The Simple Regression Model

- Some questions

  - How often (Y) would a person like to go eat at the Fuji restaurant if the price (X) drops?

  - How many more cars (Y) would be on the street in Bangkok if the government waives car tax (X)?

  - How much would my income (Y) increase if I obtain a master's degree (X)?

- The concept

  - How would you measure the "change in the number of new cars" when the "tax rate" changes.

  - What would be the "change in the number of new cars" when the "tax rate" is reduced by 1 percent?

$$\beta_1 = \frac{\Delta \text{ new cars}}{\Delta \text{ tax rate}}$$

- Or, using calculus, you can get

$$\frac{dY_i}{dX_i} = \beta_1.$$

- ** The linear relationship between Y and X is assumed.

1 Principle, assumptions and derivation of ordinary least squares (OLS) estimators

1.1 Terminology for the Linear Regression

- The simple regression model can be defined as

TABLE 2.1

Terminology for Simple Regression

y	x
Dependent variable	Independent variable
Explained variable	Explanatory variable
Response variable	Control variable
Predicted variable	Predictor variable
Regressand	Regressor

- Y_i and X_i are variables
- β_0 and β_1 are parameters

1.2 Derivation of Ordinary Least Squares (OLS) Estimators

- If we do not have the data point of the entire population, but rather on a subset of samples, what should we do to derive β_0 and β_1?

  Example:

  How often would a TU undergraduate student go eat at Fuji restaurant (Y) if he/she receive an (X) percentage discount in price?

TABLE 3.1. Number if visits to Fuji Restaurant per year and percentage discount

i	Y_i = visit to Fuji restaurant (times/year)	X_i = percentage discount in price
1	95	78
2	42	55
3	56	67
4	83	70
....	...	...
100	32	46

Source: data collected from a random survey of 100 TU undergraduate students

- We surveyed 100 TU undergraduate students (sample)
- We hope that we can use this data to explain the frequency of TU undergraduate students' (population) visit to Fuji restaurant given the price discount.

- The relation of the frequency of TU undergraduate students' (population) visit to Fuji restaurant (Y) given the price discount (X) can be expressed as:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- But since we don't have the data of the entire population and we don't know the true value of β_0 and β_1, we need to find estimators of β_0 and β_1.
- The estimators are often called $\hat{\beta}_0$ and $\hat{\beta}_1$. (like $\bar{X}$ is an estimator of μ and S^2 is an estimator of σ^2).

  How do we estimate $\hat{\beta}_0$ and $\hat{\beta}_1$?
- The OLS suggests that we can find $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimizes the sum of squared errors (deviation from the regression line).

- Mathematical Derivation of OLS (minimization the sum of squared errors)

$$\underset{\hat{\beta}_0,\hat{\beta}_1}{\arg\min} \sum_{i=1}^{n} (\hat{u}_i^2) = \underset{\hat{\beta}_0,\hat{\beta}_1}{\arg\min} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

First Order Condition (F.O.C):

$$w.r.t.\hat{\beta}_0 \quad \Rightarrow \quad 0 = -2(\Sigma_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)) \tag{3.1}$$

$$w.r.t.\hat{\beta}_1 \quad \Rightarrow \quad 0 = -2(\Sigma_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)) \tag{3.2}$$

Divide 3.1 by -2, we get

2 Properties of OLS estimators

2.1 Algebraic Properties

- Thus far, the OLS estimators are

$$\begin{aligned}\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\ \hat{\beta}_1 &= \frac{\Sigma_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\Sigma_{i=1}^n (X_i - \bar{X})^2}\end{aligned}$$

  this gives

$$\begin{aligned}\hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i \\ \hat{u}_i &= Y_i - \hat{Y}_i.\end{aligned}$$

  This implies the following algebriac properties

1. $\Sigma_{i=1}^n \hat{u}_i = 0$ – the calculation of OLS $\hat{\beta}_0$ and $\hat{\beta}_1$ is done such that it minimizes $\Sigma_{i=1}^n \hat{u}_i^2$. This is true when $\Sigma_{i=1}^n \hat{u}_i = 0$
2. $\Sigma_{i=1}^n X_i \hat{u}_i = 0$ – we get this from the first order condition deriving $\hat{\beta}_1$. This implies that $Cov(X_i, \hat{u}_i) = 0$. (Why?)
3. The point $\bar{Y}$ and $\bar{X}$ are always on the regression line – we know this from $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$.

