University of Essex — Department of Economics
Session: 2011/12 — João Santos Silva

EC501 Econometric Methods and Applications

Problem Set 2: Sketch Solutions

The Classical Linear Regression Model (continued)

1. With these data e have

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 \\ 0 & 26 \end{bmatrix}, \qquad (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{26} \end{bmatrix}, \qquad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 9 \\ 10 \end{bmatrix},$$

hence

$$\mathbf{b} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{26} \end{bmatrix} \begin{bmatrix} 9 \\ 10 \end{bmatrix} = \begin{bmatrix} \frac{9}{4} \\ \frac{5}{13} \end{bmatrix}.$$

2. (a) A symmetric idempotent matrix $\mathbf{Q}$ satisfies the properties $\mathbf{Q} = \mathbf{Q}'$ (symmetric) and $\mathbf{Q} = \mathbf{Q}\mathbf{Q} = \mathbf{Q}^2$ (idempotent). Here, $\mathbf{P}$ is clearly symmetric because $\mathbf{P}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}$, while

$$\mathbf{P}^2 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{P}$$

and so $\mathbf{P}$ is idempotent. Now $\mathbf{M}' = \mathbf{I}' - \mathbf{P}' = \mathbf{I} - \mathbf{P} = \mathbf{M}$ (using the previously established symmetry of $\mathbf{P}$) and so $\mathbf{M}$ is also symmetric, while

$$\mathbf{M}^2 = (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) = \mathbf{I} - 2\mathbf{P} + \mathbf{P}^2 = \mathbf{I} - \mathbf{P} = \mathbf{M}$$

which uses the idempotency of $\mathbf{P}$ to show that $\mathbf{M}$ is idempotent.

(b) Taking each in turn and using the properties in (a):

$$\mathbf{PM} = \mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{P} - \mathbf{P}^2 = \mathbf{P} - \mathbf{P} = 0,$$

$$\mathbf{MP} = (\mathbf{I} - \mathbf{P})\mathbf{P} = \mathbf{P} - \mathbf{P}^2 = \mathbf{P} - \mathbf{P} = 0,$$

$$\mathbf{M}'\mathbf{P} = \mathbf{MP} = 0, \qquad \mathbf{P}'\mathbf{M} = \mathbf{PM} = 0,$$

$$\mathbf{PX} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}$$

$$\mathbf{MX} = (\mathbf{I} - \mathbf{P})\mathbf{X} = \mathbf{X} - \mathbf{PX} = \mathbf{X} - \mathbf{X} = 0.$$

3. Recall that $\mathbf{e} = \mathbf{M}\epsilon$ where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric and idempotent and that $E(\mathbf{b}) = \beta$ so that $\mathbf{b} - E(\mathbf{b}) = \mathbf{b} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon$. Thus

$$\begin{aligned} E\{e[\mathbf{b} - E(\mathbf{b})]'|\mathbf{X}\} &= E\{\mathbf{M}\epsilon[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon]'|\mathbf{X}\} \\ &= E[\mathbf{M}\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}] \\ &= \sigma^2\mathbf{MX}(\mathbf{X}'\mathbf{X})^{-1} = 0 \end{aligned}$$

because $E(\epsilon\epsilon'|\mathbf{X}) = \sigma^2 I_n$ and $\mathbf{MX} = 0$. Then,

$$\begin{aligned} cov(\mathbf{e}, \mathbf{b}) &= E\{e[\mathbf{b} - E(\mathbf{b})]'\} \\ &= E_{\mathbf{X}}\{\mathbf{e}[\mathbf{b} - E(\mathbf{b})]'|\mathbf{X}\} \\ &= E_{\mathbf{X}}\{0\} = 0. \end{aligned}$$

