

EC501 Econometric Methods and Applications

Problem Set 3: Sketch Solutions

Inference in the Classical Linear Regression Model

1. (a) The model is $y = X\beta + \epsilon$ where $X = [1, x_2]$, 1 being an $n \times 1$ vector of ones. We need to compute $b = (X'X)^{-1}X'y$.

$$X'y = \begin{bmatrix} 10 \\ 9 \end{bmatrix}, \quad X'X = \begin{bmatrix} 100 & 120 \\ 120 & 150 \end{bmatrix}, \quad |X'X| = 15,000 - 14,400 = 600$$

$$\Rightarrow (X'X)^{-1} = \frac{1}{600} \begin{bmatrix} 150 & -120 \\ -120 & 100 \end{bmatrix} = \begin{bmatrix} 1/4 & -1/5 \\ -1/5 & 1/6 \end{bmatrix},$$

$$b = \begin{bmatrix} 1/4 & -1/5 \\ -1/5 & 1/6 \end{bmatrix} \begin{bmatrix} 10 \\ 9 \end{bmatrix} = \begin{bmatrix} (10/4) - (9/5) \\ -(10/5) + (9/6) \end{bmatrix} = \begin{bmatrix} 0.7 \\ -0.5 \end{bmatrix}.$$

- (b) Recall that $\text{var}(b) = s^2(X'X)^{-1}$. We need to first compute $s^2 = e'e/(n - K)$ where $e = y - Xb$. Note that

$$\begin{aligned} e'e &= (y - Xb)'(y - Xb) = y'y - 2b'X'y + b'X'Xb \\ &= y'y - 2b'X'y + b'X'X(X'X)^{-1}X'y \\ &= y'y - b'X'y, \end{aligned}$$

$$\Rightarrow e'e = 74 - \begin{bmatrix} 0.7 & -0.5 \end{bmatrix} \begin{bmatrix} 10 \\ 9 \end{bmatrix} = 74 - 2.5 = 71.5,$$

$$\Rightarrow s^2 = \frac{71.5}{98} = 0.7296,$$

$$\text{var}(b) = s^2(X'X)^{-1} = 0.7296 \begin{bmatrix} 1/4 & -1/5 \\ -1/5 & 1/6 \end{bmatrix} = \begin{bmatrix} 0.1824 & -0.1459 \\ -0.1459 & 0.1216 \end{bmatrix}.$$

- (c) First note that $se(b_2) = \sqrt{0.1216} = 0.3487$, hence

$$t_2 = \frac{-0.5}{0.3487} = -1.4338.$$

The 5% critical value is $t_{98}^{0.025} = 1.98$. Because $|t_2| < t_{98}^{0.025}$ we do not reject $H_0 : \beta_2 = 0$.

- (d) It suggests that x_2 is *not* a significant determinant of y because the evidence suggests that $\beta_2 = 0$.

2. (a) The model is $y = X\beta + \epsilon$ where $X = [x_1, x_2, x_3]$.

$$X'y = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix},$$

$$b = (X'X)^{-1}X'y = \frac{1}{26} \begin{bmatrix} 7 & 2 & -1 \\ 2 & 8 & -4 \\ -1 & -4 & 15 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 16/26 \\ 12/26 \\ -6/26 \end{bmatrix}.$$

To calculate standard errors we need $\text{var}(b) = s^2(X'X)^{-1}$. We have

$$s^2 = \frac{e'e}{n - K} = \frac{y'y - b'X'y}{n - K}.$$

Here, $n = 63$, $K = 3$, $y'y = 10$ and

$$b'X'y = \frac{1}{26} [16 \quad 12 \quad -6] \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = 44/26.$$

