**Econ 203B: Single Equation Models**

**Solutions for Problem Set 1**

**Michael Powell**
Department of Economics, UCLA
January 14th, 2006

# 1 Greene Chapter 3

**1. The Two Variable Regression.** For the regression model $y = \alpha + \beta x + \varepsilon$.

**a.** Show that the least squares normal equations imply $\sum_i e_i = 0$ and $\sum_i x_i e_i = 0$.

**Solution** To estimate $\alpha$ and $\beta$ using OLS, we want to find

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Taking first order conditions, we have:

$$\begin{aligned}(a) &: \quad -2\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = 0 \\ (b) &: \quad -2\sum_{i=1}^{n} x_i\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = 0\end{aligned}$$

These are the normal equations. Dividing both sides by $-2$, we can rewrite these:

$$0 = \sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = \sum_{i=1}^{n} e_i \qquad (1)$$

$$0 = \sum_{i=1}^{n} x_i\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = \sum_{i=1}^{n} x_i e_i \qquad (2)$$

Which gives us the desired results.

**b.** Show that the solution for the constant term is $a = \bar{y} - b\bar{x}$.

**Solution** Using (1), we have:

$$\begin{aligned}\sum_{i=1}^{n} y_i &= \sum_{i=1}^{n} \hat{\alpha} + \sum_{i=1}^{n} \hat{\beta}x_i \\ \frac{1}{n}\sum_{i=1}^{n} y_i &= \hat{\alpha}\frac{1}{n}\sum_{i=1}^{n} 1 + \hat{\beta}\frac{1}{n}\sum_{i=1}^{n} x_i \\ \bar{y} &= \hat{\alpha} + \hat{\beta}\bar{x}\end{aligned}$$

Solving for $\hat{\alpha}$, we have:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

**c.** Show that the solution for $b$ is $b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

**Solution** Using (2), we have:

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i \hat{\alpha} + \sum_{i=1}^{n} \hat{\beta} x_i^2$$

Or

$$\hat{\beta} \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \hat{\alpha} \sum_{i=1}^{n} x_i$$

If we substitute the result from part $(b)$: $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, we get:

$$\hat{\beta} \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \left(\bar{y} - \hat{\beta}\bar{x}\right) \sum_{i=1}^{n} x_i$$

$$\hat{\beta} \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \hat{\beta}\bar{x} \sum_{i=1}^{n} x_i$$

Rearranging,

$$\hat{\beta} \left( \sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i \right) = \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i$$

Or

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

Finally, recognizing that

$$\begin{aligned}
\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n} (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y}) \\
&= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i - \bar{x} \sum_{i=1}^{n} y_i + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}
\end{aligned}$$

And

$$\begin{aligned}
\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x}) &= \sum_{i=1}^{n} \left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2
\end{aligned}$$

We get that

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

**d.** Prove that these 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 $4n\left[\left(\sum_{i=1}^{n}x_i^2\right)-n\bar{x}^2\right]=4n\left[\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right]$, which is positive unless all values of $x$ are the same.

**Solution** The Hessian matrix of $g\left(a,b\right)=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2$ is

$$\begin{aligned}\left[\begin{array}{cc}\frac{\partial^2 g}{\partial g^2} & \frac{\partial^2 g}{\partial a \partial b}\\ \frac{\partial^2 g}{\partial a \partial b} & \frac{\partial^2 g}{\partial b^2}\end{array}\right] &= \left[\begin{array}{cc}\frac{\partial}{\partial a}\left(-2\sum_{i=1}^{n}\left(y_i-a-bx_i\right)\right) & \frac{\partial}{\partial b}\left(-2\sum_{i=1}^{n}\left(y_i-a-bx_i\right)\right)\\ \frac{\partial}{\partial a}\left(-2\sum_{i=1}^{n}x_i\left(y_i-a-bx_i\right)\right) & \frac{\partial}{\partial b}\left(-2\sum_{i=1}^{n}x_i\left(y_i-a-bx_i\right)\right)\end{array}\right]\\ &= \left[\begin{array}{cc}2n & 2\sum_{i=1}^{n}x_i\\ 2\sum_{i=1}^{n}x_i & 2\sum_{i=1}^{n}x_i^2\end{array}\right]\\ &= \left[\begin{array}{cc}2n & 2n\bar{x}\\ 2n\bar{x} & 2\sum_{i=1}^{n}x_i^2\end{array}\right]\end{aligned}$$

Since $n>0$ and $x_i^2\geq 0\ \forall i$, we have that $2n, 2\sum_{i=1}^{n}x_i^2>0$ as long as $x_i\neq 0$ for some $i$. Further, we have that

$$\begin{aligned}\det\left[\begin{array}{cc}2n & 2n\bar{x}\\ 2n\bar{x} & 2\sum_{i=1}^{n}x_i^2\end{array}\right] &= 4n\sum_{i=1}^{n}x_i^2-4n^2\bar{x}^2\\ &= 4n\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)\\ &= 4n\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)\end{aligned}$$

Which, as per the instructions in the question, we note is strictly positive as long as $x_i\neq x_j$ for some $i\neq j$. Putting this together, we have shown that this Hessian matrix is positive definite whenever the $x_i$'s vary. This tells us that the objective function is strictly convex, which ensures that the first order conditions pin down a unique minimizer.

**3. Linear Transformations of the data.** 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 in the regressions of $y$ on $X$ and $y$ on $Z$ 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?

