Econ 203B Econometrics

Raphael Lam

1 Problem Set 1 Solutions[1]

Greene Chapter 3 Question 1:

For the regression model:

$$y = \alpha + \beta X + \varepsilon,$$

1. Show that the least squares normal equations imply $\Sigma_i \hat{\varepsilon}_i = 0$ and $\Sigma_i x_i \hat{\varepsilon}_i = 0$.

   The least-squares minimization problem:

$$\begin{aligned}
\min_{\alpha,\widetilde{\beta}} S(\alpha,\widetilde{\beta}) &= \min_{\alpha,\widetilde{\beta}} \hat{\varepsilon}'\hat{\varepsilon} \\
&= \min_{\alpha,\widetilde{\beta}} \left(y - \alpha i - \widetilde{\beta} X\right)' \left(y - \alpha i - \widetilde{\beta} X\right) \\
&= \min_{\alpha,\widetilde{\beta}} \left(y'y + \alpha^2 i'i + \widetilde{\beta}^2 X'X - 2y'ai - 2y'\widetilde{\beta}X + 2\widetilde{\beta}X'\alpha i\right) \qquad (1)
\end{aligned}$$

   Take the first order derivatives w.r.t. α, $\tilde{\beta}$:

$$\begin{aligned}
\alpha : \frac{\partial S(\widehat{\alpha},\widehat{\beta})}{\partial \alpha} &= y'i - \hat{\alpha} i'i - \hat{\beta} X'i = 0 \\
&= i'\left(y - \hat{\alpha} i + \hat{\beta} X\right) = i'\hat{\varepsilon} = \Sigma_i \hat{\varepsilon}_i \\
&\therefore \Sigma_i \hat{\varepsilon}_i = 0 \qquad (2)
\end{aligned}$$

$$\begin{aligned}
\beta : \frac{\partial S(\widehat{\alpha},\widehat{\beta})}{\partial \beta} &= X'y - X'\hat{\alpha} i - X'X\hat{\beta} \\
&= X'\left(y - \hat{\alpha} i - X\hat{\beta}\right) = \Sigma_i X_i \hat{\varepsilon}_i \\
\therefore \Sigma_i X_i \hat{\varepsilon}_i &= 0
\end{aligned}$$

---

[1] I have edited and added comments on the source files from Juan Patano, which built on top of Myer Bretts and CM Yuan's version. All errors are mine. The question on MSE of an estimator is important and you are expected to know (it shows up everywhere), although I was told that Professor didn't mention it in class.

2. Show that the solution for the constant term is $\hat{\alpha} = \bar{y} - \hat{\beta}\overline{X}$.

   From the normal equations:

$$\begin{aligned} y'i - \hat{\alpha}i'i - \hat{\beta}X'i &= 0 \\ \hat{\alpha}i'i &= y'i - \hat{\beta}X'i \\ \hat{\alpha}n &= \Sigma_{i=1}^{n} y_i - \hat{\beta}\Sigma_{i=1}^{n} X_i \\ \hat{\alpha} &= \frac{1}{n}\Sigma_{i=1}^{n} y_i - \hat{\beta}\frac{1}{n}\Sigma_{i=1}^{n} X_i \\ &= \bar{y} - \hat{\beta}\overline{X} \end{aligned}$$

3. Show that the solution for $\hat{\beta}$ is $\hat{\beta} = \left[\Sigma_i \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right] / \left[\Sigma_i^2 \left(x_i - \bar{x}\right)\right]$.

   Again, from the normal equations:

$$\begin{aligned} X'y - X'\hat{\alpha}i - X'X\hat{\beta} &= X'y - X'\left(\bar{y} - \hat{\beta}\overline{X}\right)i - X'X\hat{\beta} \\ \hat{\beta}X'\left(X - Xi\right) &= X'\left(y - \bar{y}i\right) \\ \hat{\beta}\left(X - \overline{X}i\right)'\left(X - \overline{X}i\right) &= \left(X - \overline{X}i\right)'\left(y - \bar{y}i\right) \\ \hat{\beta} &= \frac{\Sigma_{i=1}^{n}\left(X_i - \overline{X}\right)\left(y_i - \bar{y}\right)}{\Sigma_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \end{aligned} \tag{3}$$

4. Prove that this two values uniquely minimize the sum of squares by showing that the diagonal elements of the second derivatives matrix of the sum of squares with respect to the parameters are both positive and that the determinant is $4n\left[\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right] = 4n\left[\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right]$,which is positive unless all values of x are the same.

