{1,WLS}$:

$$\begin{aligned} Var(\hat{\beta}_{1,WLS}) &= Var \left(\frac{1}{N} \sum_{i=1}^N \frac{Y_i}{X_i} \right) = \frac{1}{N^2} Var \left(\sum_{i=1}^N \frac{Y_i}{X_i} \right) = \frac{1}{N^2} \sum_{i=1}^N Var \left(\frac{Y_i}{X_i} \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N \frac{1}{X_i^2} Var(Y_i) = \frac{1}{N^2} \sum_{i=1}^N \frac{1}{X_i^2} \sigma^2 X_i^2 = \frac{N \sigma^2}{N^2} = \frac{\sigma^2}{N} \end{aligned}$$

Comparing the variances, we see that we're really comparing N to $(\sum_{i=1}^N X_i^2)^2 / \sum_{i=1}^N X_i^4$. If N is always greater than that ratio, then the WLS estimator for β_1 has smaller variance. By the Cauchy Swartz Inequality

we know that this ratio will always be less than one, while N must always be at least equal to one. Therefore, the WLS estimate of β_1 has smaller variance than the OLS estimate of β_1 .

Problem 3

Derive the inverse of matrix V in the Example 4.5.1.

In Example 4.5.1 we are given

$$\mathbf{V} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{bmatrix}$$

Let's look at some examples of the inverse using a variety of N s and values for $|\rho| < 1$ using the following function:

```
inv_V.fn <- function(N,rho){
  V <- diag(N)
  rowii <- NULL
  for(jj in c(-N:-1,2:N)) rowii <- c(rowii,rho^(abs(jj)-1))
  for(ii in 1:N) V[ii,] <- rowii[(N-(ii-1)):(length(rowii)-(ii-1))]

  return(as.fractions(solve(V)))
}
```

$N=5$, $\rho = 0.1$:

```
inv_V.fn(5,0.1)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 100/99 -10/99      0      0      0
## [2,] -10/99 101/99 -10/99      0      0
## [3,]      0 -10/99 101/99 -10/99      0
## [4,]      0      0 -10/99 101/99 -10/99
## [5,]      0      0      0 -10/99 100/99
```

$N=8$, $\rho = 0.1$:

```
inv_V.fn(8,0.1)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,] 100/99 -10/99      0      0      0      0      0      0
## [2,] -10/99 101/99 -10/99      0      0      0      0      0
## [3,]      0 -10/99 101/99 -10/99      0      0      0      0
## [4,]      0      0 -10/99 101/99 -10/99      0      0      0
## [5,]      0      0      0 -10/99 101/99 -10/99      0      0
## [6,]      0      0      0      0 -10/99 101/99 -10/99      0
## [7,]      0      0      0      0      0 -10/99 101/99 -10/99
## [8,]      0      0      0      0      0      0 -10/99 100/99
```

$N=5$, $\rho = 0.6$:

```
inv_V.fn(5,0.6)
```

```
##      [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 25/16 -15/16    0    0    0
## [2,] -15/16 17/8  -15/16    0    0
## [3,]    0 -15/16 17/8  -15/16    0
## [4,]    0    0 -15/16 17/8  -15/16
## [5,]    0    0    0 -15/16 25/16
```

N=5, $\rho = 0.9$:

```
inv_V.fn(5,0.9)
```

```
##      [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 100/19 -90/19    0    0    0
## [2,] -90/19 181/19 -90/19    0    0
## [3,]    0 -90/19 181/19 -90/19    0
## [4,]    0    0 -90/19 181/19 -90/19
## [5,]    0    0    0 -90/19 100/19
```

The numerator for nonzero elements is always equal to $100 \times (1 - \rho^2)$ (multiplied by 100 because of the fractions function I'm using): $1 - (0.1)^2 = 0.99$, $1 - (0.6)^2 = 0.64$, and $1 - (0.9)^2 = 0.19$. The size of the matrix doesn't effect the values of the elements. In the $[1, 1]$ and $[N, N]$ element numerators, we consistently have a 1 – the numerator of these elements isn't affected by ρ . The remaining diagonal elements are the same (depending on ρ). Looking at the cases where $\rho = 0.1$ and $\rho = 0.9$, it is easy to verify that numerator of diagonal elements (not in the corners) is $1 + 0.1^2$ and $1 + 0.9^2$, specifically $1 + \rho^2$. This can be verified by the case where $\rho = 0.6$ as well after scaling. Finally, the first-off-diagonal element numerators are all the same and simply equal to $-\rho^2$ (in the example where $\rho = 0.6$, we see this after scaling the fraction by 4). Thus, \mathbf{V}^{-1} is

$$\mathbf{V}^{-1} = \begin{bmatrix} \frac{1}{1-\rho^2} & \frac{-\rho}{1-\rho^2} & 0 & \dots & 0 \\ \frac{-\rho}{1-\rho^2} & \frac{1+\rho^2}{1-\rho^2} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \frac{1+\rho^2}{1-\rho^2} & \frac{-\rho}{1-\rho^2} \\ 0 & \dots & 0 & \frac{-\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix}$$
