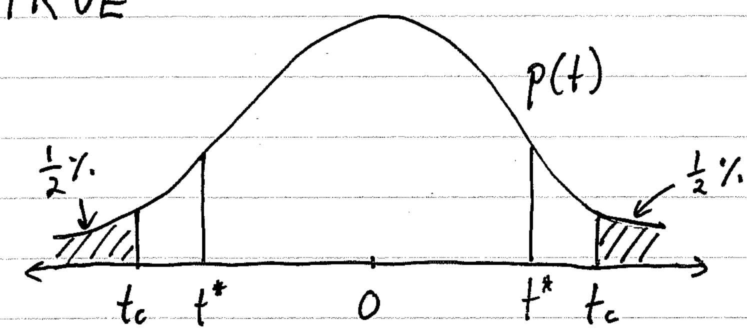
Sample Midterm - Solutions

1. C 2. C 3. C 4. D 5. C

6. a) TRUE



Reject H_0 at 1% level if $|t| > t_c$. Reject at the 5% or 10% level requires $|t| > t^*$, for some t^* as shown - this will automatically occur if we are rejecting at the 1% level.

b) FALSE. In each case the null hypothesis is that the regressors are asymptotically uncorrelated with the errors - $H_0: \text{plim}(\frac{1}{n}X'\varepsilon) = 0$. Further, rejection of the null hypothesis would lead us to use I.V. estimation, and not OLS.

c) FALSE. $E[b_1/b_2] \neq E(b_1)/E(b_2)$. So, $E(b_1/b_2) \neq [\beta_1/\beta_2]$. However, by Slutsky's Theorem, $\text{plim}(b_1/b_2) = \text{plim}(b_1)/\text{plim}(b_2) = (\beta_1/\beta_2)$.

d) FALSE. $e_1 = -e_n$. This is because $X'e = 0$.

So, $\begin{bmatrix} - & - & - & - & - \\ 1 & 0 & 0 & \cdots & 0 & 1 \\ - & - & - & - & - \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$, and so $e_1 + e_n = 0$, or $e_1 = -e_n$

e) TRUE: $MSE = var + Bias^2 \geq var$

7 (a) – (c) See assignment #1.

d) $\hat{\varepsilon}_i = y_i - \hat{\beta} x_i = y_i - (\bar{y}/\bar{x}) x_i$

$$\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = \sum_{i=1}^{n} x_i (y_i - \bar{y}/\bar{x}\, x_i) = \sum_{i=1}^{n} x_i y_i - \frac{\bar{y}}{\bar{x}} \sum_{i=1}^{n} x_i^2 \neq 0$$

(in general)

e) If $x_i = 1$; $\forall i$, then $\bar{x} = 1$.

$$\text{So, } x'\hat{\varepsilon} = \sum_{i=1}^{n} x_i \hat{\varepsilon}_i = \sum_{i=1}^{n} x_i y_i - \bar{y}/\bar{x} \sum_{i=1}^{n} x_i^2$$

$$= \sum_{i=1}^{n} y_i - \bar{y} \sum_{i=1}^{n} 1^2 = n\bar{y} - n\bar{y} = 0$$

f) $$\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} \left[y_i - \frac{\bar{y}}{\bar{x}} x_i \right] = \sum_{i=1}^{n} y_i - \frac{\bar{y}}{\bar{x}} \sum_{i=1}^{n} x_i = n\bar{y} - \left(\frac{\bar{y}}{\bar{x}}\right) n\bar{x}$$

$$= n\bar{y} - n\bar{y} = 0.$$

g) $\hat{y}_i = \hat{\beta} x_i$. Let $x_i = \bar{x}$.

Then $\hat{y}_i = \hat{\beta} \bar{x} = \frac{\bar{y}}{\bar{x}} \bar{x} = \bar{y}$. Yes, it does pass through $(\bar{x}, \bar{y})$.

8 a) $e_{IV} = y - Xb_{IV} = y - X(Z'X)^{-1}Z'y$

$$= X\beta + \varepsilon - X(Z'X)^{-1}Z'(X\beta + \varepsilon)$$
$$= X\beta + \varepsilon - X\underbrace{(Z'X)^{-1}Z'X}_{I}\beta - X(Z'X)^{-1}Z'\varepsilon$$
$$= X\beta + \varepsilon - X\beta - X(Z'X)^{-1}Z'\varepsilon$$
$$= \varepsilon - X(Z'X)^{-1}Z'\varepsilon = \underbrace{(I - X(Z'X)^{-1}Z')}_{W}\varepsilon$$
$$= W\varepsilon$$

b) $e_{IV}'e_{IV} = \varepsilon'W'W\varepsilon = \varepsilon'[I - Z(X'Z)^{-1}X'][I - X(Z'X)^{-1}Z']\varepsilon$

$$= \varepsilon'\varepsilon - \varepsilon'Z(X'Z)^{-1}X'\varepsilon - \varepsilon'X(Z'X)^{-1}Z'\varepsilon + \varepsilon'Z(X'Z)^{-1}X'X(Z'X)^{-1}Z'\varepsilon$$

c) $s_{IV}^2 = \dfrac{e_{IV}'e_{IV}}{n}$

$$= \frac{\varepsilon'\varepsilon}{n} - \frac{\varepsilon'Z(X'Z)^{-1}X'\varepsilon}{n} - \frac{\varepsilon'X(Z'X)^{-1}Z'\varepsilon}{n} + \frac{\varepsilon'Z(X'Z)^{-1}X'X(Z'X)^{-1}Z'\varepsilon}{n}$$