To check that we are in fact taking a minimum value, we take the derivatives on the FOCs above:

$$\text{Noramal Equation:} - 2X'y + 2X'X\widehat{\beta} = 0 \tag{4}$$

(Check **Simon and Blume** text for the necessary and sufficient conditions for an optimization problem!)

$$\frac{\partial SSR}{\partial\widehat{\beta}\partial\widehat{\beta}'} = 2X'X = \begin{bmatrix} 2n & 2\Sigma_{i=1}^{n}X_i \\ 2\Sigma_{i=1}^{n}X_i & 2\Sigma_{i=1}^{n}X_i^2 \end{bmatrix}$$

$$\begin{aligned} \left|2X'X\right| &= 4n\Sigma_{i=1}^{n}X_i^2 - 4\left(\Sigma_{i=1}^{n}X_i\right)^2 \\ &= 4n\left(\Sigma_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right) \geq 0 \end{aligned} \tag{5}$$

Note that second order derivatives is positive semi-definite. Further to note that the determinant is equal to zero iff all observations are the same, i.e. $x_1 = ... = x_n$, which is ruled out if we have a full-rank assumption at the beginning.

Greene Chapter 3 Question 3:

Consider the least-squares regression of y on K variables (with a constant) X. Consider an alternative set of regressors $Z = XP$ where P is a nonsingular matrix. Thus, each column of Z is a mixture of some of the columns of X. Prove that the residual vectors $\hat{\varepsilon}_X$ and $\hat{\varepsilon}_Z$ in the regressions of y on X and y on Z respectively are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variables?

Express the disturbance $\hat{\varepsilon}_X = (I - X\left(X'X\right)^{-1}X')y$ and $\hat{\varepsilon}_Z = (I - Z\left(Z'Z\right)^{-1}Z')y$ for the two regressions respectively. Rearranging with matrix algebra:

$$\begin{aligned}
\hat{\varepsilon}_Z &= (I - Z\left(Z'Z\right)^{-1}Z')y \\
&= (I - XP\left((XP)'XP\right)^{-1}(XP)')y \\
&= (I - XP\left(P'X'XP\right)^{-1}P'X')y \\
&= (I - XPP^{-1}\left('X'X\right)^{-1}P'^{-1}P'X')y \\
&= (I - X\left(X'X\right)^{-1}X')y \\
\therefore \hat{\varepsilon}_X &= \hat{\varepsilon}_Z \qquad (6)
\end{aligned}$$

This question is similar to one of the questions in the past exam that we discussed in class.

Greene Chapter 3 Question 6:

A data set contains n observations on X_n and y_n. The least squares estimator based on these n observations is $\hat{\beta}_n = \left(X_n'X_n\right)^{-1}X_n'y_n$. Another observation, x_s and y_s, becomes available. Prove that the least squares estimator computed using this additional observation is

$$\hat{\beta}_{n,s} = \hat{\beta}_n + \frac{1}{1 + x_s\left(X_n'X_n\right)^{-1}x_s'}\left(X_n'X_n\right)^{-1}x_s'\left(y_s - x_s\hat{\beta}_n\right)$$

***It is preferable to work with x_s as a row vector of dimension $K \times 1$ and that is why the formula above looks slightly different from Greens excercise.

The last term is $\hat{\varepsilon}_s$, the residual from the prediction of y_s using the coefficients based on X_n and $\hat{\beta}_n$. Since the new observation is likely to change the estimation of the OLS estimator, we have to adjust the initial estimator $\hat{\beta}_n$ with a second term listed on above equation. If the new observational variables on the right hand side has perfectly predicted the new observation using the initial $\hat{\beta}_n$, then we can see that $\hat{\varepsilon}_s$ is zero and therefore we don't need any adjustment.

