Problem Set 2

Nima Mohammadi

February 27, 2020

Question 1: Orthogonality

Consider a CLRM with sample size n and k regressors. Show that if all columns in the data matrix $\mathbf{X}$ are orthogonal to one another, the resulting solution to the Least Square problem is equivalent to running k separate regressions, one per explanatory variable.

(Hint: Partition $\mathbf{X}$ into columns and express the solution formula for $\mathbf{b}$ in terms of column interactions. Then examine what this solution looks like under the stated orthogonality assumption. Note: This is NOT asking you to do a partitioned regression. Just a basic regression, with $\mathbf{X}$ expressed in terms of its columns.)

$$\mathbf{X} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_k\end{bmatrix}$$

$$\mathbf{X}' = \begin{bmatrix}\mathbf{c}_1' \\ \mathbf{c}_2' \\ \vdots \\ \mathbf{c}_k'\end{bmatrix}$$

Due to orthogonality columns of$\mathbf{X}$, we have$\mathbf{c}_i'\mathbf{c}_j = 0fori \neq j$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix}\mathbf{c}_1'\mathbf{c}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{c}_2'\mathbf{c}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mathbf{c}_k'\mathbf{c}_k\end{bmatrix} \tag{1}$$

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \begin{bmatrix}\mathbf{c}_1'\mathbf{c}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{c}_2'\mathbf{c}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \mathbf{c}_k'\mathbf{c}_k\end{bmatrix}^{-1} = \begin{bmatrix}(\mathbf{c}_1'\mathbf{c}_1)^{-1} & 0 & \cdots & 0 \\ 0 & (\mathbf{c}_2'\mathbf{c}_2)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & (\mathbf{c}_k'\mathbf{c}_k)^{-1}\end{bmatrix}$$

Also

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{c}_1' \\ \mathbf{c}_2' \\ \vdots \\ \mathbf{c}_k' \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{c}_1'\mathbf{y} \\ \mathbf{c}_2'\mathbf{y} \\ \vdots \\ \mathbf{c}_k'\mathbf{y} \end{bmatrix}$$

Then

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} (\mathbf{c}_1'\mathbf{c}_1)^{-1} & 0 & \cdots & 0 \\ 0 & (\mathbf{c}_2'\mathbf{c}_2)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & (\mathbf{c}_k'\mathbf{c}_k)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1'\mathbf{y} \\ \mathbf{c}_2'\mathbf{y} \\ \vdots \\ \mathbf{c}_k'\mathbf{y} \end{bmatrix} = \begin{bmatrix} (\mathbf{c}_1'\mathbf{c}_1)^{-1}\mathbf{c}_1'\mathbf{y} \\ (\mathbf{c}_2'\mathbf{c}_2)^{-1}\mathbf{c}_2'\mathbf{y} \\ \vdots \\ (\mathbf{c}_k'\mathbf{c}_k)^{-1}\mathbf{c}_k'\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_k \end{bmatrix} \tag{2}$$

In this (rather rare) occasion, the coefficients of the regression of $\mathbf{y}$ on $\mathbf{X}$ may be obtained via separate regressions of $\mathbf{y}$ on $\mathbf{c}_i$. In other words, the solution of the Least Squares problem would be equivalent to k separate regressions, one for each explonatory variable.

Question 2: Regression on a constant

Consider a classical linear regression model that regresses the dependent variable, $\mathbf{y}$, against a column of ones (call it $\mathbf{i}$), and no other explanatory variables. Assume the underlying theoretical population model has the usual OLS properties.

(a) Write down the underlying theoretical model for a single observation, and for the full sample of n observations. Clearly specify the vector dimensions of each element. Call the unknown population parameter μ).

For single observation:

$$y_j = i_j\mu + \epsilon_j \tag{3}$$

For full sample of n observations:

$$\mathbf{y}_{n\times 1} = \mathbf{i}_{n\times 1}\mu + \epsilon_{n\times 1} \tag{4}$$

- μ: unknown population parameter to be estimated (as $\mathbf{b}$), scalar value
- $\mathbf{y}$: dependent variable, $\begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}'$
- $\mathbf{i}$: independent "variable", column vector of ones
- ϵ: disturbance term

(b) Derive the OLS estimator for this model (Call it b). (Hint: Start with the known formula of the OLS estimator and replace the $\mathbf{X}$ with the regressor for the current model. Then simplify until you're left with a scalar that has a very well known form).

$$\begin{aligned} \mathbf{b} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\ \Rightarrow \mathbf{b} &= \left(\mathbf{i}'\mathbf{i}\right)^{-1}\mathbf{i}'\mathbf{y} = n^{-1}\sum_{j=1}^{n} y_j = \bar{y} \end{aligned} \tag{5}$$

The OLS estimator for this model finds the sample mean of $\mathbf{y}$ as the estimation of μ.