Now, (i) $\text{plim}\left(\dfrac{\varepsilon'Z(X'Z)^{-1}X'\varepsilon}{n}\right) = \text{plim}\left[\left(\dfrac{Z'\varepsilon}{n}\right)'\left(\dfrac{X'Z}{n}\right)^{-1}\left(\dfrac{X'\varepsilon}{n}\right)\right]$

$$= 0'Q_{ZX}'^{-1}\gamma = 0$$

(ii) $\dfrac{\varepsilon'X(Z'X)^{-1}Z'\varepsilon}{n} = \left(\dfrac{X'\varepsilon}{n}\right)'\left(\dfrac{Z'X}{n}\right)^{-1}\left(\dfrac{Z'\varepsilon}{n}\right)$

$$\text{plim} = \gamma' Q_{ZX}^{-1} 0' = 0$$

(iii) $$\frac{\varepsilon' Z(X'Z)^{-1}X'X(Z'X)^{-1}Z'\varepsilon}{n} = \left(\frac{Z'\varepsilon}{n}\right)'\left(\frac{Z'X}{n}\right)'^{-1}\left(\frac{X'X}{n}\right)\left(\frac{Z'X}{n}\right)^{-1}\left(\frac{Z'\varepsilon}{n}\right)$$

$$\text{plim} = 0' Q_{ZX}'^{-1} Q_{XX} Q_{ZX}^{-1} 0 = 0$$

So, $\text{plim } s_{IV}^2 = \text{plim}\left(\frac{\varepsilon'\varepsilon}{n}\right) = \text{plim}\left[\frac{1}{n}\sum_i \varepsilon_i^2\right]$

Now, if the ε_i's are independent, so are the ε_i^2. By Khintchine's Theorem,

$$\text{plim}\left(\frac{1}{n}\sum \varepsilon_i^2\right) = E(\varepsilon_i^2) = \sigma^2$$

So, $\text{plim}(s_{IV}^2) = \sigma^2$.

9 (a) Yes they do. Larger CITES and CLAIMS means a more complex application, requiring more checking, and a longer duration until accepted. The negative coefficient on APPYEAR may reflect an increase in the efficiency of the patent office over time.

(b) They are essentially identical because n is nearly 2 million.

$$R^2 = 1 - \frac{e'e}{y'M_0 y} \qquad \bar{R}^2 = 1 - \frac{e'e/(n-k)}{y'M_0 y/(n-1)}$$

and $(n-1)$, $(n-k)$ basically equal if n very large.

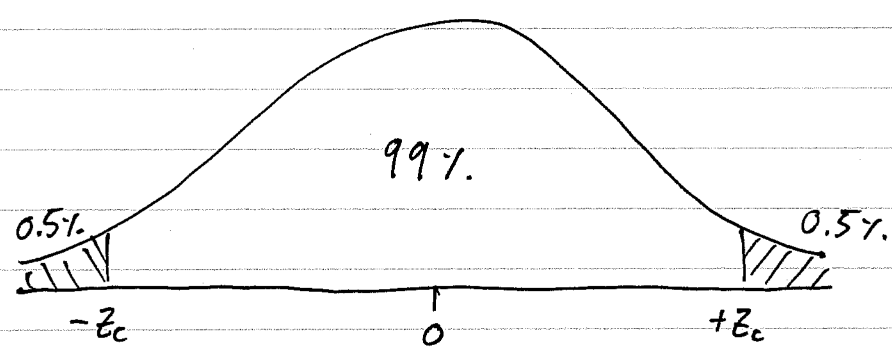
The model explains only 5.2% of the variation.

(c) S.E. Regression $= s = \sqrt{s^2} = \sqrt{\frac{e'e}{n-k}} = \frac{406426.7}{(1983420-9)} = 0.45267$

$t_i = \frac{b_i}{s.e.(b_i)}$, so $s.e.(b_i) = \left(\frac{b_i}{t_i}\right)$ and,

$s.e.(b_7) = \frac{0.179390}{135.3954} = 0.001325$

(d) Because n is so large, the t-statistics are essentially standard-normally distributed, so use the z-table (or t-table w/ dof $= \infty$).



Need to have area to left of $+z_c$ to be 0.995, So, $z_c = 2.576$.
The 99% C.I. is

$$0.056065 \pm (2.576)(0.001001), \quad \text{or:}$$

$$(0.053486,\ 0.058644)$$

The units are log-years. (The dummy variable has no units, so the coefficient's units are the same as for $y = \log(time)$).
Interpretation: if we constructed intervals like this, many times, then 99% of such random intervals would cover the true value of β_5.

(e) $\log(\widehat{TIME}) = 23.24623 + (8)(0.003776) + (12)(0.002786)$
$- (1986)(0.011506) + 0.056065$

$= 0.515019$

So, $\widehat{TIME} = \exp(0.515019) = 1.674$ years