**Solution** Recall that if $y=X\beta+\varepsilon$, the OLS estimator is

$$\hat{\beta}_X=\left(X'X\right)^{-1}X'y$$

And the residuals are

$$e_X=y-X\hat{\beta}_X=y-X\left(X'X\right)^{-1}X'y=\left(I-X\left(X'X\right)^{-1}X'\right)y$$

Thus, if $y=Z\beta+\varepsilon$, the OLS estimator is:

$$\begin{aligned}\hat{\beta}_Z &= \left(Z'Z\right)^{-1}Z'y=\left(\left(XP\right)'XP\right)^{-1}\left(XP\right)'y\\ &= \left(P'X'XP\right)^{-1}P'X'y\\ &= P^{-1}\left(X'X\right)^{-1}\left(P'\right)^{-1}P'X'y\\ &= P^{-1}\left(X'X\right)^{-1}X'y\end{aligned}$$

And the residuals are

$$\begin{aligned} e_Z &= y - Z\hat{\beta}_Z = y - XPP^{-1}\left(X'X\right)^{-1}X'y \\ &= y - X\left(X'X\right)^{-1}X'y = \left(I - X\left(X'X\right)^{-1}X'\right)y \end{aligned}$$

A change in the units of measurements of the independent variables is equivalent to saying that the matrix $P$ is diagonal. Applying the above analysis, we see that a change in units of measurement of the independent variables does not affect the values of the residuals and hence would not affect the fit of the regression, which is a function only of the dependent variables and the residuals.

**6. Adding an observation.** A data set consists of $n$ observations on $X_n$ and $y_n$. The least squares estimator based on these $n$ observations is $b_n = \left(X_n'X_n\right)^{-1}X_n'y_n$. Another observation, $x_s$ and $y_s$, becomes available. Prove that the least squares estimator computed using this additional information is

$$b_{n,s} = b_n + \frac{1}{1 + x_s'\left(X_n'X_n\right)^{-1}x_s}\left(X_n'X_n\right)^{-1}x_s\left(y_s - x_s'b_n\right).$$

Note that the last term is $e_s$, the residual from the prediction of $y_s$ using the coefficients based on $X_n$ and $b_n$ Conclude that the new data change the results of least squares only if the new observation on $y$ cannot be perfectly predicted using the information already in hand.

**Solution** Define $X_{n,s} = \begin{bmatrix} X_n \\ x_s' \end{bmatrix}$, $y_{n,s} = \begin{bmatrix} y_n \\ y_{s'} \end{bmatrix}$, and $\varepsilon_{n,s} = \begin{bmatrix} \varepsilon_n \\ \varepsilon_s \end{bmatrix}$. The model which includes this extra observation is therefore,

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

Following the brilliant advice of Masa, let

$$\begin{aligned} \tilde{\beta}_{n,s} &= \left(\begin{bmatrix} X_n' & x_s \end{bmatrix}\begin{bmatrix} X_n \\ x_s' \end{bmatrix}\right)^{-1}\begin{bmatrix} X_n' & x_s \end{bmatrix}\begin{bmatrix} y_n \\ y_s \end{bmatrix} \\ &= \left(X_n'X_n + x_s x_s'\right)^{-1}\left(X_n'y_n + x_s y_s\right) \end{aligned}$$

And let

$$b_{n,s} = b_n + \frac{1}{1 + x_s'\left(X_n'X_n\right)^{-1}x_s}\left(X_n'X_n\right)^{-1}x_s\left(y_s - x_s'b_n\right).$$

Since $\left(X_n'X_n + x_s x_s'\right)^{-1}$ exists, we know that $\tilde{\beta}_{n,s} = b_{n,s}$ if and only if

$$\left(X_n'X_n + x_s x_s'\right)\tilde{\beta}_{n,s} = \left(X_n'X_n + x_s x_s'\right)b_{n,s} \tag{1}$$

It thus suffices to verify that (1) holds. Working on the left hand side of (1), we see that

$$\begin{aligned} \left(X_n'X_n + x_s x_s'\right)\tilde{\beta}_{n,s} &= \left(X_n'X_n + x_s x_s'\right)\left(X_n'X_n + x_s x_s'\right)^{-1}\left(X_n'y_n + x_s y_s\right) \\ &= X_n'y_n + x_s y_s \end{aligned} \tag{2}$$

Working on the right hand side of (1), we have:

$$\begin{aligned} \left(X_n'X_n + x_s x_s'\right)b_{n,s} &= \left(X_n'X_n + x_s x_s'\right)\left[b_n + \frac{1}{1 + x_s'\left(X_n'X_n\right)^{-1}x_s}\left(X_n'X_n\right)^{-1}x_s\left(y_s - x_s'b_n\right)\right] \\ &= X_n'X_n b_n + x_s x_s' b_n + \frac{\left(X_n'X_n\right)\left(X_n'X_n\right)^{-1}x_s\left(y_s - x_s'b_n\right)}{1 + x_s'\left(X_n'X_n\right)^{-1}x_s} \\ &\quad + \frac{x_s x_s'\left(X_n'X_n\right)^{-1}x_s\left(y_s - x_s'b_n\right)}{1 + x_s'\left(X_n'X_n\right)^{-1}x_s} \end{aligned}$$

Cleaning this up a little, we have:

$$\begin{aligned}
& \left(X_n' X_n + x_s x_s'\right) b_{n,s} \\
= \; & X_n' X_n b_n + x_s x_s' b_n + \frac{x_s y_s - x_s x_s' b_n}{1 + x_s' \left(X_n' X_n\right)^{-1} x_s} + \frac{x_s x_s' \left(X_n' X_n\right)^{-1} x_s y_s - x_s x_s' \left(X_n' X_n\right)^{-1} x_s x_s' b_n}{1 + x_s' \left(X_n' X_n\right)^{-1} x_s} \\
= \; & X_n' X_n b_n + x_s x_s' b_n + \frac{x_s y_s + x_s x_s' \left(X_n' X_n\right)^{-1} x_s y_s - x_s x_s' b_n - x_s x_s' \left(X_n' X_n\right)^{-1} x_s x_s' b_n}{1 + x_s' \left(X_n' X_n\right)^{-1} x_s} \\
= \; & X_n' X_n b_n + x_s x_s' b_n + \frac{x_s \left(1 + x_s' \left(X_n' X_n\right)^{-1} x_s\right) y_s}{1 + x_s' \left(X_n' X_n\right)^{-1} x_s} - \frac{x_s \left(1 + x_s' \left(X_n' X_n\right)^{-1} x_s\right) x_s' b_n}{1 + x_s \left(X_n' X_n\right)^{-1} x_s}
\end{aligned}$$