2.2 Properties proving BLUE

- The OLS estimator is the Best Linear Unbiased Estimator (BLUE). How?
- The OLS' $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1 respectively

$$\begin{aligned} E(\hat{\beta}_0) &= \beta_0 \text{ and } E(\hat{\beta}_1) = \beta_1. \\ \text{or } E(\hat{\beta}_{OLS}) &= \beta. \end{aligned}$$

- The OLS' $\hat{\beta}_0$ and $\hat{\beta}_1$ are the most efficient among all the linear estimators

$$Var(\tilde{\beta}_{non-OLS}|\ X) - Var(\hat{\beta}_{OLS}|\ X) \geq 0.$$

2.3 Assumptions on the simple linear regression (SLR) model

- Some assumptions (or certain conditions) are required for the OLS to be BLUE. This set of assumptions are often called the "Gauss-Markov Assuptions for Simple Regression"

  Assumption SLR1. Linear in Parameter – Y is linear in X.

  Assumption SLR2. Random Sampling – We have a random sample size n, $\{(x_i, y_i) : i = 1, ..., n\}$, This, the model _____ becomes:

  Assumption SLR3. Sample Variation in the Explanatory Variable – The sample outcomes are not all the same value

  Assumption SLR4. Zero Conditional Mean –

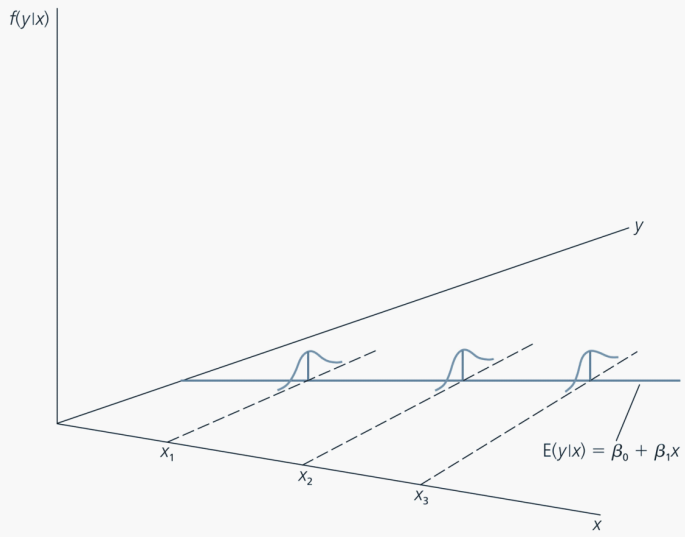
  Assumption SLR5. Homoskedasticity –

- In other words, The OLS estimator of β in the linear model when u_i is $i.i.d.(0, \sigma^2)$ is the best (minimum variance) estimator within the class of linear unbiased estimator.
- The conditional concept, which implies that X_i is predetermined (being conditional upon, or fixed) is very crucial for the OLS estimators to be unbiased.
- We must not forget that there is no reason that all these assumptions should be true!

2.4 *Homoskedasticity VS. Heteroskedasticity*

FIGURE 2.8

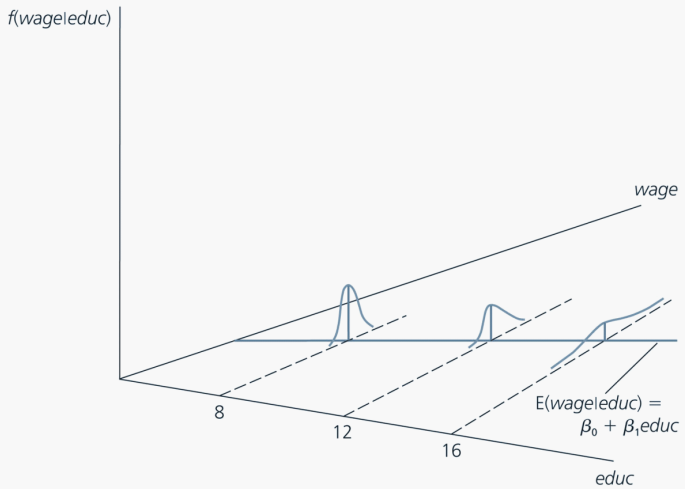
The simple regression model under homoskedasticity.