4. (a) First note that

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon) = \mathbf{X}\beta + \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\epsilon$$

so that $\hat{\mathbf{y}} = \mathbf{X}\beta + \mathbf{P}\epsilon$ where $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ is a projection matrix. It follows that

$$\begin{aligned}
E(\hat{\mathbf{y}}|\mathbf{X}) &= \mathbf{X}\beta, \\
var(\hat{\mathbf{y}}|X) &= var(\mathbf{X}\beta + \mathbf{P}\epsilon|X) \\
&= var(\mathbf{P}\epsilon|X) \text{ because we are conditioning on } X \\
&= \mathbf{P}var(\epsilon|X)\mathbf{P}' \\
&= \sigma^2\mathbf{P}\mathbf{P}' \text{ because } E(\epsilon\epsilon'|\mathbf{X}) = \sigma^2 I_n \\
&= \sigma^2\mathbf{P} \text{ because } \mathbf{P} \text{ is symmetric and idempotent}
\end{aligned}$$

Therefore $var(\hat{\mathbf{y}}|\mathbf{X}) = \sigma^2\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$.

(b) Let f denote the forecast error:

$$\begin{aligned}
f &= \text{“actual minus predicted”} \\
&= y_n - \hat{y}_n \\
&= x_n'\beta + \epsilon_n - x_n'\mathbf{b}
\end{aligned}$$

i.e. $f = \epsilon_n - x_n'(\mathbf{b} - \beta)$. Clearly $E(f) = 0$ because $E(\epsilon_n) = 0$ and $E(\mathbf{b} - \beta) = 0$. Then, noting that $E\left[\epsilon_n x_n'(\mathbf{b} - \beta)|X\right] = 0$ (so the cross-product terms in the variance are zero):

$$\begin{aligned}
var(f|X) &= var(\epsilon_n|X) + var[x_n'(\mathbf{b} - \beta)|\mathbf{X}] \\
&= \sigma^2 + E[x_n'(\mathbf{b} - \beta)(\mathbf{b} - \beta)'x_n|\mathbf{X}] \\
&= \sigma^2 + \sigma^2 x_n'\left(\mathbf{X}'\mathbf{X}\right)^{-1}x_n,
\end{aligned}$$

i.e. $var(f|X) = \sigma^2[1 + x_n'\left(\mathbf{X}'\mathbf{X}\right)^{-1}x_n]$ (a scalar).

5. (a) In the worst possible case, where the model explains nothing, $RSS = TSS$ and hence $R^2 = 0$. In the best possible case, where the model explains *all* the variation in y, the $RSS = 0$ and hence $R^2 = 1$. In practice, R^2 lies between these bounds (but do not forget that if it is computed this way in a model not containing an intercept, it can be negative!)

(b) Note that we can write

$$\bar{R}^2 = 1 - \frac{RSS}{TSS}\frac{(n-1)}{(n-K)}.$$

From the definition of R^2, $RSS/TSS = 1 - R^2$. The result is then immediate by substitution.

(c) Let $\bar{R}_a^2$ denote the $\bar{R}^2$ in the augmented model which contains $K+H$ regressors, and let RSS_a denote the corresponding RSS. Then we are interested in

$$\begin{aligned}
\bar{R}_a^2 - \bar{R}^2 &= \left[1 - \frac{RSS_a}{TSS}\frac{(n-1)}{(n-K-H)}\right] - \left[1 - \frac{RSS}{TSS}\frac{(n-1)}{(n-K)}\right] \\
&= \frac{RSS}{TSS}\frac{(n-1)}{(n-K)} - \frac{RSS_a}{TSS}\frac{(n-1)}{(n-K-H)} \\
&= \frac{(n-1)}{TSS}\left[\frac{RSS}{(n-K)} - \frac{RSS_a}{(n-K-H)}\right] \\
&> 0 \text{ if } \frac{RSS}{(n-K)} > \frac{RSS_a}{(n-K-H)} \\
&\Rightarrow \frac{RSS_a}{RSS} < \frac{(n-K-H)}{(n-K)} \Rightarrow \frac{RSS - RSS_a}{RSS} > \frac{H}{(n-K)}.
\end{aligned}$$

In words, the percentual reduction in RSS achieved by the additional regressors must be more than the percentual reduction in the number of degrees of freedom, in order for $\bar{R}_a^2 > \bar{R}^2$.