Which simplifies to:

$$\begin{aligned}
\left(X_n' X_n + x_s x_s'\right) b_{n,s} &= X_n' X_n b_n + x_s x_s' b_n + x_s y_s - x_s x_s' b_n \\
&= X_n' X_n b_n + x_s y_s
\end{aligned}$$

Since $y_n = X_n b_n + e_n$ or $(X_n b_n = y_n - e_n)$, we have:

$$\begin{aligned}
\left(X_n' X_n + x_s x_s'\right) b_{n,s} &= X_n' \left(y_n - e_n\right) + x_s y_s \\
&= X_n' y_n + x_s y_s - X_n' e_n
\end{aligned}$$

But, the first order conditions from the derivation of the OLS estimators gives us $X_n' e_n = 0$. Therefore,

$$\left(X_n' X_n + x_s x_s'\right) b_{n,s} = X_n' y_n + x_s y_s \tag{3}$$

Since (2) = (3), we can conclude that $\tilde{\beta}_{n,s} = b_{n,s}$.

As for the second part of the question, clearly if the new data cannot be perfectly predicted using the previous estimators, we will have that $e_s \neq 0$ and therefore,

$$\begin{aligned}
b_{n,s} &= b_n + \frac{1}{1 + x_s' \left(X_n' X_n\right)^{-1} x_s} \left(X_n' X_n\right)^{-1} x_s \left(y_s - x_s' b_n\right) \\
&= b_n + \frac{1}{1 + x_s' \left(X_n' X_n\right)^{-1} x_s} \left(X_n' X_n\right)^{-1} x_s \left(e_s\right) \\
&\neq b_n
\end{aligned}$$

**7. Deleting an observation.** 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 $b$ 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.

**Solution** To simplify the analysis for the first part of the question, I will not include a constant term, and I will assume that for the $y_s$ observation, we have no data for $x_{sj}$, $j = 1, \ldots, k$. Define $X_n^* = \begin{bmatrix} X_n & 0 \\ 0 & 1 \end{bmatrix}$, where $X_n$ is the original data matrix, $y_n^* = \begin{bmatrix} y_n \\ y_s \end{bmatrix}$, $\varepsilon_n^* = \begin{bmatrix} \varepsilon_n \\ \varepsilon^* \end{bmatrix}$ and $\beta_n^* = \begin{bmatrix} \beta_n \\ \beta^* \end{bmatrix}$. Our goal here is to estimate $\beta_n^*$ of the model

$$y_n^* = X_n^* \beta_n^* + \varepsilon_n^*$$

The OLS solution will be:

$$\begin{aligned}
\hat{\beta}_n^* &= \left(X_n^{*\prime} X_n^*\right)^{-1} X_n^{*\prime} y_n^* \\
&= \left( \begin{bmatrix} X_n' & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_n & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_n' & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_n \\ y_s \end{bmatrix} \\
&= \left( \begin{bmatrix} X_n' X_n & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} X_n' y_n \\ y_s \end{bmatrix}
\end{aligned}$$

Recall that if $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is a block diagonal matrix and $A$ and $B$ are invertible, then $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}$. This gives us:

$$\begin{aligned}
\hat{\beta}_n^* &= \begin{bmatrix} (X_n' X_n)^{-1} & 0 \\ 0 & 1^{-1} \end{bmatrix} \begin{bmatrix} X_n' y_n \\ y_s \end{bmatrix} \\
&= \begin{bmatrix} (X_n' X_n)^{-1} X_n' y_n \\ y_s \end{bmatrix} \\
&= \begin{bmatrix} \hat{\beta}_n \\ y_s \end{bmatrix}
\end{aligned}$$

Where $\hat{\beta}_n$ is the OLS estimator for $\beta_n$ in the model where the value has been discarded:

$$y_n = X_n \beta_n + \varepsilon_n$$

Does this affect $R^2$? Here, I will assume that we are asking whether or not $R_C^{2*} = R_C^2$ where $R_C^{2*}$ is the centered $R^2$ value for the model with the missing data and $R_C^2$ is the centered $R^2$ value for the model where the value has been discarded. Recall the formula for $R_C^2$ :

$$R_C^2 = 1 - \frac{\hat{y}_n' \hat{y}_n}{y_n' M^0 y_n} = \frac{\hat{y}_n' M^0 \hat{y}_n}{y_n' M^0 y_n}$$

Where, if we let $i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^2$, $M^0$ is defined as $M^0 = I_n - i\left(i'i\right)^{-1} i'$. Here, we see that

$$R_C^2 = \frac{\left(X_n \hat{\beta}_n\right)' M^0 \left(X_n \hat{\beta}_n\right)}{\left(X_n \beta\right)' M^0 \left(X_n \beta\right)}$$