Homoskedasticity

FIGURE 2.9

Var(*wage*|*educ*) increasing with *educ*.



Heteroskedasticity

2.5 Expectation of Estimators

- Proof for $E(\hat{\beta}_1) = \beta_1$: We need to use assumption SLR 1,2,3,4.

  From

$$\hat{\beta}_1 = \frac{\Sigma_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\Sigma_{i=1}^{n}(X_i - \bar{X})^2} \tag{3.3}$$

  For calculation tractability, let

2.6 Variance of OLS Estimators

- From eq._____, we can write

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + (\Sigma_{i=1}^n u_i k_i) \\ Var(\hat{\beta}_1) &= Var(\beta_1) + Var(\Sigma_{i=1}^n u_i k_i) \end{aligned}$$

  Here, β_1 (the true β_1) is a constant. And since we are conditioning on X_i, the values of k_i are also non-random. (This is not to be confused with the random sampling of X_i, $i = 1, ..., n$.) When you are conditioning on some variables, you take those variables as non-random (or as given). In this case, we can write

The prove that $Var(\hat{\beta}_1) = \frac{\sigma^2}{SST}$ is the minimum under the class of "linear unbiased estimator" is complicated without relying on some metrix simplifications. Therefore, we will do this proof using the metrix notation.

2.7 Some Concepts to be exphasized (Population vs. Sample; PRF vs. SRF; error vs. residual)

- Population – is the "truth" and in econometrics, we believe that there is one set of "truth". Therefore, the Population Regression Function (PRF) is fixed. In almost all cases, we do not know "exactly" what PRF is.

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{3.4}$$

  where $i = 1, 2, ..., \Psi$. Ψ represents the total number of the "population" we are interested in.

- Sample – is a subset of "truth", a subset of "population". We run, or construct, the Sample Regression Function (SRF) to estimate the PRF.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

  where $i = 1, 2, ..., n$, $n < \Psi$.

- The true "error term" u_i in eq.3.4 is **unobserved** because we never know what β_0 and β_1 is.

- We can, however, observe the "residual" or "predicted residual" $\hat{u}_i$ from the following calculation

$$Y_i - \hat{Y}_i = \hat{u}_i.$$

Examples of some simple linear regressions:

1. $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ where $\beta_0 = 120, \beta_1 = -9.8$, $Y_i =$ quantity, $X_i =$ price.

2. $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ where $\hat{\beta}_0 = 164, \hat{\beta}_1 = 0.27, Y_i =$ housing expenditure, $X_i =$ monthly income. Here, $\hat{\beta}_1$ is equivalent to the marginal propensity to consume housing out of income.

2.8 Goodness of Fit (R^2)

- We can never measure how well SRF can estimate PRF because PRF is not known to allow us to compare.
- But we can assess how well different SRFs fit with the "sample" data that we collect – this can also be called the "goodness of fit".

Residual Concepts:-

$$\begin{aligned} \text{Total Sum of Squares } (SST) &= \\ \text{Explained Sum of Squares } (SSE) &= \\ \text{Residual Sum of Squares } (SSR) &= \\ SST &= \end{aligned}$$

- The R^2, or the coefficient of determination, is defined as

2.9 Incorporating Nonlinearities in Sample Regression

- So far, we have only mentioned the "linear" relation.
- OLS can actually incorporate non-linearities as long as the non-linearities is in the "variables".
- For examples:

$$\log Y_i = \beta_0 + \beta_1 \log X_i + u_i \quad \text{(This is called a constant elasticity model. Why?)}$$

or

$$\log Y_i = \beta_0 + \beta_1 X_i^2 + u_i$$

or

$$Y_i = \beta_0 + \beta_1 \frac{1}{X_i} + u_i.$$

etc. etc...

---

3 Regression Through the Origin

- If you know that the regression line goes through the origin ($X = 0, Y = 0$) for sure, we can impose this restriction
- This restriction makes sense in some contexts. For example, $X = income$ and $Y = income\ tax$.
- The SRF then becomes

$$Y_i = \tilde{\beta}_1 X_i$$

where $\tilde{\beta}_1$ is an OLS estimator. It can be calculated through finding $\tilde{\beta}_1$ that minimizes the sum of squared residual