Hence $s^2 = [10 - (44/26)]/60 = 8.3077/60 = 0.1385$ and so

$$\text{var}(b) = \frac{0.1385}{26} \begin{bmatrix} 7 & 2 & -1 \\ 2 & 8 & -4 \\ -1 & -4 & 15 \end{bmatrix} = \begin{bmatrix} 0.0373 & 0.0107 & -0.0053 \\ & 0.0426 & -0.0213 \\ & & 0.0799 \end{bmatrix}.$$

The standard errors are the square-root of the diagonal elements yielding

$$se(b_1) = 0.1931, \quad se(b_2) = 0.2064, \quad se(b_3) = 0.2827.$$

- (b) The t -statistic is $t_1 = (16/26)/0.1931 = 3.1869$. The critical value is $t_{60}^{0.025} = 2$. Because $|t_1| > t_{60}^{0.025}$ we reject $H_0 : \beta_1 = 0$ in favour of $H_1 : \beta_1 \neq 0$.
- (c) Under $H_0 : \beta_1 = \beta_3 = 0$ the model becomes $y_i = \beta_2 x_{i2} + \epsilon_i$. We shall conduct an F -test based on the RSS in the unrestricted (S) and restricted (S_R) models, given by

$$F = \left(\frac{S_R - S}{S} \right) \left(\frac{n - K}{J} \right),$$

where J denotes the number of restrictions. Here, $n = 63$, $K = 3$, $J = 2$, and from part (a) we have $S = 8.3077$. Under H_0 ,

$$b_{2R} = \frac{\sum x_{i2}y_i}{\sum x_{i2}^2} = \frac{1}{4} = 0.25$$

and so $S_R = e'_R e_R = y'y - b_{2R}(\sum x_{i2}y_i) = 10 - 0.25 = 9.75$. Hence

$$F = \frac{9.75 - 8.3077}{8.3077} \cdot \frac{60}{2} = 5.2083.$$

The 5% critical value is $\bar{F}_{2,60} = 3.15$ and, because $F > \bar{F}_{2,60}$ we reject $H_0 : \beta_1 = \beta_3 = 0$ in favour of $H_1 : \beta_1 \neq 0$ or $\beta_3 \neq 0$ or both.

Apart from rounding errors, the same result can be obtained as follows:

$$H_0 : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

then

$$\begin{aligned}
 F &= (Rb - q)' [s^2 R(X'X)^{-1} R']^{-1} (Rb - q) / J \\
 &= 0.5 \begin{bmatrix} \frac{8}{13} & -\frac{3}{13} \end{bmatrix} \left[\frac{0.1385}{26} \begin{bmatrix} 7 & -1 \\ -1 & 15 \end{bmatrix} \right]^{-1} \begin{bmatrix} \frac{8}{13} \\ -\frac{3}{13} \end{bmatrix} \\
 &= 5.2069
 \end{aligned}$$

3. (a) The required t -statistic can be calculated as

$$t = \frac{0.9817 - 1}{0.0465} = -0.3935,$$

for which the 5% critical value is $t_{32}^{0.025} = 2.04$. Because $|t| < t_{32}^{0.025}$ we do not reject $H_0 : \beta_2 = 1$.

- (b) From the regression output we find that the relevant t -statistic is 0.780 with a p-value of 0.441. This implies that a total of 44.1% of the $t_{32}^{0.025}$ distribution lies beyond ± 0.78 and hence we do not reject $H_0 : \beta_3 = 0$ at the 5% level of significance.
- (c) We want to test the restriction

$$H_0 : \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_1 \end{bmatrix} - 1 = 0.$$

We know that

$$F = (Rb - q)' [s^2 R(X'X)^{-1} R']^{-1} (Rb - q) / J.$$

Now

$$(Rb - q) = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} .9816755 \\ .0313708 \\ .0020919 \end{bmatrix} - 1 = 0.013046$$

$$s^2 R(X'X)^{-1} R' = 0.00215810 + 0.00161916 - 2 \times 0.00186195 = 0.00005336$$

and therefore

$$F = \frac{(Rb - q)^2}{s^2 R(X'X)^{-1} R'} = \frac{0.013046^2}{0.00005336} = 3.1896$$

The 5% critical value is $\bar{F}_{1,32} = 4.149$ and, because $F < \bar{F}_{1,32}$, we do not reject $H_0 : \beta_2 + \beta_3 = 1$.

