Lecture 4

Bivariate Regression Model (Cross-Section Data only)

$$y_i = \beta_0 + \beta_1 X_i + u_i \quad , \quad i = 1, 2, \cdots, N \qquad (SLR.1)$$

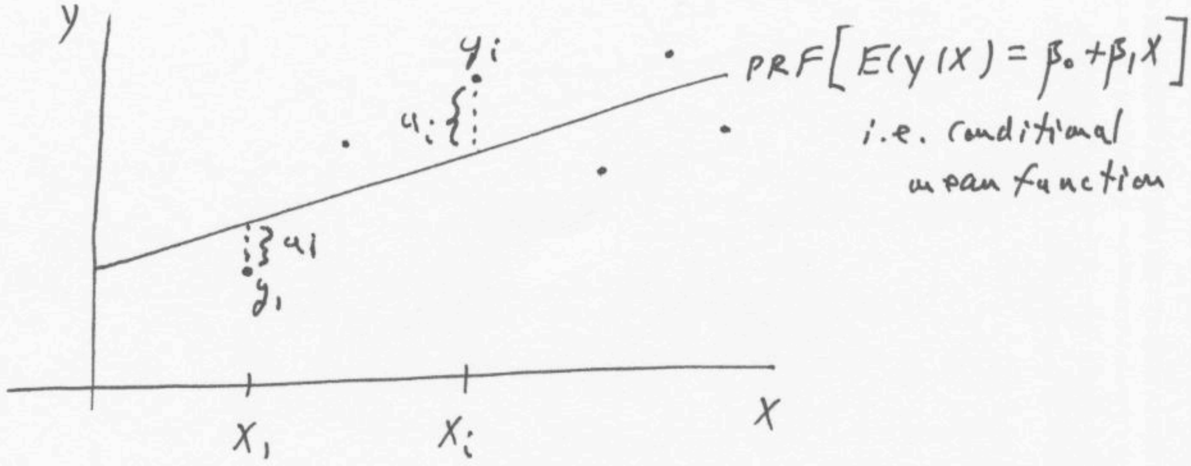
(X_i, y_i) $i = 1, 2, \cdots, N$ constitute a random sample from the population of (X, y) values (SLR.2)

$$E(u_i \mid X = X_i) = 0 \quad \text{for all } i \qquad (SLR.3)$$

Note: SLR.3 in conjunction with SLR.2 allows for a convenient technical simplification. In particular, we can derive the statistical properties of the OLS estimators as <u>conditional</u> on the values of the X_i in our sample. Technically, in statistical derivations, conditioning on the sample values of the independent (explanatory) variable X is the same as treating the X_i as <u>fixed in repeated samples</u>. (Woolaridge, p. 48)

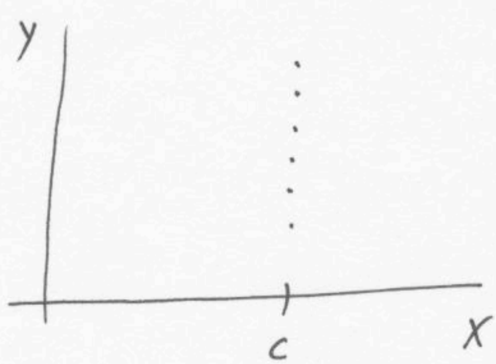
Also SLR.3 in conjunction with SLR.2 implies that the random errors u_i and u_j for two different individuals i and j are uncorrelated with each other. That is $E(u_i u_j) = 0$ for all i and j.

Graph of $y_i = \beta_0 + \beta_1 X_i + u_i$



There is some variation in the X_i, i.e. $X_i = c \;\; \forall i$ is not allowed. (SLR.4)