It is convenient to stack the matrix into two subgroups: the old observations from $(1...n)$ and the new observations from $(n+1...n+s)$. So, the dimensions of the vectors will be n and s respectively. Write it as the following:

$$\begin{bmatrix} y_n \\ y_s \end{bmatrix} = \begin{bmatrix} X_n\beta_{n,s} \\ x_s\beta_{n,s} \end{bmatrix} + \begin{bmatrix} \varepsilon_n \\ \varepsilon_s \end{bmatrix} \tag{7}$$

Do the standard minimization problem (Expanding the matrix and take first-order derivatives)

$$\begin{aligned} \min_{\beta} S(\beta_{n,s}) &= \min_{\beta} \hat{\varepsilon}'\hat{\varepsilon} = \min_{\beta} \left(\begin{bmatrix} y_n \\ y_s \end{bmatrix} - \begin{bmatrix} X_n\beta_{n,s} \\ x_s\beta_{n,s} \end{bmatrix} \right)' \left(\begin{bmatrix} y_n \\ y_s \end{bmatrix} - \begin{bmatrix} X_n\beta_{n,s} \\ x_s\beta_{n,s} \end{bmatrix} \right) \\ &= y_n'y_n + y_s^2 - 2\left(X_n\beta_{n,s}\right)' y_n - 2\left(x_s\beta_{n,s}\right)' y_s + \beta_{n,s}'X_n'X_n\beta_{n,s} + \beta_{n,s}'x_s'x_s\beta_{n,s} \end{aligned}$$

FOC:

$$\begin{aligned} \frac{\partial S\left(\beta_{n,s}\right)}{\partial \beta_{n,s}} &= -X_n'y_n - x_s'y_s + X_n'X_n\hat{\beta}_{n,s} + x_s'x_s\hat{\beta}_{n,s} \\ \left(X_n'X_n\right)^{-1}\left(X_n'X_n\hat{\beta}_{n,s} + x_s'x_s\hat{\beta}_{n,s}\right) &= \left(X_n'X_n\right)^{-1}\left(X_n'y_n + x_s'y_s\right) \\ &\text{[Premultiply with } \left(X_n'X_n\right)^{-1}\text{]} \\ \hat{\beta}_{n,s} + \left(X_n'X_n\right)^{-1}x_s'x_s\hat{\beta}_{n,s} &= \hat{\beta}_n + \left(X_n'X_n\right)^{-1}x_s'y_s \text{[Simplify and noting } \hat{\beta}_n = \left(X_n'X_n\right)^{-1} \\ \hat{\beta}_{n,s} &= \hat{\beta}_n + \left(X_n'X_n\right)^{-1}x_s'y_s \text{ - } + \left(X_n'X_n\right)^{-1}x_s'x_s\hat{\beta}_{n,s} \\ &\text{[We're close and premultiply with } x_s\text{]} \\ x_s\hat{\beta}_{n,s} - x_s\hat{\beta}_n + x_s\left(X_n'X_n\right)^{-1}x_s'x_s\hat{\beta}_{n,s} &= x_s\left(X_n'X_n\right)^{-1}x_s'y_s \\ &\text{[To get the term } 1 + \ldots \text{ in above eq., try } -x_s\left(X_n'X_n\right)^{-1} \\ \left(1 + x_s\left(X_n'X_n\right)^{-1}x_s'\right)\left(x_s\hat{\beta}_{n,s} - x_s\hat{\beta}_n\right) &= x_s\left(X_n'X_n\right)^{-1}x_s'\left(y_s - x_s\hat{\beta}_n\right) \\ &\text{[Drop off the } x_s\text{]} \\ \hat{\beta}_{n,s} - \hat{\beta}_n &= \frac{1}{1 + x_s\left(X_n'X_n\right)^{-1}x_s'}\left(X_n'X_n\right)^{-1}x_s'\left(y_s - x_s\hat{\beta}_n\right) \quad \text{[Done!]} \end{aligned}$$

Note: This is a tedious algebra but you should be able to do it once in a long while.

Intereptation: Having an additional observation doesn't necessarily change the $\hat{\beta}$ estimation, if the second term is equal to zero. How? The very last term is the residual error from the new observation. So, if the residuals are equal to zero, the two $\hat{\beta}$ are the same. If the residuals are not equal to zero, it would change the new estimator $\hat{\beta}_{n,s}$ with a term related to two things: 1) the estimated error of the $\hat{\beta}_n$. 2) the prediction error from using the initial estimator $\hat{\beta}_n$ to infer the new observation. The denominator term looks like the prediction error you learn in the lecture notes.

