2.1 Accounting Nonlinearity

! The regression model requires linearity in parameters.

! Similar to the Simple Regression Model, the Multiple Regression Model can also take into account the nonlinear relationships between variables.

Example: In the CEO Salary example, we could write the relations between CEO salary ($salary$), firm sales ($sales$) and CEO age (age) as follows:

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 age + \beta_3 age^2 + u$$

! - This model has $k = 3$ because there are 3 regressors. $Y = \log(salary), X_1 = \log(sales), X_2 = age, X_3 = age^2$.

- β_1 measures the change in $\log(salary)$ with respect to $\log(sales)$, holding other factors fixed. β_1 is the sales elasticity of CEO salary.

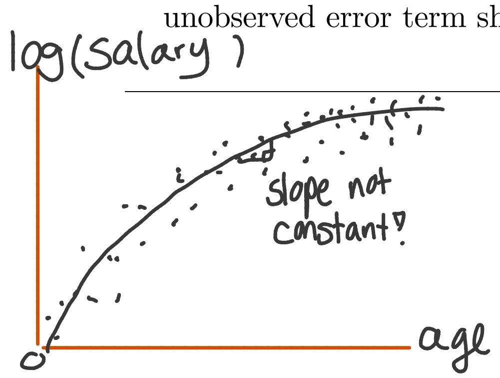
- How do we measure the change in $\log(salary)$ with respect to age, holding other factors fixed?

- How do we measure the change in $salary$ with respect to age, holding other factors fixed?

! In any case, the OLS estimates of β would be unbiased if

$$E(u"X_1, X_2, ..., X_k) = 0.$$

This is the Multiple Regression version of assumption SLR 4–all factors in the unobserved error term should be uncorrelated with the explanatory variables.



$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 age$

incorrect because relationship between log(salary) & age is not linear.

BUT! We can account for the non-linearity by adding (age^2)

$\log(salary) = \beta_0 + \beta_1 \underbrace{\log(sales)}_{X_1} + \beta_2 \underbrace{age}_{X_2} + \beta_3 \underbrace{age^2}_{X_3}$

This regression has 3 explanatory variables

$X_1 = \log(sales)$
$X_2 = age$
$X_3 = age^2$
$Y = \log(salary)$

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 age + \beta_3 age^2 + u$$

Find $\frac{d\, salary}{d\, age}$ → from $\frac{d \log(salary)}{d\, age} = \beta_2 + 2\beta_3 age$

The impact of 1-year age increase on salary.

In econometrics
$\log(x) \approx \ln(x)$
So, $\frac{d \log x}{dx} = \frac{1}{x}$
$d \log x = \frac{1}{x} dx$

$$\frac{\frac{1}{salary} d\, salary}{d\, age} = \beta_2 + 2\beta_3 age$$

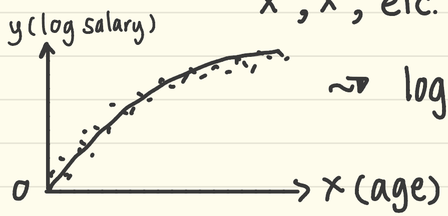
$$\frac{d\, salary}{d\, age} = salary(\beta_2 + 2\beta_3 age)$$

- The impact of a 1-year age increase is not constant. It depends on the salary level and age of each worker!

☆ Regarding the assumption SLR 1 - Linear in parameter (β). What should we do if the relationship between X & Y is non-linear?

Option 1: transform variable(s) Y and X (See page 44)

Option 2: add polynomial terms of X variables, e.g. x^2, x^3, etc. The most popular term to add is x^2.



$$\rightarrow \log(salary) = \beta_0 + \beta_1 \log(sale) + \beta_2 age + \beta_3 age^2$$
β_2: (+), β_3: (−)

No one knows which option is correct because the true population function is not observed. But we can compare R^2 of different models (using option 1 vs. option 2). The model with higher R^2 is preferred.