Compare this to:

$$\begin{aligned}
R_C^{2*} &= \frac{\hat{y}_n^{*\prime} M^0 \hat{y}_n^*}{y_n^{*\prime} M^0 y_n^*} = \frac{\left(X_n^* \hat{\beta}_n^*\right)' M^0 \left(X_n^* \hat{\beta}_n^*\right)}{y_n^{*\prime} M^0 y_n^*} \\
&= \frac{\left( \begin{bmatrix} X_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_n \\ y_s \end{bmatrix} \right)' M^0 \left( \begin{bmatrix} X_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_n \\ y_s \end{bmatrix} \right)}{\begin{bmatrix} y_n \\ y_s \end{bmatrix}' \begin{bmatrix} y_n \\ y_s \end{bmatrix}}
\end{aligned}$$

$$= \frac{\left(\begin{bmatrix} X_n\hat{\beta}_n \\ y_s \end{bmatrix}\right)' M^0 \left(\begin{bmatrix} X_n\hat{\beta}_n \\ y_s \end{bmatrix}\right)}{\begin{bmatrix} y_n' & y_s \end{bmatrix} M^0 \begin{bmatrix} y_n \\ y_s \end{bmatrix}}$$

$$= \frac{\begin{bmatrix} \hat{\beta}_n' X_n' & y_s \end{bmatrix} M^0 \begin{bmatrix} X_n\hat{\beta}_n \\ y_s \end{bmatrix}}{\begin{bmatrix} y_n' & y_s \end{bmatrix} M^0 \begin{bmatrix} y_n \\ y_s \end{bmatrix}}$$

It can therefore be shown that

$$R_C^2 = R_C^{2*}$$

If and only if $y_s = 0$.

For the second part of the question, consider the model

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Recall from question 1 that

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x} \\ \hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{aligned}$$

Now, suppose we have an additional $k$ observations of $y$ but the corresponding data for the $x$'s are missing. Let $x_i = \bar{x}$ for $i = n+1, \ldots, n+k$, and define the new sample mean of the $y_i's$ for this larger sample: $\bar{y}^* = \frac{1}{n+k}\sum_{i=1}^{n+k} y_i$. Then we have:

$$\begin{aligned} \hat{\beta}^* &= \frac{\sum_{i=1}^{n+k} (x_i - \bar{x})(y_i - \bar{y}^*)}{\sum_{i=1}^{n+k} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}^*) + \sum_{i=n+1}^{n+k} \overbrace{(\bar{x} - \bar{x})}^{=0}(y_i - \bar{y}^*)}{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=n+1}^{n+k} \underbrace{(\bar{x} - \bar{x})^2}_{=0}} \\ &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}^*)}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \end{aligned}$$

Unless it happens to be the case that $\bar{y}^* = \bar{y}$, then it will not be the case that $\hat{\beta}^* = \hat{\beta}$. Further, if $\hat{\beta}^* \neq \hat{\beta}$ and $\bar{y}^* \neq \bar{y}$, we will obviously have that

$$\hat{\alpha}^* = \bar{y}^* - \hat{\beta}^*\bar{x} \neq \bar{y} - \hat{\beta}\bar{x} = \hat{\alpha}$$

Of course, if $\bar{y}^* = \bar{y}$, we will have that $\hat{\alpha}^* = \hat{\alpha}$. Therefore, replacing the additional observations with the means, contrary to what is being asked in the question, does indeed affect the OLS estimates unless the sample mean of the larger sample is equal to the sample mean of the smaller sample:

$$\begin{aligned} \bar{y}^* &= \bar{y} \\ \frac{1}{n+k}\sum_{i=1}^{n+k} y_i &= \frac{1}{n}\sum_{i=1}^{n} y_i \end{aligned}$$

## 2 Greene Chapter 4

**1.** Suppose that you have two independent unbiased estimators of the same parameter $\theta$, say $\hat{\theta}_1$ and $\hat{\theta}_2$, with different variances $v_1$ and $v_2$. What linear combination $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ is the minimum variance unbiased estimator of $\theta$?

**Solution** First, note that since $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased, if we are to have that $\hat{\theta}$ is unbiased, then it must be that:

$$\begin{aligned} \theta &= E\left[\hat{\theta}\right] = E\left[c_1\hat{\theta}_1 + c_2\hat{\theta}_2\right] = c_1 E\left[\hat{\theta}_1\right] + c_2 E\left[\hat{\theta}_2\right] \\ &= c_1\theta + c_2\theta = (c_1 + c_2)\,\theta \end{aligned}$$

Since $\hat{\theta}_1 \amalg \hat{\theta}_2$, we have that $cov\left(\hat{\theta}_1, \hat{\theta}_2\right) = 0$. Therefore,

$$\begin{aligned} Var\left(\hat{\theta}\right) &= Var\left(c_1\hat{\theta}_1 + c_2\hat{\theta}_2\right) = c_1^2 Var\left(\hat{\theta}_1\right) + c_2^2 Var\left(\hat{\theta}_2\right) \\ &= c_1^2 v_1 + c_2^2 v_2 \end{aligned}$$

We want to choose $c_1$ and $c_2$ to minimize $c_1^2 v_1 + c_2^2 v_2$ subject to the constraint that $c_1 + c_2 = 1$:

$$\min\,(1 - c_2)^2 v_1 + c_2^2 v_2$$

Taking the FOC wrt $c_2$ :

$$-2\,(1 - c_2)\,v_1 + 2c_2 v_2 = 0$$

$$(1 - c_2)\,v_1 = c_2 v_2$$

$$v_1 - c_2 v_1 = c_2 v_2$$

Or

$$c_2\,(v_2 + v_1) = v_1$$

$$c_2 = \frac{v_1}{v_1 + v_2}$$

And

$$\begin{aligned} c_1 &= 1 - c_2 = \frac{v_1 + v_2}{v_1 + v_2} - \frac{v_1}{v_1 + v_2} \\ &= \frac{v_2}{v_1 + v_2} \end{aligned}$$

Is this necessarily a minimum? Taking the SOC wrt $c_2$ gives:

$$2v_1 + 2v_2 > 0$$

Since variances must be non-negative (the strict inequality holds since the statement of the question specifies that the two variances are different, and thus at least one must be strictly positive). Therefore, our solution is indeed a minimizer.

