Osama K Mahmood

Econometrics

2/March/2018

## Review

Multiple Regression Model

- Motivation
- Derivation
- Assumptions
- Multicolinearity
- Omitted Variable Bias
- $\bar{R}^2$

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{u} \qquad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{SST_x}$$

$$y = \tilde{\beta}_1 x + \tilde{u} \qquad \tilde{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$E(\tilde{\beta}_1) = \beta_1 + \beta_0 \frac{n\bar{x}}{\sum x_i^2}$$

If $\bar{x} \cdot \beta_0 < 0 \Rightarrow \text{Bias} < 0$

$\bar{x} \cdot \beta_0 > 0 \Rightarrow \text{Bias} > 0$

$$V(\hat{\beta}_1) = \frac{\sigma^2}{SST_x} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$V(\tilde{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$$

Hence $V(\hat{\beta}_1) > V(\tilde{\beta}_1)$

# Multiple Linear ~~Rg~~ Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + u$$

Why

- → Realistic
- → To estimate non-linear relations
- → Closer to ceteris paribus interpretation

Consider the case of two variables:

$$\underset{\text{wage}}{y} = \beta_0 + \underset{\text{Education}}{\beta_1 x_1} + \underset{\text{Experience}}{\beta_2 x_2} + u$$

Hence, $\beta_1$ is the effect of one more year of education on wage, keeping experience constant.

$$x_1 = \hat{\gamma}_0 + \hat{\gamma}_1 \underset{\text{Experience}}{x_2} + \hat{r}_1 = \hat{x}_1 + \hat{r}_1$$

$\Rightarrow$ reg $y$ on $\hat{r}_1$

Similarly $x_2 = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{r}_2$

$x_2 = \hat{x}_2 + \hat{r}_2$

reg $y$ on $\hat{r}_2$

Computing residuals

→ reg educ exper

→ predict r1hat, res

→ reg wage r1hat

$$\hat{\beta}_1 = \frac{\sum(\hat{r}_1 - \bar{\hat{r}}_1) y_i}{\sum(\hat{r}_1 - \bar{\hat{r}}_1)^2} = \frac{\sum(\hat{r}_j y)}{\sum \hat{r}_j^2} \quad (\text{since } \bar{\hat{r}}_i = 0)$$

$$\boxed{\hat{\beta}_j = \frac{\sum \hat{r}_j y}{\sum \hat{r}_j^2}}$$

<u>Derivation</u>

$$\min \sum (y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2)^2$$

$$\frac{\partial}{\partial \beta_1} = 0 \implies \sum x_1 (y - \beta_0 - \beta_1 x_1 - \hat{\beta}_2 x_2) = 0$$

$$\downarrow$$

$$\hat{x}_1 + \hat{r}_1$$

$$\hat{\gamma}_0 + \hat{\gamma}_1 x_2$$

(Since $\hat{x}_1$ is not correlated with residual)

$$= \sum \hat{r}_1 (y - \beta_0 - \beta_1 x_1 - \beta_2 x_2) = 0$$

$$= \sum r_1 (y - \beta_1 x_1)$$

$$= \sum r_1 (y - \beta_1 (\hat{x}_1 + \hat{r}_1))$$

$$= \sum r_1 (y - \hat{\beta}_1 \hat{r}_1) = 0 \implies \boxed{\hat{\beta}_1 = \frac{\sum \hat{r}_j y}{\sum \hat{r}_j^2}}$$

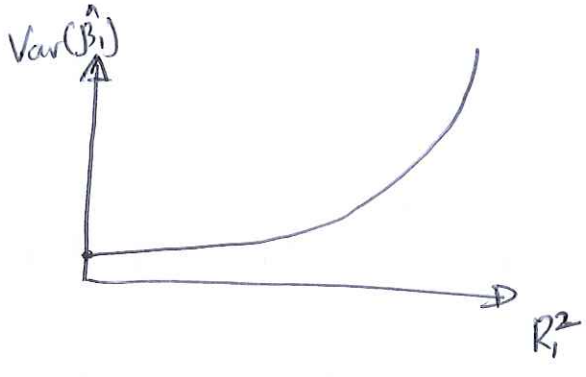
$$\hat{\beta}_1 = \frac{\sum \hat{r}_1 y}{SSR}$$

$$SST = SSE + SSR \Rightarrow R^2 = 1 - \frac{SSR}{SST}$$

$$\Rightarrow SSR = SST(1 - R_1^2)$$

Hence, $$\hat{\beta}_1 = \frac{\sum \hat{r}_1 y}{SST_1(1 - R_1^2)}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST(1 - R_1^2)}$$




Multicolinearity

Perfect $R_j^2$ is 1

$R_j^2$ close to 1

$\Rightarrow$ High variance (High Standard Error)

$\Rightarrow$ Small t-stat

$\Rightarrow$ Less (non) significant co-efficients

$\Rightarrow$ High instability in the co-efficients

VIF → Variance Inflation Factor

If VIF > 10 (problem with multicolinearity) [rule of thumb]

$$VIF_j = \frac{1}{1 - R_j^2}$$

## Solutions to Multicolinearity

1. Ignore it (If highly correlated variables are not of interest)
2. Increase the sample size
3. Increase $SST_j$
4. Create a new variable : $X_{new} = \alpha_1 x_3 + \alpha_2 x_4 + (1 - \alpha_1 - \alpha_2) x_5$ ; $\alpha_1 + \alpha_2 \leq 1$

3: (Take a larger variance in the same sample)

## Assumptions

A1. Linearity in paramater

A2. Randomnes of Sample

A3. Conditional Independence

$$E(u | x_{1i}, x_{2i}, x_{3i} \ldots x_{ni}) = 0$$

A4. No perfect Multicolinearity (*)

A5. Homoskedasticity

A6. No large outliers