Greene Chapter 3 Question 7:

A common strategy for handling a case in which an observation is missing data for one or more variables is to fill those missing variables with 0s and add a variable to the model that takes the value 1 for that one observation and 0 for all other observations. Show that this "strategy" is equivalent to discarding the observation as regards the computation of $\hat{\beta}$ but it *does* have an effect on R^2. Consider the special case in which X contains only a constant and one variable. Show that replacing missing values of x with the mean of the complete observations has the same effect as adding the new variable.Note that we only need to focus on the case where only *one* observation has this missing data problem

The question suggests that in face of missing variables omitted from the regression we have (for X with n rows and k columns):

$$\begin{aligned}(X'X)^{-1}X'y &= \hat{\beta} \qquad (k \times 1 \text{ vector})\\ &= Ay \\ &= \begin{bmatrix} A_{11} \cdots A_{1n} \\ \vdots \ \ddots \ \vdots \\ A_{k1} \cdots A_{kn}\end{bmatrix} y\end{aligned}$$

where A is $k \times n$.When X is augmented with a new column and the new row is added as described above, we have:

$$X_0 = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \cdots & x_{nk} & 0 \\ 1 & 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad y_0 = \begin{bmatrix} y_1 \\ \vdots \\ y_n \\ y_{n+1} \end{bmatrix}$$

X_0 is $(n+1) \times (k+1)$ matrix and y_0 is $(n+1) \times 1$

The first k elements of $\hat{\beta}_0$ are the same as in $\hat{\beta}$, while the last $(k+1)^{th}$ element of $\hat{\beta}_0$ is equal to difference between the "new" observation in y

and the intercept. However, we are not very interested in this new $\hat{\beta}_{k+1}$. It is not our interest in the estimation. This must be the case for the condition $\Sigma_i \hat{\varepsilon}_i = 0$ to hold. Applying the formula for the inverse of partitioned matrices, after some tedious algebra we get (Refer to the matrix algebra note or convince yourself with Matlab)

$$\begin{aligned} \left(X_0' X_0\right)^{-1} X_0' y_0 &= \hat{\beta}_0 \qquad ((k+1) \times 1 \text{ vector}) \\ &= B y_0 \\ &= \begin{bmatrix} A_{11} \cdots A_{1n} \, 0 \\ \vdots \quad \ddots \quad \vdots \quad 0 \\ A_{k1} \cdots A_{kn} \, 0 \\ B_n \ \cdots \ B_n \ \ 1 \end{bmatrix} y_0 \end{aligned}$$

The goodness of fit is improved by having an additional observation y_{n+1} that exactly the same as the predicted value $\hat{y}_{n+1}$. R^2 increases as $\hat{y}_{n+1} = y$ with $\hat{\varepsilon}_{n+1} = 0$.

The second case:

Considering the special case where X contains only a constant and one variable, and where the missing values of x are replaced by the mean $(\bar{x})$ of the complete observations of x. In case with missing values of x are omitted:

$$\begin{aligned} \hat{\beta} &= \left[\Sigma_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right] / \left[\Sigma_{i=1}^{n} \left(x_i - \bar{x}\right)^2\right] \\ \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \end{aligned}$$

By observation we can see that when any number of missing x values are added to the sample as $\bar{x}$, the estimate for $\hat{\beta}$ remains *unchanged* as the quantities $\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$ and $\left(x_i - \bar{x}\right)^2$ in the numerator and the denominator, respectively.

Greene Chapter 4 Question 1:

Suppose that you have two independent unbiased estimators of the same parameter θ, say $\hat{\theta}_1$ and $\hat{\theta}_2$, with different variances υ_1 and υ_2. What linear combination $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ is the minimum variance unbiased estimator of θ?

Minimum variance unbiased estimator of θ :
2 Features: 1) minimum variance; 2) Unbiasedness $E\left[\hat{\theta}\right] = \theta$

Working first on the second feature:

$$E\left[\hat{\theta}\right] = E\left[c_1\hat{\theta}_1 + c_2\hat{\theta}_2\right] = c_1 E\left[\hat{\theta}_1\right] + c_2 E\left[\hat{\theta}_2\right]$$

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are BOTH unbiased, it follows that $c_1 + c_2 = 1$

We need further restriction on c_1 and c_2 to achieve the first feature of min. variance (The trick is to choose those parameters to satisfy that feature)

Rewriting $c_1 = c$ and take the first order derivative w.r.t. c :

$$\min_c var\left(c\hat{\theta}_1 + (1-c)\hat{\theta}_2\right)$$
$$\min_c c^2 v_1 + (1-c)^2 v_2$$

This is because

$$Var(\hat{\theta}_1 + \hat{\theta}_2) = Var(\hat{\theta}_1) + Var(\hat{\theta}_2) + 2Cov(\hat{\theta}_1, \hat{\theta}_2)$$

The last term on covariance is equal to zero.