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

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

**Solution** Define

$$y=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ x=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},\ \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Following the hint of the problem, we want to

$$\begin{aligned}
\min_c E\left[\left(\hat{\beta}-\beta\right)^2\middle| x\right] &= \min_c\left\{Var\left(\hat{\beta}\middle| x\right)+\left(E\left[\hat{\beta}-\beta\middle| x\right]\right)^2\right\} \\
&= \min_c\left\{Var\left(c'y| x\right)+\left(E\left[c'y-\beta| x\right]\right)^2\right\} \\
&= \min_c\left\{c'Var\left(y\ x\right)c+\left(c'E\left[y| x\right]-\beta\right)^2\right\} \\
&= \min_c\left\{c'E\left[\varepsilon^2\middle| x\right]c+\left(c'E\left[\beta x+\varepsilon| x\right]-\beta\right)^2\right\} \\
&= \min_c\left\{c'\sigma^2 I_n c+\left(\beta c'x+E\left[\varepsilon| x\right]-\beta\right)^2\right\} \\
&= \min_c\left\{\sigma^2 c'c+\left(\beta c'x-\beta\right)^2\right\} \\
&= \min_c\left\{\sigma^2 c'c+\beta^2\left(c'x\right)^2-2\beta^2 c'x+\beta^2\right\}
\end{aligned}$$

Noting that $c'x$ is a scalar, we see that $c'x=(c'x)'=x'c$. Rewriting the above equation, we have:

$$\min_c E\left[\left(\hat{\beta}-\beta\right)^2\middle| x\right]=\min_c\left\{\sigma^2 c'c+\beta^2 c'xx'c-2\beta^2 c'x+\beta^2\right\}$$

Taking first order conditions with respect to $c$:

$$(c): 2\sigma^2 c+2\beta^2 xx'c-2\beta^2 x=0$$

Where I used the fact that

$$\frac{\partial c'c}{\partial c}=\frac{\partial\left(c_1^2+\cdots+c_n^2\right)}{\partial\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}=\begin{bmatrix} 2c_1 \\ \vdots \\ 2c_n \end{bmatrix}=2c$$

And for a vector $x=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$,

$$\frac{\partial c'x}{\partial c}=\frac{\partial\left(c_1x_1+\cdots+c_nx_n\right)}{\partial\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}=x$$

And

$$\frac{\partial c'xx'c}{\partial c}=\frac{\partial\left(c'x\right)^2}{\partial c}=\frac{\partial\left(c_1x_1+\cdots+c_nx_n\right)^2}{\partial\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}}=\begin{bmatrix} 2x_1\left(c_1x_1+\cdots+c_nx_n\right) \\ \vdots \\ 2x_n\left(c_1x_1+\cdots+c_nx_n\right) \end{bmatrix}=2xx'c$$

Therefore, we have:

$$\sigma^2 c + \beta^2 x x' c - \beta^2 x = 0$$

Or (since $x'c = (c'x)$ is a scalar):

$$\begin{aligned} \beta^2 c' x x - \beta^2 x &= -\sigma^2 c \\ \beta^2 (c'x - 1) x &= -\sigma^2 c \end{aligned} \quad (1)$$

If we premultiply (1) by $x'$, we get (remembering that $\beta^2 (c'x - 1)$ is a scalar):

$$\beta^2 (c'x - 1) x'x = -\sigma^2 x'c = -\sigma^2 c'x$$

Rearranging in order to solve for $c'x$:

$$\sigma^2 c'x + \beta^2 x'x c'x = \beta^2 x'x$$

$$c'x \left(\sigma^2 + \beta^2 x'x\right) = \beta^2 x'x$$

$$c'x = \frac{\beta^2 x'x}{\sigma^2 + \beta^2 x'x} \quad (2)$$

Also, from (1), we can solve for $c$ in terms of $c'x$ by dividing through by $-\sigma^2$:

$$c = -\frac{\beta^2 x}{\sigma^2} (c'x - 1)$$

Substituting in (2) gives us:

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

Therefore, our solution is

$$\hat{\beta} = c'y = \frac{\beta^2}{\sigma^2 + \beta^2 x'x} x'y$$

**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 $b$ is

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

Note that $\tau$ is the square of the population analog to the "t ratio" for testing the hypothesis that $\beta = 0$, which is given in $(4 - 14)$. How do you interpret the behavior of this ratio as $\tau \to \infty$?

**Solution** I will proceed by calculating both $MSE\left[\hat{\beta}\right]$ and $MSE\left[b\right]$. First, recall that for the OLS estimator, $Var\left[b|\, x\right] = \sigma^2 (x'x)^{-1}$ and $E\left[b\right] = \beta = E\left[\beta\right]$:

$$\begin{aligned} MSE\left[b|\, x\right] &= Var\left[b|\, x\right] + \left(E\left[b - \beta|\, x\right]\right)^2 \\ &= \sigma^2 (x'x)^{-1} + [0]^2 \\ &= \frac{\sigma^2}{x'x} = \frac{\beta^2}{\tau^2} \end{aligned}$$

Deriving the MSE for the estimator from part $(a)$ is a bit more complicated. I will derive $Var\left[\hat{\beta}\middle| x\right]$ and $E\left[\hat{\beta}-\beta\middle| x\right]$ separately. First, substituting in $y = x\beta + \varepsilon$ into the estimator gives us:

$$\hat{\beta} = \frac{\beta^2}{\sigma^2 + \beta^2 x'x} x'\left(x\beta + \varepsilon\right) = \frac{\beta^2}{\sigma^2 + \beta^2 x'x}\beta x'x + \frac{\beta^2}{\sigma^2 + \beta^2 x'x} x'\varepsilon$$

