Ch: 3 Multiple Regression

y | indep. vars. $x_1, x_2, \ldots$ | u more likely to satisfy exogeneity

wage | educ. IQ exper.

crime | police unemp. graduation

k indep. vars. e.g. $k = 2$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

β_1 : Δy for $\Delta x_1 = 1$ $\quad \Delta x_2 = 0$

$\Delta u = 0$

Estimation

Ass^{n}s :

$$E(u) = 0$$

$$E(x_1 u) = 0$$

$$E(x_2 u) = 0$$

OLS estimates of β_0, β_1, and β_2 $\longrightarrow$ $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$

$$\frac{1}{n} \sum_{i=1}^{n} \boxed{\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right)} = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} x_{i1} \boxed{} = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} x_{i2} \boxed{} = 0$$

Example

y (wage)	x_1 (educ)	x_2 (exper)
3.1	11	2
3.2	12	22
3	11	2
6	8	44
5.3	12	7
8.8	16	9
11	18	15
5	12	5
3.6	12	26
18	17	22

$$
\begin{aligned}
\overline{y} - \widehat{\beta}_0 - \widehat{\beta}_1\overline{x_1} - \widehat{\beta}_2\overline{x_2} &= 0 \\
\overline{x_1 y} - \widehat{\beta}_0\overline{x_1} - \widehat{\beta}_1\overline{(x_1)^2} - \widehat{\beta}_2\overline{x_1 x_2} &= 0 \\
\overline{x_2 y} - \widehat{\beta}_0\overline{x_2} - \widehat{\beta}_1\overline{x_1 x_2} - \widehat{\beta}_2\overline{(x_2)^2} &= 0
\end{aligned}
$$

$$
\begin{aligned}
6.742 - \widehat{\beta}_0 - 12.9\widehat{\beta}_1 - 15.4\widehat{\beta}_2 &= 0 \\
97.234 - 12.9\widehat{\beta}_0 - 175.1\widehat{\beta}_1 - 190.4\widehat{\beta}_2 &= 0 \\
115.064 - 15.4\widehat{\beta}_0 - 190.4\widehat{\beta}_1 - 396.8\widehat{\beta}_2 &= 0
\end{aligned}
$$

$\widehat{\beta}_0 = -12.317 \qquad \widehat{\beta}_1 = 1.312 \qquad \widehat{\beta}_2 = 0.138$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \quad \text{(fitted value)}$$

$$\hat{u}_i = y_i - \hat{y}_i$$

Goodness of fit: $R^2 = \dfrac{SSE}{SST}$

$$= 1 - \frac{SSR}{SST}$$ non $\downarrow$ in k (# x's)

Adjusted R^2 : $\bar{R}^2 = 1 - \dfrac{\frac{SSR}{(n-k-1)}}{\frac{SST}{(n-1)}}$

As $k \uparrow$ $\quad SSR \downarrow \quad (n-k-1) \downarrow$

Expected value

unbiased OLS estimators $\qquad E(\hat{\beta}_j) = \beta_j \qquad j = 0, 1, \ldots, k$

(# x's)

under certain assumptions

- linear in parameters
- random sampling
- Variation in each x

  No linear rel^{n}ship among regressors.
- u unrelated to each x

  $E(u \mid x_1, \ldots, x_k) = E(u) = 0$

$$wage = \beta_0 + \beta_1\, age + \beta_2\, exper + \beta_3\, educ + \ldots + u$$

$$age = 6 + educ + exper$$

$$n \geq k + 1$$

Omitted Variable Bias

True model satisfying condⁿs for unbiasedness :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