(c) Derive the variance of this estimator. Assume that the error variance σ is known. (Hint: Proceed as above - start with the known form of V(b), then replace the $\mathbf{X}$ with the regressor for the current model).

$$\mathbf{V}(\mathbf{b}) = \sigma^2 \mathbb{E}(\left(\mathbf{i}'\mathbf{i}\right)^{-1}) = \frac{\sigma^2}{n} \tag{6}$$

(d) Show that the OLS estimator for your model is unbiased for the underlying population parameter.

$$\begin{aligned}\mathbb{E}(\mathbf{b}) &= \mathbb{E}(\bar{y}) = \mathbb{E}(\frac{1}{n}\sum_{j=1}^{n} y_j) = \frac{1}{n}\mathbb{E}(\sum_{j=1}^{n} y_j) \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(y_j) = \frac{1}{n}\sum_{j=1}^{n}\mu = \frac{1}{n}n\mu = \mu\end{aligned} \tag{7}$$

(e) In light of your results, can you make a general statement regarding the best linear unbiased estimator for the population mean when no explanatory variables are available?

In part (b) we saw that in absence of explanatory variables, the OLS estimator is equal to the sample mean $\bar{y}$. The Gauss-Markov theorem shows that this estimator of population mean is 'BLUE'. Part (d) showed us the unbiasedness of this estimator whereas in part (d) the consistency of the model is shown. Comparing the variance with thereof other estimators indicates the relative efficiency of the OLS estimator. In conclusion, the OLS estimator of poulation mean gives the lowest variance of the estimate, as compared to other unbiased, linear estimators.

Question 3: Small Sample Properties of a Linear regression Model with non-zero-mean-error

Consider the TRUE linear regression model $\mathbf{y} = \mathbf{i}\beta_1 + \mathbf{X}\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ where $\mathbf{i}$ is an n by 1 column of 1's, and $\mathbf{X}$ is an n by (k-1) matrix of regressors. Assume that the n elements of the random error vector $\boldsymbol{\epsilon}$ are distributed i.i.d normal with variance σ^2, but with mean $\mu \neq 0$. As usual, assume that the elements in $\boldsymbol{\epsilon}$ are uncorrelated with the elements in $\mathbf{X}$.

1. Because the mean of ϵ is not zero, hence assumptions 3 and assumption 5 are violated by this specification.

2. We want to show that via Partitioned Regression, the estimator of the coefficient parameters for this model will produce unbiased slope estimation, but that the estimation for the intercept will be biased. We will indicate the estimated intercept as b_1 and the vector of remaining estimated parameters as b_2 with true parameter values β_1 and β_2, respectively. The functional

form for the partitioned regression will be $y = X_1\beta_1 + X_2\beta_2$, where X_1 is a vector of ones. Then the estimator of β_2 will be $b_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y$ where $\mathbf{M}_1 = \mathbf{I} - i(i'i)^{-1}i'$ represents the deviation-from-the-mean matrix.

$$
\begin{aligned}
&\mathrm{b}_2 = \left(X_2' M_1 X_2'\right)^{-1} X_2' M_1 \left(X_1\beta_1 + X_2\beta_2 + \epsilon\right) \\
&\quad = \left(X_2' M_1 X_2\right)^{-1} X_2' M_1 X_1 \beta_1 + \left(X_2' M_1 X_2\right)^{-1} X_2' M_1 X_2 \beta_2 + \left(X_2' M_1 X_2\right)^{-1} X_2' M_1 \epsilon \\
&M_2 = I - X_2\left(X_2' X_2\right)^{-1} X_2' \rightarrow M_2 X_2 = 0 \\
&\text{and} \\
&M_1 = I - X_1\left(X_1' X_1\right)^{-1} X_1' \rightarrow M_1 X_1 = 0 \\
&\quad \mathbf{b}_2 = \left(X_2' M_1 X_2\right)^{-1} X_2' M_1 X_2 \beta_2 + \left(X_2' M_1 X_2\right)^{-1} X_2' M_1 \epsilon \\
&\quad \text{Also we have } M_1\epsilon = \epsilon - i\bar{\epsilon} = \epsilon - i\mu \longrightarrow E_\epsilon\left(\mathbf{b}_2 | X\right) = \beta_2 + \left(\left(X_2' M_1 X_2\right)^{-1} X_2' E_\epsilon(\epsilon - i\mu) | X\right) \\
&\quad E_\epsilon((\epsilon - i\mu)|X) = i\mu - i\mu = 0 \longrightarrow E\left(b_2\right) = \beta_2 \longrightarrow \text{Then } b_2 \text{ is an unbiased estimator for the real} \beta_2.
\end{aligned} \tag{8}
$$