Proceeding to derive $Var\left[\hat{\beta}\middle| x\right]$ :

$$\begin{aligned}
Var\left[\hat{\beta}\middle| x\right] &= Var\left[\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\beta x'x + \frac{\beta^2}{\sigma^2 + \beta^2 x'x} x'\varepsilon\middle| x\right] \\
&= \left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 x'\underbrace{Var\left(\varepsilon\middle| x\right)}_{=\sigma^2 I_n}x \\
&= \left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 x'\sigma^2 x \\
&= \sigma^2\left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 x'x
\end{aligned}$$

And deriving $E\left[\hat{\beta}-\beta\middle| x\right]$ :

$$\begin{aligned}
E\left[\hat{\beta}-\beta\middle| x\right] &= E\left[\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\beta x'x + \frac{\beta^2}{\sigma^2 + \beta^2 x'x} x'\varepsilon - \beta\middle| x\right] \\
&= \frac{\beta^2}{\sigma^2 + \beta^2 x'x}\beta x'x - \beta + \frac{\beta^2}{\sigma^2 + \beta^2 x'x} x'\underbrace{E\left[\varepsilon\middle| x\right]}_{=0} \\
&= \frac{\beta^3 x'x}{\sigma^2 + \beta^2 x'x} - \frac{\beta\sigma^2 + \beta^3 x'x}{\sigma^2 + \beta^2 x'x} \\
&= -\frac{\beta\sigma^2}{\sigma^2 + \beta^2 x'x}
\end{aligned}$$

Putting these two together gives us:

$$\begin{aligned}
MSE\left[\hat{\beta}\middle| x\right] &= Var\left[\hat{\beta}\middle| x\right] + \left(E\left[\hat{\beta}-\beta\middle| x\right]\right)^2 \\
&= \sigma^2\left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 x'x + \left(-\frac{\beta\sigma^2}{\sigma^2 + \beta^2 x'x}\right)^2 \\
&= \frac{\sigma^2\beta^4 x'x + \beta^2\sigma^4}{\left(\sigma^2 + \beta^2 x'x\right)^2}
\end{aligned}$$

If we substitute in $\tau^2 = \frac{\beta^2 x'x}{\sigma^2}$, we have:

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

Therefore, we have the result:

$$\frac{MSE\left[\hat{\beta}\middle|\, x\right]}{MSE\left[b|\, x\right]} = \frac{\frac{\beta^2}{1+\tau^2}}{\frac{\beta^2}{\tau^2}} = \frac{\tau^2}{1+\tau^2}$$

As for the limiting conditions, we have:

$$\lim_{\tau\to\infty} \frac{MSE\left[\hat{\beta}\middle|\, x\right]}{MSE\left[b|\, x\right]} = \lim_{\tau\to\infty} \frac{\tau^2}{1+\tau^2} = \lim_{\tau\to\infty} \frac{1}{\frac{1}{\tau^2}+1} = 1$$

Since $\lim_{\tau\to\infty} \frac{1}{\tau^2} = 0$. By a couple manipulations, we can get this into a more intuitive form:

$$1 = \lim_{\tau\to\infty} \frac{MSE\left[\hat{\beta}\middle|\, x\right]}{MSE\left[b|\, x\right]} = \frac{\lim_{\tau\to\infty} MSE\left[\hat{\beta}\middle|\, x\right]}{\lim_{\tau\to\infty} MSE\left[b|\, x\right]}$$

Or

$$\lim_{\tau\to\infty} MSE\left[\hat{\beta}\middle|\, x\right] = \lim_{\tau\to\infty} MSE\left[b|\, x\right]$$

Or, the OLS estimator and the minimum MSE estimator from part $(a)$ have the same asymptotic MSE.

**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.

**Solution** Here, we are asked to estimate two models:

$$(1) \quad : \quad Y^1 = X^1\beta^1 + \varepsilon^1,\ Y^1 = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ X^1 = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},\ \beta^1 = \beta_1^1,\ \varepsilon^1 = \begin{bmatrix} \varepsilon_1^1 \\ \vdots \\ \varepsilon_n^1 \end{bmatrix}$$

$$(2) \quad : \quad Y^2 = X^2\beta^2 + \varepsilon^2,\ Y^2 = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ X^2 = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix},\ \beta^2 = \begin{bmatrix} \beta_1^2 \\ \beta_2^2 \end{bmatrix},\ \varepsilon^2 = \begin{bmatrix} \varepsilon_1^2 \\ \vdots \\ \varepsilon_n^2 \end{bmatrix}$$

And we are to compare $Var\left(\hat{\beta}_1^1\middle|\, X\right)$ with $Var\left(\hat{\beta}_2^2\middle|\, X\right)$.

Recall that for a general OLS estimator, $Var\left(\hat{\beta}\middle|\, X\right) = \sigma^2 \left(X'X\right)^{-1}$.