FOC:

$$\begin{aligned} 2cv_1 - 2\left(1-c\right)v_2 &= 0 \\ c\left(v_1 + v_2\right) &= v_2 \\ c &= \frac{v_2}{v_1 + v_2} \end{aligned}$$
$$c_1 = \frac{v_2}{v_1 + v_2}; \qquad c_2 = \frac{v_1}{v_1 + v_2}$$

Greene Chapter 4 Question 2

Consider the simple regression $y_i = \beta x_i + \varepsilon_i$ where $E\left[\varepsilon|x\right] = 0$ and $E\left[\varepsilon^2|x\right] = \sigma^2$.

1. a) What is the minimum mean squared error linear estimator of β? (Hint: Let the estimator be $\hat{\beta} = c'y$ and choose c to minimize $var\left(\hat{\beta}\right) + E\left[\hat{\beta} - \beta\right]^2$. The answer is a function of the unknown parameters.)

   Remark 1 *Again we break down into two features: 1) Minimum MSE. 2) Linear estimator of* β*. Here we only consider a class of estimators that satisfies those two features. There may be another **nonlinear** estimator of* β *that achieves a lower minimum MSE. Keep in mind that the OLS estimator* $\hat{\beta}$ *is only one of the many types of estimators. This question explores whether OLS estimator* $\hat{\beta}$ *satisfies these two features.*

As we discussed in the section, the mean squared error (MSE) of an estimator $\tilde{\beta}$ is equal to the squared derivation from the **true parameter value** β (not its own expected value $E(\tilde{\beta})$). One important thing to note is that $E(\tilde{\beta})$ may NOT equal to β,i.e. $\tilde{\beta}$ is unbiased.

$$\begin{aligned}
MSE(\tilde{\beta}) &= E(\tilde{\beta}-\beta)^2 \\
&= E\left[\tilde{\beta}-E(\tilde{\beta})+E(\tilde{\beta})-\beta\right]^2 \\
&= E[\tilde{\beta}-E(\tilde{\beta})]^2 + E[E(\tilde{\beta})-\beta]^2 + 2E\left\{[\tilde{\beta}-E(\tilde{\beta})][E(\tilde{\beta})-\beta]\right\}
\end{aligned}$$

Interpretation: We come up with three terms in the above equation. The first term is the variance of the estimator $\tilde{\beta} : Var(\tilde{\beta})$; the second term we call it a Bias. Note that this term is positive. Applying law of iterated expectation we arrive at the bias is equal to $E(\tilde{\beta})-\beta$ which is a constant. Try to convince yourself at your leisure that the third term is equal to zero (by applying the law of iterated expectation). Here we go:

$$\begin{aligned}
2E\left\{[\tilde{\beta}-E(\tilde{\beta})][E(\tilde{\beta})-\beta]|x\right\} &= 2E\left\{\tilde{\beta}E(\tilde{\beta})-\tilde{\beta}\beta-E(\tilde{\beta})E(\tilde{\beta})+E(\tilde{\beta})\beta\right\} \\
&= 2\{E(\tilde{\beta})E(\tilde{\beta})-E(\tilde{\beta})\beta-E(\tilde{\beta})E(\tilde{\beta})+E(\tilde{\beta})\beta\} \\
&\quad \text{(Applying Law of iterated Expectation)} \\
&= 0
\end{aligned}$$

Note that the true parameter β is a constant and the above expectation is conditional on x.

$$\begin{aligned}
MSE(\tilde{\beta}) &= Var(\tilde{\beta}) + Bias \\
&= Var(\tilde{\beta}) + E[E(\tilde{\beta})-\beta]^2 \quad \text{(same equation as in the question)}
\end{aligned}$$

Here we consider the estimator $\tilde{\beta}$ to be linear (2nd feature):

$$\begin{aligned}
MSE(\tilde{\beta}) &= var\,(c'y) + E\,[c'y-\beta|x]^2 \\
&= var\,(c'\beta x + c'\varepsilon) + E\,[c'\beta x + c'\varepsilon - \beta|x]^2 \\
&= var\,(c'\varepsilon) + E\,[c'\beta x + c'\varepsilon - \beta|x]^2 \\
&= \sigma^2 c'c + \beta^2 c'xx'c - 2\beta^2 c'x + \beta^2
\end{aligned}$$