For b_1 we have $b_1 = \left(X_A' M_2 X_1\right)^{-1} X_1' M_2 y = \left(X_1' M_2 X_1\right)^{-1} X_1' M_2 \left(X_1\beta_1 + X_2\beta_2 + \epsilon\right)$

Since we have $M_2 X_2 = 0$:

$$
E\left(b_1\right) = \beta_1 + E\left[\left(X_1' M_2 X_1\right)^{-1} X_1' M_2 \epsilon\right] = \beta_1 + \left(X_1' M_2 X_1\right)^{-1} X_1' E\left(M_2 \epsilon\right) = \beta_1 + \mu
$$

Since the expected value of $M_2\epsilon$ is not zero in this case, expected value of b_1 is not equal to β_1, so it is a biased estimator.

3. Now we assume the true model is as before, but the analyst chooses to estimate it without an intercept. Continuing to assume the above given properties of the error term, we want to show that in this case the OLS estimator $\mathbf{b}_2$ will be biased.

$$
\begin{aligned}
&b = \left(X'X\right)^{-1} X'y = \left(X'X\right)^{-1} X'(X\beta + \epsilon) = \beta + \left(X'X\right)^{-1} X'\epsilon \\
&\Rightarrow E_\epsilon(b|X) = E_\epsilon\left(\left(\beta + \left(X'X\right)^{-1} X'\epsilon\right) | X\right) \\
&= \beta + \left(X'X\right)^{-1} X' E_\epsilon(\epsilon|X) = \beta + \left(X'X\right)^{-1} X'\mu \\
&\Rightarrow E(b) = E_X\left(E_\epsilon \langle b|X)\right) = E_X(\beta) + E_X\left(\left(X'X\right)^{-1} X'\mu\right) = \beta + (X'X)^{-1}X'\mu \neq \beta
\end{aligned} \tag{9}
$$

So the estimator is biased.

4. The reason for biasedness of b in the previous part is that we do not have an intercept in our model preventing us from capturing the effects of a non-zero mean error term in the intercept. But, in part 2 we could split the model and the effects of the non-zero mean would be capture by b_1 and therefore b_2 is an unbiased estimator, but now in this setting we do not have intercept and the whole b as the estimator of beta matrix is consequently biased.

Question 4: Variance of OLS estimator

Consider script `mod1s3`, which deals with the small sample properties of the OLS estimator.
In this exercise you will show that the empirical standard deviations ("standard errors") of the OLS estimator depend on the variability in $\mathbf{X}$. Specifically, perform the following simulation tasks (make sure all your tables and figures are labeled clearly and correctly):

```
R> set.seed(37)
R> N <- 1000 #population size
R> n <- 1000 #sample size
R> r1 <- 500 #number of error vectors to draw

R> bvec<-c(1, .5, 1.2)
R> sig <- 1
R> x1 <- rep(1, N)
R> x2 <- rnorm(N, 3, 1)
R> x3 <- rnorm(N, -2, 1)
R> X1 <- cbind(x1, x2, x3)
R> k <- ncol(X1)
R> x2 <- rnorm(N, 3, 3)
R> x3 <- rnorm(N, -2, 3)
R> X2 <- cbind(x1, x2, x3)
R> x2 <- rnorm(N, 3, 4)
R> x3 <- rnorm(N, -2, 4)
R> X3 <- cbind(x1, x2, x3)
R> x2 <- rnorm(N, 3, 6)
R> x3 <- rnorm(N, -2, 6)
R> X4 <- cbind(x1, x2, x3)

R> bmat1 <- matrix(0, k, r1)
R> bmat2 <- matrix(0, k, r1)
R> bmat3 <- matrix(0, k, r1)
R> bmat4 <- matrix(0, k, r1)
R> for (i in 1:r1){
         eps <- rnorm(n, 0, sig)
         y1 <- X1 %*% bvec + eps
         y2 <- X2 %*% bvec + eps
         y3 <- X3 %*% bvec + eps
         y4 <- X4 %*% bvec + eps

         b1 <- solve((t(X1)) %*% X1) %*% (t(X1) %*% y1)
         b2 <- solve((t(X2)) %*% X2) %*% (t(X2) %*% y2)
         b3 <- solve((t(X3)) %*% X3) %*% (t(X3) %*% y3)
         b4 <- solve((t(X4)) %*% X4) %*% (t(X4) %*% y4)

         bmat1[, i] <- b1
```

```
        bmat2[, i] <- b2
        bmat3[, i] <- b3
        bmat4[, i] <- b4
}
```