For (1), we have:

$$\begin{aligned} Var\left(\hat{\beta}^1\middle|\, X\right) &= Var\left(\hat{\beta}_1^1\middle|\, X\right) = \sigma^2 \left( \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right)^{-1} \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \end{aligned}$$

For (2), we have:

$$
\begin{aligned}
Var\left(\hat{\beta}^2\middle|X\right) &= \sigma^2\left(\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}\right)^{-1} = \sigma^2\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} \\
&= \frac{\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \\
&= \sigma^2\begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} & \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} \\ \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} & \frac{n}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix}
\end{aligned}
$$

And thus

$$
\begin{aligned}
Var\left(\hat{\beta}_2^2\middle|X\right) &= Var\left(\begin{bmatrix} 0 & 1 \end{bmatrix}\hat{\beta}^2\middle|X\right) \\
&= \begin{bmatrix} 0 & 1 \end{bmatrix} Var\left(\hat{\beta}^2\middle|X\right)\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \sigma^2\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} & \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} \\ \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} & \frac{n}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \frac{\sigma^2 n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2}
\end{aligned}
$$

Since $\frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2 \geq 0$, this gives us:

$$
Var\left(\hat{\beta}_2^2\middle|X\right) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2} \geq \frac{\sigma^2}{\sum_{i=1}^n x_i^2} = Var\left(\hat{\beta}_1^1\middle|X\right)
$$

**4.** Suppose that the regression model is $y_i = \alpha + \beta x_i + \varepsilon_i$, where the disturbances $\varepsilon_i$ have $f(\varepsilon_i) = \frac{1}{\lambda}\exp(-\lambda\varepsilon_i)$, $\varepsilon_i \geq 0$. 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.

**Solution** Let $n$ be the fixed size of the sample. Then if we define:

$$
Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix},\ \gamma = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},\ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}
$$

Our model becomes $Y = X\gamma + \varepsilon$. Recognizing that:

$$
\begin{aligned}
(X'X)^{-1} &= \left(\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}\right)^{-1} \\
&= \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} \\
&= \frac{1}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} & \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} \\ \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} & \frac{n}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix}
\end{aligned}
$$

We have:

$$\begin{aligned}
E[\hat{\gamma}|X] &= E\left[(X'X)^{-1}X'Y\middle|X\right] = (X'X)^{-1}X'E[Y|X] \\
&= (X'X)^{-1}X'E[X\gamma+\varepsilon|X] = (X'X)^{-1}X'X\gamma + (X'X)^{-1}X'E[\varepsilon|X] \\
&= \gamma + (X'X)^{-1}X'\begin{bmatrix} E[\varepsilon_1|X] \\ \vdots \\ E[\varepsilon_n|X] \end{bmatrix} = \gamma + (X'X)^{-1}X'\begin{bmatrix} \lambda \\ \vdots \\ \lambda \end{bmatrix}
\end{aligned}$$

Simplifying:

$$\begin{aligned}
E[\hat{\gamma}|X] &= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} & \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ \frac{-\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} & \frac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix} \begin{bmatrix} n\lambda \\ \lambda\sum_{i=1}^n x_i \end{bmatrix} \\
&= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \frac{n\lambda\sum_{i=1}^n x_i^2 - \lambda\sum_{i=1}^n x_i \sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ \frac{n\lambda\sum_{i=1}^n x_i - n\lambda\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \end{bmatrix} \\
&= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \frac{\lambda\left(n\sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha+\lambda \\ \beta \end{bmatrix}
\end{aligned}$$

This gives us $E\left[\hat{\beta}\middle|X\right] = \beta$ and, since $\lambda \neq 0$, we have that $E[\hat{\alpha}|X] \neq \alpha$. That is, the slope estimator is unbiased and the intercept estimator is biased.

# 3 Other Questions

**1.** When dealing with time series data it is often observed that economic variables exhibit time trends, i.e. a tendency to grow (positive trend) or decline (negative trend) over time. For example, in the table below, the variable $X$, the number of deaths of children under age 1 (in thousands) exhibits a negative time trend, while the variable $Y$, the consumption of beer (in bulk barrels) exhibits a positive time trend.

| Year | 1935 | 1936 | 1937 | 1938 | 1939 | 1940 | 1941 | 1942 | 1943 | 1944 | 1945 | 1946 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | 60 | 62 | 61 | 55 | 53 | 60 | 63 | 53 | 52 | 48 | 49 | 43 |
| $Y$ | 23 | 23 | 25 | 25 | 26 | 26 | 29 | 30 | 30 | 32 | 33 | 31 |

The presence of such trends may produce spurious results when trying to estimate the relationship between two or more variables. It is thus common practice to detrend the variables first. If a variable appears to grow (or decline) linearly with time, it is reasonable to fit a linear time trend. A linear time trend may be fitted to $X$ (or $Y$) by calculating a LS regression of $X$ (or $Y$) on a constant and time $t$. The detrended values are then the residuals from that regression.

**a.** Fitting a trend requires choosing an origin and a unit of measurement for the time variable. For example, if the origin is set at mid-1935 and the unit of measurement is 1 year, then the year 1942 corresponds to $t = 7$, and so forth for the other years. If the origin is set at end-1940 (beginning of 1941) and the unit of measurement is 6 months, then 1937 corresponds to $t = -7$. Show that any computed trend values $\hat{X}_t = \hat{a} + \hat{b}t$ are unaffected by the choice of origin and unit of measurement, where $\hat{a}$ and $\hat{b}$ are the LS estimates. Do $\hat{a}$ and $\hat{b}$ change when we change the origin and unit of measurement?

**Solution** The estimators $\hat{a}$ and $\hat{b}$ for the trend line are given by the formulae:

$$\begin{aligned}
\hat{a} &= \bar{Y} - b\bar{t} \\
\hat{b} &= \frac{\sum_{t\in T}\left(Y_t - \bar{Y}\right)(t-\bar{t})}{\sum_{t\in T}\left(Y_t - \bar{Y}\right)^2}
\end{aligned}$$

Where $T$ is the index set for time, which contains $n$ elements. The choice of origin and unit of measurement is simply a linear transformation of the time variable. Define $t^* \equiv \alpha + \beta t$ and define $T^* = \{\alpha + \beta t, t \in T\}$ to be the transformed index set for time.

The first part of the question asks us to show that our estimates using this transformation of the time variable: $\hat{Y}_{t^*} = \hat{a} + \hat{b}t^*$ are equal to $\hat{Y} = \hat{a} + \hat{b}t$. I will proceed by first completing the second part of the question and then coming back to establish the claim from the first part.