Remember that we want that $\tilde{\beta}$ satisfies the first feature of minimum MSE, so we are choosing c such that $MSE(\tilde{\beta})$ achieves the minimum.

$$\begin{aligned}
\frac{\partial MSE(\hat{\beta})}{\partial c} &= \beta^2 c' x x' + \sigma^2 c' - \beta^2 x' = 0 \\
\beta^2 x' &= \beta^2 c' x x' + \sigma^2 c' \\
&= c' \left(\beta^2 x x' + I\sigma^2\right) \\
c' &= \beta^2 x' \left(\beta^2 x x' + I\sigma^2\right)^{-1} \\
c &= \frac{\beta^2 x}{\left(\beta^2 x' x + \sigma^2\right)} \\
\hat{\beta} &= c' y = \frac{\beta^2 x}{\left(\beta^2 x' x + \sigma^2\right)} y
\end{aligned}$$

Remark 2 *An important thing to notice is that the mean squared error of an estimator consists of two terms: variance of estimator and a bias. There exists a tradeoff between the two. You can construct a unbiased estimator but usually at a tradeoff of having a higher variance of that estimator (an example will be an OLS estimator, which is unbiased). Here when we consider both at the same time to arrive at an estimator with minimum MSE. There may exist a third estimator that achieves a lower variance but at a cost of increasing the bias. We will see more of it over the entire course like the maximum likelihood estimator* (MLE) *and their properties.*

b) For the estimator in part a), show that the ratio of the mean squared error of $\hat{\beta}$ to that of the ordinary least squares estimator b (with $MSE(b) = \sigma^2 / x'x = \beta^2/\tau^2$) is:

$$\frac{MSE\left(\hat{\beta}\right)}{MSE(b)} = \frac{\tau^2}{1+\tau^2}, \quad \text{where } \tau^2 = \frac{\beta^2}{\sigma^2 / x'x}$$

How do you interpret the behavior of this ratio as $\tau \longrightarrow \infty$?

1. Substituting the calculated value of c into the above yields:

$$\begin{aligned}
MSE(\tilde{\beta}) &= \frac{\beta^4 x' x \sigma^2 + \beta^2 \sigma^4}{\left(\beta^2 x' x + \sigma^2\right)^2} \\
&= \frac{\frac{\beta^4 \beta^2}{\tau^2} + \frac{\beta^2 \beta^4}{\tau^4}}{\left(\beta^2 + \frac{\beta^2}{\tau^2}\right)^2} = \frac{\beta^2\left(\frac{1}{\tau^2} + \frac{1}{\tau^4}\right)}{\left(1 + \frac{1}{\tau^2}\right)^2} \\
&= \frac{\beta^2\left(1+\tau^2\right)}{\left(1+\tau^2\right)^2} = \frac{\beta^2}{\left(1+\tau^2\right)}
\end{aligned}$$

Note that $MSE(b) = \sigma^2(x'x)^{-1} = \beta^2/\tau^2$

$$\frac{MSE\left(\hat{\beta}\right)}{MSE\left(b\right)} = \frac{\beta^2/\left(1+\tau^2\right)}{\beta^2/\tau^2}$$
$$= \frac{\tau^2}{1+\tau^2}$$

As $\tau \longrightarrow \infty$ this ratio converges to 1. (Take the limit and you may refer to Rudin 1976), i.e. asymptotically OLS estimator is the minimum MSE linear estimator of β.

Greene Chapter 4 Question 3.

Suppose that the classical regression model applies but that the true value of the constant is zero. Compare the variance of the least squares slope estimator computed without a constant term with that of the estimator computed with an unnecessary constant term.

Consider a special case with one independent variable, both with and without a constant term.