Table 1: Sampling distribution, std(x2)=std(x3)=1

variable	true value	mean (samp.dist.)	std (samp.dist)
constant	1.0000	0.9939	0.1138
x2	0.5000	0.5018	0.0316
x3	1.2000	1.2002	0.0306

Table 2: Sampling distribution, std(x2)=std(x3)=3

variable	true value	mean (samp.dist.)	std (samp.dist)
constant	1.0000	1.0001	0.0484
x2	0.5000	0.4999	0.0110
x3	1.2000	1.2005	0.0098

Table 3: Sampling distribution, std(x2)=std(x3)=4

variable	true value	mean (samp.dist.)	std (samp.dist)
constant	1.0000	1.0019	0.0412
x2	0.5000	0.4991	0.0081
x3	1.2000	1.2002	0.0076

Table 4: Sampling distribution, std(x2)=std(x3)=6

variable	true value	mean (samp.dist.)	std (samp.dist)
constant	1.0000	1.0019	0.0349
x2	0.5000	0.4991	0.0055
x3	1.2000	1.2002	0.0052

As implied by $Var(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, with increase of the variance of covariates $\mathbf{X}$ we will have decrease of variance for $\mathbf{b}$. Interestingly there is an inverse relation between the variance of the explonatory variables and the standard deviation of parameter estimations. This suggests that with higher explanatory power the distribution of of estimated values become narrower. Consequently, we will have more significant estimations and observe increase in t-value. The figures and tables above confirm our findings.

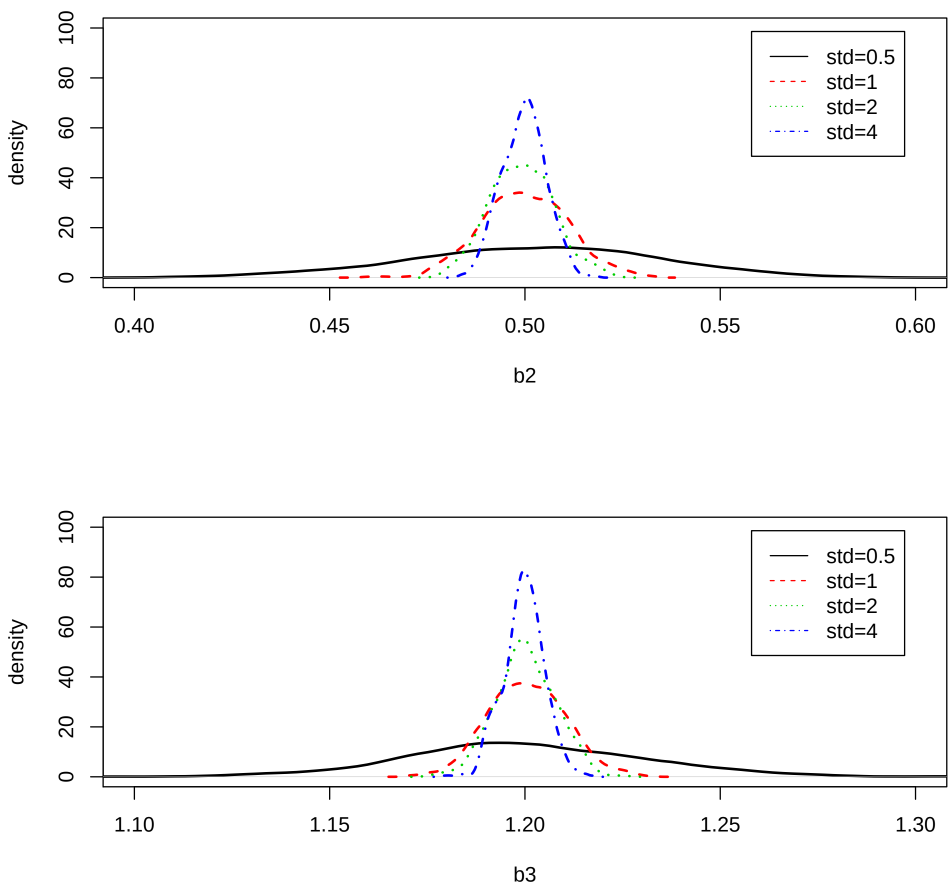

Figure 1: CONDITIONAL Sampling distribution for OLS estimates at different settings for the coefficients of x2 and x3