For the second part of the question, first note that

$$\bar{t}^* = \frac{1}{n}\sum_{t^* \in T^*} t^* = \frac{1}{n}\sum_{t \in T}(\alpha + \beta t) = \frac{1}{n}\sum_{t \in T}\alpha + \frac{1}{n}\beta\sum_{t \in T} t = \alpha + \beta\bar{t}$$

Therefore, we have:

$$\begin{aligned}
\hat{b}^* &= \frac{\sum_{t^* \in T^*}\left(Y_t - \bar{Y}\right)(t^* - \bar{t}^*)}{\sum_{t^* \in T^*}(t^* - \bar{t}^*)^2} = \frac{\sum_{t \in T}\left(Y_t - \bar{Y}\right)(\alpha + \beta t - (\alpha + \beta\bar{t}))}{\sum_{t \in T}(\alpha + \beta t - (\alpha + \beta\bar{t}))^2} \\
&= \frac{1}{\beta}\frac{\sum_{t \in T}\left(Y_t - \bar{Y}\right)(t - \bar{t})}{\sum_{t \in T}(t - \bar{t})^2} = \frac{\hat{b}}{\beta}
\end{aligned}$$

Similarly, we have:

$$\begin{aligned}
\hat{\alpha}^* &= \bar{Y} - \hat{b}^*\bar{t}^* = \bar{Y} - \frac{\hat{b}}{\beta}(\alpha + \beta\bar{t}) = \bar{Y} - \frac{\alpha}{\beta}\hat{b} - \hat{b}\bar{t} \\
&= -\frac{\alpha}{\beta}\hat{b} + \left(\bar{Y} - \hat{b}\bar{t}\right) = -\frac{\alpha}{\beta}\hat{b} + \hat{a}
\end{aligned}$$

Clearly, for any non-trivial transformation of the time variable, $\hat{a} \neq \hat{a}^*$ and $\hat{b} \neq \hat{b}^*$.

Returning to the first part of the question (does $\hat{Y}_{t^*} = \hat{Y}_t$?), let us now use the results just established.

$$\begin{aligned}
\hat{Y}_{t^*} &= \hat{a}^* + \hat{b}^*t^* = \hat{a}^* + \hat{b}^*(\alpha + \beta t) = -\frac{\alpha}{\beta}\hat{b} + \hat{a} + \frac{\hat{b}}{\beta}(\alpha + \beta t) \\
&= -\frac{\alpha}{\beta}\hat{b} + \hat{a} + \frac{\alpha}{\beta}\hat{b} + \hat{b}t = \hat{a} + \hat{b}t = \hat{Y}_t
\end{aligned}$$

Indeed, $\hat{Y}_{t^*} = \hat{Y}_t$.

**b.** Calculate the sample correlation coefficient between $X$ and $Y$ and the LS slope coefficient of the regression of $X$ on $Y$ and a constant. Is there a positive or negative linear relationship between the two variables? What is the predicted mean of the number of children deaths if the government prohibited beer consumption? By how much does this number increase or decrease if beer consumption increases by one bulk barrel?

**Solution** According to MATLAB, $corr(X, Y) = -0.7375$. Similarly, if we regress the number of child deaths on beer consumption, that is regress $X$ on $Y$ and a constant, we have:

$$\hat{Y}_t = 92.1751 - 1.3426X_t$$

That is, there is a negative linear relationship between child deaths and alcohol consumption, quite a counterintuitive (and likely wrong!) result. If the government were to prohibit beer consumption, our model would estimate that we have $Y_i|_{X_i=0} = 92.1751$ thousand child deaths. For each bulk barrel of beer consumed, our model predicts that there will be 1343 fewer child deaths.

**c.** Even though we may expect in this example some correlation between $X$ and $Y$ (drunk driving may be correlated with children's deaths), the results from ($b$) above may be exaggerating the strength of the relationship. To see if this conjecture is correct, calculate the correlation coefficient of the detrended values of the two variables.

**Solution** Clearly the result above is spurious. After detrending, we have:

$$corr\,(\text{detrendX,detrendY}) = 0.2518$$

Which is a much more intuitive result.

**d.** Run the least squares regression of the detrended values of $X$ on the detrended values of $Y$ (without including a constant in the regression). Interpret your results.

**Solution** If we regress the detrended values of $X$ on the detrended values of $Y$, that is, if we regress detrendX on detrendY, we get:

$$\text{detrendX} = 1.0116\text{detrendY}$$

That is, after accounting for the time trend of the two variables, our model associates an increase of one bulk barrel of beer consumption with an increase of approximately 1012 deaths of children under the age of 1.

**e.** What happens if we include a constant in the LS regression above?

**Solution** If we regress detrendX on detrendY and a constant, we get the initially surprising result that:

$$\text{detrendX} = 0 + 1.0116\text{detrendY}$$

This should not be so surprising however, since the inclusion of a constant term measures the tendency for $X$ to be systematically above (or below) its trend line on average, which, by the definition of the trend line, is impossible. It would be more surprising if the constant were not zero!

**f.** Run the least squares regression of the raw (i.e. non-detrended) values of $X$ on the raw values of $Y$, a constant and time $t$. How does the slope coefficient of $Y$ compare to that computed in part ($d$)?

**Solution** If we regress $X$ on $Y$, a constant, and time $t$ (measured in years beyond 1934), we get:

$$X = 42.2326 - 2.3675t + 1.0116Y$$

Obviously, the slope coefficient of $Y$ in this part is the same as that of the result in part ($d$). This is to be expected since in this regression, as in the regressions in parts ($c$) and ($d$), we are controlling for the trend. In the model in this question, we are just doing so more explicitly.