The Simple Regression Model

$y = \beta_0 + \beta_1 x + u$

Ch.2 The simple regression model

1. Definition of the simple regression model
2. Deriving the OLS estimates
3. Mechanics of OLS
4. Units of measurement & functional form
5. Expected values & variances of OLSE
6. Regression through the origin

2.1 Definition of the model

- Equation (2.1), $y = \beta_0 + \beta_1 x + u$, defines the *Simple Regression model*.
- In the model, we typically refer to
  - y as the Dependent Variable
  - x as the Independent Variable
  - βs as parameters, and
  - u as the error term.

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

The Concept of Error Term

- u represents factors other than x that affect y.
- If the other factors in u are held fixed, so that $\Delta u = 0$, then $\Delta y = \beta_1 \Delta x$.
- Ex. 2.1: $yield = \beta_0 + \beta_1 fertilizer + u$ (2.3)
  - u includes land quality, rainfall, etc.
- Ex. 2.2: $wage = \beta_0 + \beta_1 educ + u$ (2.4)
  - u includes experience, ability, tenure, etc.

A Simple Assumption for *u*

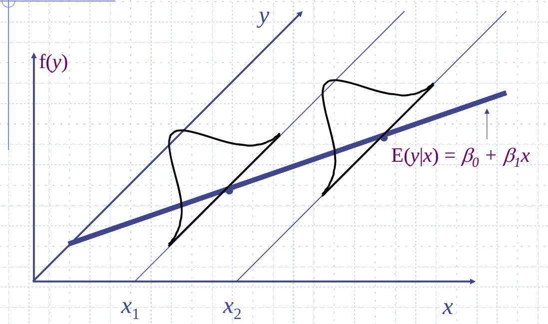
- The average value of u, the error term, in the population is 0. That is, $E(u) = 0$.
  - This is not a restrictive assumption, since we can always use β_0 to normalize $E(u)$ to 0.
- To draw ceteris paribus conclusions about how x affects y, we have to hold all other factors (in u) fixed.

Zero Conditional Mean

- We need to make a crucial assumption about how u and x are related.
- We want it to be the case that knowing something about x does not give us any information about u, so that they are completely unrelated. That is, that
  - $E(u|x) = E(u) = 0$ (2.5&2.6), which implies
  - $E(y|x) = \beta_0 + \beta_1 x$ **(PRF)** (2.8)

Special Lecture 7

E(y|x) as a linear function of x, where for any x the distribution of y is centered about E(y|x)



Special Lecture 8

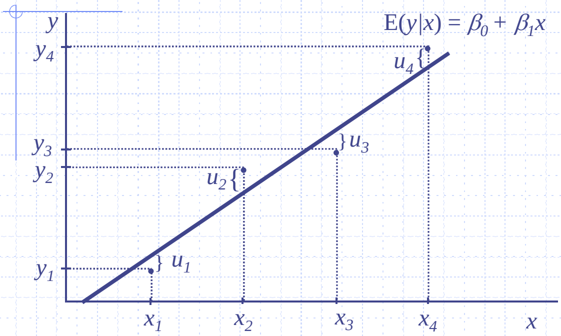
2.2 Deriving the OLSE

- Basic idea of regression is to estimate the ***population parameters*** from a ***sample***.
- Let $\{(x_i,y_i): i = 1, \ldots, n\}$ denote a random sample of size n from the population.
- For each observation in this sample, it will be the case that

$$y_i = \beta_0 + \beta_1 x_i + u_i. \quad (2.9)$$

Special Lecture 9

Population regression line, sample data points and the associated error terms



Special Lecture 10

Deriving OLSE using MM

- To derive the OLS estimates, we need to realize that our main assumption of
  - $E(u|x) = E(u) = 0$ also implies that
  - $Cov(x,u) = E(xu) = 0$
    - Because Cov(X,Y) = E(XY) – E(X)E(Y) (B.27)
- Now we prepare 2 restrictions to estimate βs.
  - $E(u) = 0$ (2.10)
  - $E(xu) = 0$ (2.11)

Special Lecture 11

Cont. Deriving OLSE using MM

- Since $u = y - \beta_0 - \beta_1 x$, we can rewrite;

$$E(u) = E(y - \beta_0 - \beta_1 x) = 0 \quad (2.12)$$

$$E(xu) = E[x(y - \beta_0 - \beta_1 x)] = 0 \quad (2.13)$$

- These are called moment restrictions
  - The approach to estimation implies imposing the population moment restrictions on the sample moments. It means, a sample estimator of E(X), the mean of a population distribution, is simply the arithmetic mean of the sample.