In case where the slope is estimated without a constant term the variance is equal to $var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{x_2'x_2}$

In case where the slope is estimated with a constant term (x_1 is a vector of 1) the covariance matrix is:

$$var\left(\hat{\beta}^*\right) = \sigma^2\left(X'X\right)^{-1}$$
$$= \sigma^2\begin{bmatrix} x_1'x_1 & x_1'x_2 \\ x_1'x_2 & x_2'x_2 \end{bmatrix}^{-1} = \sigma^2\frac{1}{|X'X|}\begin{bmatrix} x_2'x_2 & -x_1'x_2 \\ -x_1'x_2 & x_1'x_1 \end{bmatrix}$$

The variance of the estimator for the slope parameter is therefore

$$var\left(\hat{\beta}_2^*\right) = \frac{\sigma^2 x_1'x_1}{x_1'x_1x_2'x_2 - x_1'x_2x_2'x_1}$$
$$= \frac{\sigma^2 n}{n\sum x_{2i}^2 - \left(\sum x_{2i}^2\right)\left(\sum x_{2i}^2\right)} = \frac{n\sigma^2}{n\left[\sum x_{2i}^2 - n\overline{x}_2^2\right]}$$
$$= \frac{\sigma^2}{\left[\sum x_{2i}^2 - n\overline{x}_2^2\right]}$$

Since $\overline{x}_2$ is in general different form zero, the denominator in $var\left(\hat{\beta}_2^*\right)$ (the estimate with a constant) is smaller than the denominator in $var\left(\hat{\beta}_2\right)$ (the estimate without a constant). Therefore, it is the case that:

$$var\left(\hat{\beta}_2^*\right) \geq var\left(\hat{\beta}_2\right)$$

Remark 3 *When a constant term is deleted from the regression, it can be shown that the overall goodness of fit R^2 may not be bounded between zero and one. It may be negative.*

Greene Chapter 4 Question 4.

Suppose that the regression model is $y_i = \alpha + \beta x_i + \varepsilon_i$, where the disturbances ε_i have $f(\varepsilon_i) = (1/\lambda)\exp(-\lambda\varepsilon_i)$, $\varepsilon_i \geq 0$. This model is rather peculiar in that all the disturbances are assumed to be positive. Note that the disturbances have $E[\varepsilon_i|x_i] = \lambda$ and $var(\varepsilon_i|x_i) = \lambda^2$. Show that the least squares slope is unbiased but that the intercept is biased.

We know that the estimator takes the form $(X'X)^{-1}X'y$

$$\begin{aligned}
\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} &= (X'X)^{-1}X'y = (X'X)^{-1}X'\left(X\begin{matrix}\alpha\\ \beta\end{matrix} + \varepsilon\right)\\
E\left[\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}|X\right] &= \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + (X'X)^{-1}X'E[\varepsilon|X]\\
&= \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + (X'X)^{-1}X'\begin{bmatrix}\lambda\\ \vdots\\ \lambda\end{bmatrix} \qquad \text{(Expand the matrix)}\\
&= \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + \begin{bmatrix}n & \Sigma_i x_i\\ \Sigma_i x_i & \Sigma_i x_i^2\end{bmatrix}^{-1}\begin{bmatrix}1\cdots & 1\\ x_i\cdots & x_n\end{bmatrix}'\begin{bmatrix}\lambda\\ \vdots\\ \lambda\end{bmatrix}\\
&= \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + \frac{1}{n\Sigma_i x_i^2 - \Sigma_i x_i\Sigma_i x_i}\begin{bmatrix}\Sigma_i x_i^2 & -\Sigma_i x_i\\ -\Sigma_i x_i & n\end{bmatrix}\begin{bmatrix}n\lambda\\ \lambda\Sigma_i x_i\end{bmatrix}\\
&= \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + \begin{bmatrix}\frac{n\lambda\Sigma_i x_i^2 - \lambda\Sigma_i x_i\Sigma_i x_i}{n\Sigma_i x_i^2 - \Sigma_i x_i\Sigma_i x_i}\\ \frac{n\lambda(\Sigma_i x_i - \Sigma_i x_i)}{n\Sigma_i x_i^2 - \Sigma_i x_i\Sigma_i x_i}\end{bmatrix} = \begin{bmatrix}\alpha\\ \beta\end{bmatrix} + \begin{bmatrix}\frac{n\lambda\Sigma_i x_i^2 - \lambda\Sigma_i x_i\Sigma_i x_i}{n\Sigma_i x_i^2 - \Sigma_i x_i\Sigma_i x_i}\\ 0\end{bmatrix}
\end{aligned}$$

Numerical Problem 1 (Here I use one of your solutions)

Matlab code attached.

In the entire question , we have X as the dependent variable and Y as the explanatory one. We want to find the impact of drunk driving on child

mortality. Regression result doesn't point to the way that A causes B or B causes A; it only shows there is a correlation (positive or negative) and whether the correlation is signifcant or not.