wage educ. IQ

bwght smoking alcohol

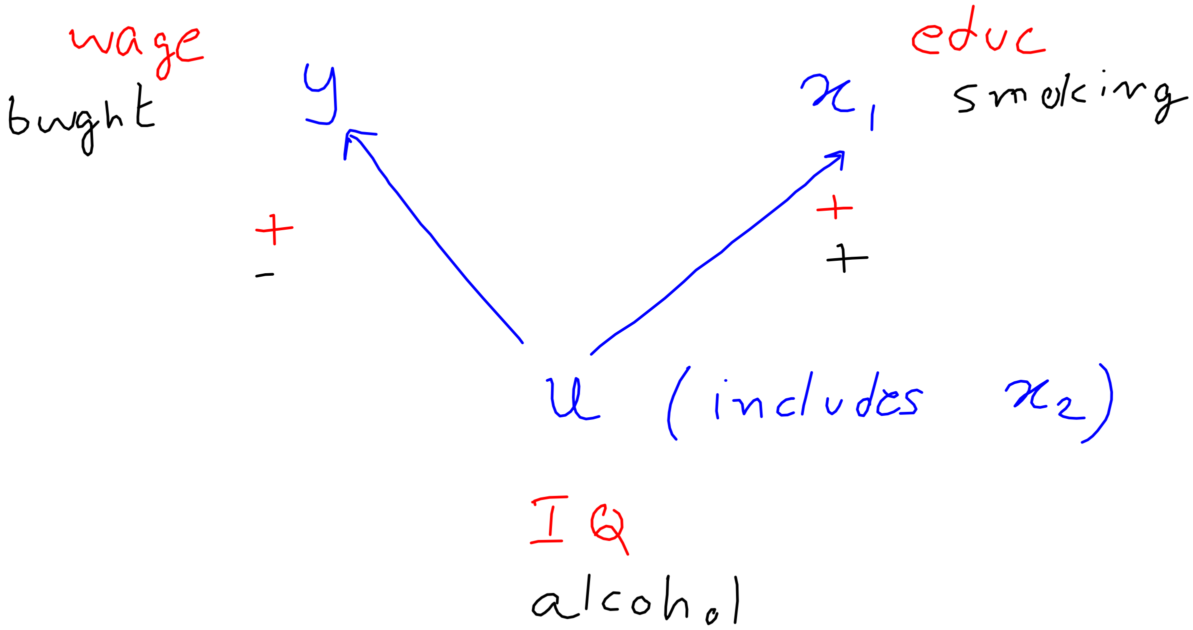
x_2 : omitted

estimate : $y = \beta_0 + \beta_1 x_1 + u$

obtain : $\tilde{\beta}_0$ and $\tilde{\beta}_1$ $\longrightarrow$ biased

$$E(\tilde{\beta}_j) \neq \beta_j \qquad j = 0, 1$$

Bias depends on β_2 and corr. b/w x_2 (omitted) and x_1 (included).



bwght

wage

Dep. var.

educ. smoking

indep. var.

+

−

+

+

u (includes omitted vars.) alcohol

Bias > 0

IQ

Bias < 0

Inclusion of irrelevant regressors :

- Exercise caution

y	x
# accidents	alcohol laws, " consumption, . . .
wage	educ, occupation, discrimination

Variance

$k = 3$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

y : price

x_1 : sq. ft.

x_2 : bdrms

x_3 : lot size

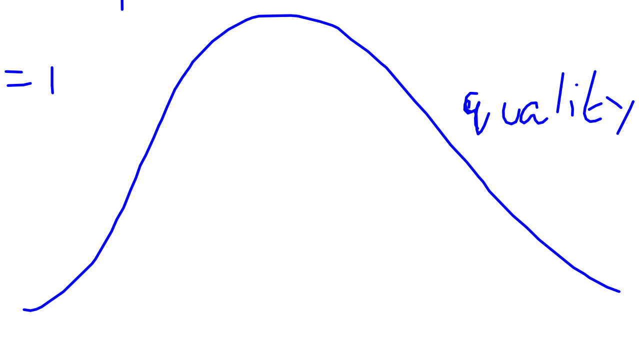
u : quality

Homoskedasticity : $Var(u \mid x_1, x_2, x_3) = \sigma^2$

$x_1 = 2000$

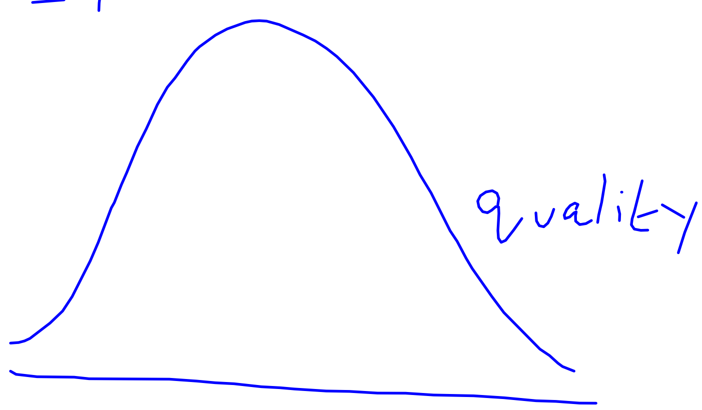
$x_2 = 4$

$x_3 = 1$



$x_1 = 1000$ $\quad x_3 = 0.5$

$x_2 = 1$



$$var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)} \qquad j = 1, 2, 3$$

$$SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

R_j^2 : R^2 from reg. of x_j on other x's

y : price

x_1 : sq.ft.

x_2 : bdrms

x_3 : lot size

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\text{Total variation in sq.ft} \left(1 - R^2 \text{ from reg. of sq.ft. on bdrms \& lot size}\right)}$$

R_j^2 close to 1 $\longrightarrow$ multicollinearity

does not violate assn's reqd. for unbiasedness

Under the assumptions up to homosk.

$\hat{\sigma}^2 \longrightarrow$ unbiased estimator of σ^2

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k-1} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-3-1}$$

$k \hookrightarrow$ # of regressors

$(k = 3)$

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j (1 - R_j^2)}}$$

$j = 1, 2, 3$