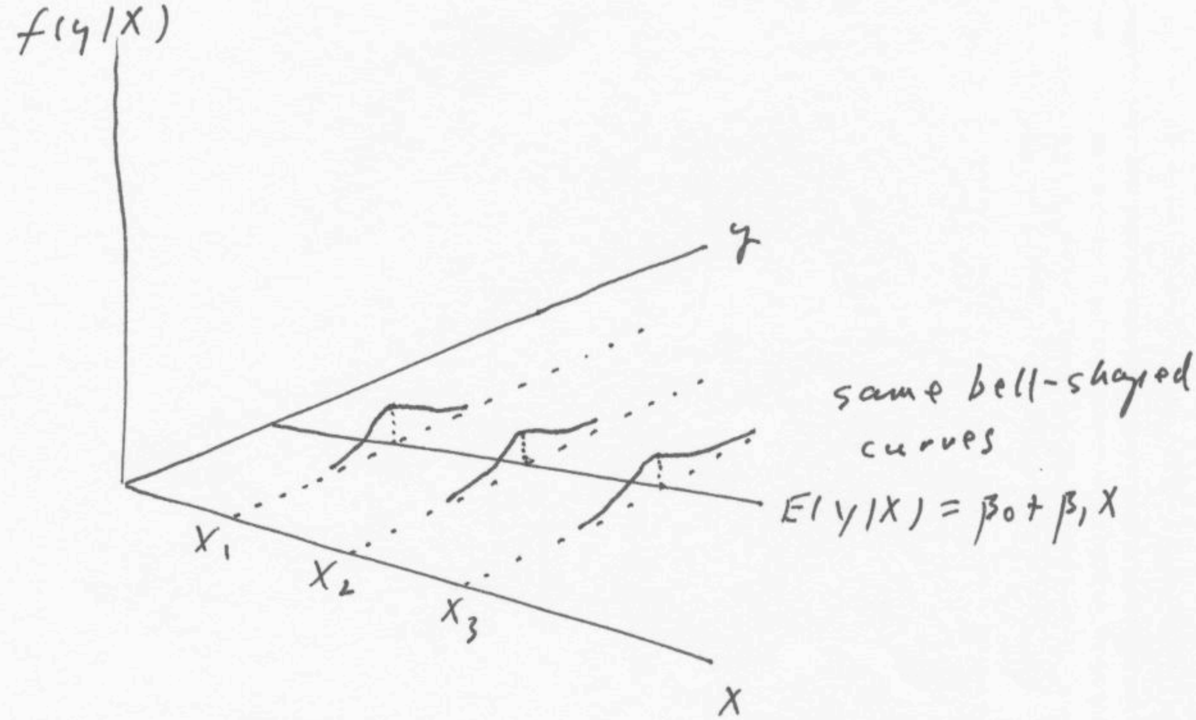
Disallowed



$$X_1 = X_2 = \cdots = X_N = c.$$

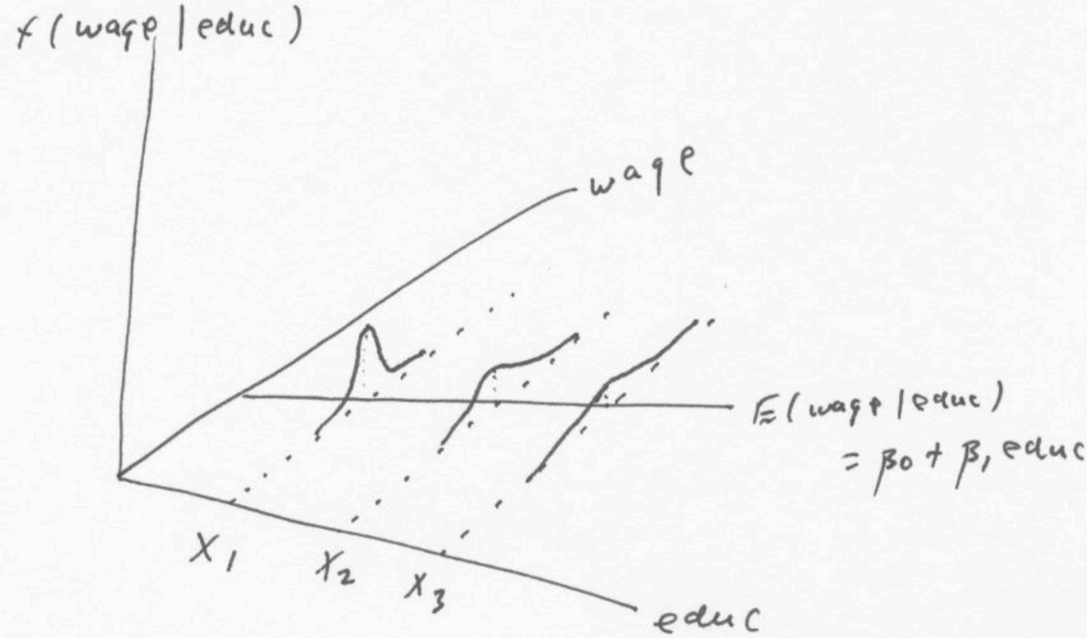
$Var(u_i | X = X_i) = \sigma^2$ for all i (SLR.5)

(Homoskedasticity)



Example of Heteroskedasticity in Wage equation

Var (wage | educ) increasing with education



$(u_i | X = X_i) \sim N(0, \sigma^2)$ (SLR.6)

Each error is normally distributed.

Note: This last assumption is needed when deriving the <u>small sample</u> sampling distributions of the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ and that of the t-statistics used in hypothesis tests of the β_0 and β_1.