Dudin George, group 201, data: 04

№1

Model: $M_i = \beta K_i + u_i$, where: K_i - # of orders
M_i - money earnd

1. $b = \dfrac{\widehat{Cov}(M, K)}{\widehat{Var}(K)}$

We have no intercept, therefore I think, that the better choice would be:

$$b^*_{OLS} = \frac{\sum_{i=1}^{n} M_i K_i}{\sum_{i=1}^{n} K_i^2}$$

a) Linear? $b^*_{OLS} = \sum_{i=1}^{n} M_i K_i \cdot \underbrace{\frac{1}{\sum_{i=1}^{n} K_i^2}}_{\text{non-stochastic}} \Rightarrow$

$\Rightarrow$ it's linear

b) Unbiased?

$$E(b^*_{OLS}) = E\left(\frac{\sum_{i=1}^{n} M_i K_i}{\sum_{i=1}^{n} K_i^2}\right) =$$

$$= E\left(\frac{\sum_{i=1}^{n} B \cdot \cancel{K_i^2}}{\cancel{\sum_{i=1}^{n} K_i^2}}\right) + E\left(\frac{\sum_{i=1}^{n} K_i \overset{\to 0}{u_i}}{\sum_{i=1}^{n} K_i^2}\right) = B \Rightarrow$$

$\Rightarrow$ Unbiased

Let's check the same for b:

a) Linear? $b = \sum_i a_i m_i$ ← from presentation

$$a_i = \frac{M_i - \bar{M}}{\sum_{j=1}^{n} (K_j - \bar{K})^2} \qquad k_i = K_i - \bar{K} \qquad m_i = M_i - \bar{M}$$

$$b = \frac{\sum_i (K_i - \bar{K})(M_i - \bar{M})}{\sum (K_i - \bar{K})} = \frac{\sum k_i m_i}{\sum_i k_i^2} = \sum a_i m_i$$

$$\Rightarrow b = \sum_i a_i \cdot (M_i - \bar{M}) = \sum_i a_i M_i - \sum_i a_i \bar{M} \Rightarrow$$

$\Rightarrow$ linear

b) Unbiased? $E(b) = E(B + \sum a_i u_i) =$

Unbiased. $\Longleftarrow = B + \underbrace{E(\sum a_i u_i)}_{=0} = B$

2. $$Var(b^*_{OLS}) = E((b^*_{OLS} - B)^2) = E\left(\left(\frac{\sum k_i u_i}{\sum k_i^2}\right)^2\right) =$$

$$= \frac{E\left(\sum k_i^2 u_i^2 + \sum_{z=1}^{n} \sum_{j \neq z}^{n} k_z k_j u_z u_j\right)}{(\sum k_i^2)^2} =$$

$$= \frac{\sum k_i^2 \cdot \sigma_u^2}{(\sum k_i^2)^2} = \frac{\sigma_u^2}{\sum k_i^2}$$

$$Var(b) = \frac{\sigma_u^2}{\sum_i (k_i - \bar{k})^2} = \frac{\sigma_u^2}{\sum k_i^2 - n\bar{k}} \geq Var(b^*_{OLS})$$

=> b* is better estimator, then b

from Gauss-Markov theorem : b* _ BLUE

3. If $n \to \infty$: is b^*_{OLS} - consistent?

$$plim(b^*_{OLS}) = plim\left(\frac{\sum k_i M_i}{\sum k_i^2}\right) =$$

$$= plim\left(\frac{B\sum k_i^2 + \sum k_i u_i}{\sum k_i^2}\right) =$$

$$= plim\left(B + \frac{E(ku)}{E(k)}\right) =$$

$$= plim\left(B + \frac{E(k)\overbrace{E(u)}^{=0} + \overbrace{Cov(k,u)}^{=0}}{E(k^2)}\right) =$$

$$= B \rightarrow \text{consistent.}$$

№2.

1. $y_i = \beta_1 + \beta_2 X_i + u_i$

$$TSS_n = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$R^2_n = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$TSS_{n+1} - TSS_n = (y_{n+1} - \bar{y})^2 \geq 0 \Rightarrow$$

$\Rightarrow$ TSS won't decrease

$$R^2_{n+1} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + (\hat{y}_{n+1} - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2 + (y_{n+1} - \bar{y})^2} \Rightarrow$$

$\Rightarrow$ R^2 could either increase or decrease

2. a) Of course don't throw away observations. They could represent a „new" class of schools, which were not in the base-data set.

b) Researcher should check if the data is correct

c) Researched should try to increase the dataset to cover more schools

d) This would slightly affect the OLS slope.