- (d) From the regression output, we have that $F = 30523.24$. The 5% critical value is $\bar{F}_{2,32} = 3.295$ and, because $F > \bar{F}_{2,32}$, we reject $H_0 : \beta_2 = 0$ and $\beta_3 = 0$.
- (e) The non-rejection of the null hypothesis in part (a) implies that there is a unit elasticity of output prices with respect to labour costs, while the non-rejection of the null in part (b) implies that import costs do not affect output prices. Part (c) tests for whether the sum of the two elasticities is unity. In view of the results in (a) and (b) it is not surprising that we do not reject this hypothesis. Note, however, that the null in (c) does not impose $\beta_2 = 1$ and $\beta_3 = 0$. In fact, any combination of values of β_2 and β_3 that sum to unity satisfies this hypothesis.

4. (a) Pre-multiplying the estimated model by $X'M_Z$ yields

$$X'M_Z y = X'M_Z X \hat{\beta} + X'M_Z Z \hat{\gamma} + X'M_Z e.$$

We know that $M_Z Z = 0$ and find that

$$X'M_Z e = X'(I - Z(Z'Z)^{-1}Z')e = X'e - X'Z(Z'Z)^{-1}Z'e = 0$$

due to the orthogonality of e with X and Z with e i.e., $X'e = 0$ and $Z'e = 0$. Hence $X'M_Z y = X'M_Z X \hat{\beta}$ which yields $\hat{\beta} = (X'M_Z X)^{-1} X'M_Z y$ as required.

- (b) The estimators $\hat{\beta}$ and $\hat{\gamma}$ are best linear unbiased estimators (BLUEs) of β and γ , respectively, because the model satisfies the classical assumptions.
- (c) If $E(y|X, Z) = X\beta$ but (2) is estimated, the latter model includes irrelevant variables. How does this affect the expectation and variance of the estimators? We have

$$E(\hat{\beta}|X, Z) = (X'M_Z X)^{-1} X'M_Z X \beta = \beta.$$

Hence, the inclusion of irrelevant variables does not affect the unbiasedness of the OLS estimator of β . What about the variance of $\hat{\beta}$? We have that

$$\begin{aligned} \text{var}(\hat{\beta}|X, Z) &= (X'M_Z X)^{-1} X'M_Z \text{var}(\epsilon|X, Z) M_Z' X (X'M_Z X)^{-1} \\ &= \sigma^2 (X'M_Z X)^{-1}, \end{aligned}$$

using the idempotency of M_Z . It can be shown that the matrix difference $\sigma^2 (X'M_Z X)^{-1} - \sigma^2 (X'X)^{-1}$ is positive semi-definite, so that in general there is an **increase in variance resulting from including irrelevant variables** in the model. However, it is easy to show that $\sigma^2 (X'M_Z X)^{-1} = \sigma^2 (X'X)^{-1}$ when $X'Z = 0$, and therefore in this very special case there are no consequences of including irrelevant variables.

- (d) When (1) is estimated but $E(y|X, Z) = X\beta + Z\gamma$, we obtain

$$E(b|X, Z) = (X'X)^{-1} X'(X\beta + Z\gamma) = \beta + (X'X)^{-1} X'Z\gamma.$$

Hence, b is biased when relevant variables (Z) are omitted from the regression unless X and Z are orthogonal i.e., $X'Z = 0$. The variance of b is naturally

$$\text{var}(b|X, Z) = \sigma^2 (X'X)^{-1}$$

which, as noted above, can be smaller than $\text{var}(\hat{\beta}|X, Z)$. That is, **omission of relevant variables biases the OLS estimator, but may reduce its variance.**