Special Lecture 12

More Derivation of OLS

- We want to choose values of the parameters that will ensure that the sample versions of our moment restrictions are true
- The sample versions are as follows:

$$n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \quad (2.14)$$

$$n^{-1}\sum_{i=1}^{n}x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \quad (2.15)$$

Special Lecture 13

Cont. More Derivation of OLS

- Given the definition of a sample mean, and properties of summation, we can rewrite the first condition as follows

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad (2.16) \quad \text{or} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (2.17)$$

- So the OLS estimated slope is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (2.19)$$

Special Lecture 14

Summary of OLS slope estimate

- The slope estimate is the sample covariance between x and y divided by the sample variance of x.
- If x and y are positively (negatively) correlated, the slope will be positive (negative).
- x needs to vary in our sample.
  - See (2.18) & Figure (2.3)

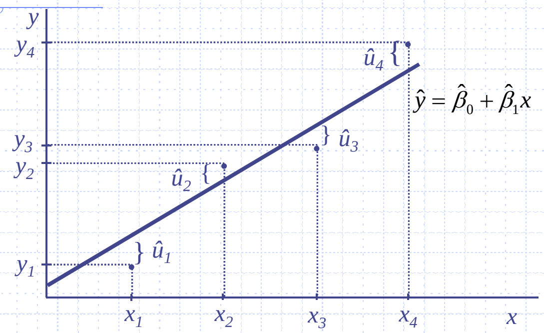
Special Lecture 15

More OLS

- Intuitively, OLS is fitting a line through the sample points such that the ***sum of squared residuals*** is as small as possible, hence the term is called least squares.
- The ***residual***, $\hat{u}$, is an estimate of the error term, u, and is the difference between the fitted line (sample regression function) and the sample point.

Special Lecture 16

Sample regression line, sample data points and the associated estimated error terms



Special Lecture 17

Alternate approach to derivation

- Given the intuitive idea of fitting a line, we can set up a formal minimization problem.

$$\sum_{i=1}^{n}(\hat{u}_i)^2 = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 \quad (2.22)$$

- The first order conditions, which are the almost same as (2.14) & (2.15),

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0, \quad \sum_{i=1}^{n}x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

Special Lecture 18

2.3 Properties of OLS

Algebraic Properties of OLS

1. The sum of the OLS residuals is zero. Thus, the sample average of the OLS residuals is zero as well.

$$\sum_{i=1}^{n} \hat{u}_i = 0 \text{ and thus, } \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0 \quad (2.30)$$

Cont. Algebraic Properties

2. The sample covariance between the regressors and the OLS residuals is zero

$$\sum_{i=1}^{n} x_i \hat{u}_i = 0 \quad (2.31)$$

3. The OLS regression line always goes through the mean of the sample

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Cont. Algebraic Properties

We can think of each observation as being made up of an explained part, and an unexplained part, $y_i = \hat{y}_i + \hat{u}_i$ (2.32) Then we define the following:

$$\sum (y_i - \bar{y})^2 \equiv \text{SST} \quad (2.33)$$

$$\sum (\hat{y}_i - \bar{y})^2 \equiv \text{SSE} \quad (2.34)$$

$$\sum \hat{u}_i^2 \equiv \text{SSR} \quad (2.35)$$

Then, $\text{SST} = \text{SSE} + \text{SSR} \quad (2.36)$

Goodness-of-Fit

- It's useful we think about how well the sample regression line fits sample data.
- From (2.36),

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \quad (2.38).$$

- R^2 indicates the fraction of the sample variation in y_i that is explained by the model.

2.4 Measurement Units & Function Form

- If we use the model $y^* = \beta_0^* + \beta_1^* x^* + u^*$ instead of $y = \beta_0 + \beta_1 x + u$, we get

$$\hat{\beta}_0^* = c\hat{\beta}_0 \text{ and } \hat{\beta}_1^* = \frac{c}{d}\hat{\beta}_1$$

where $y^* = c\, y$ and $x^* = d\, x$. Similarly,

$$\hat{\beta}_1^* = \frac{\partial y}{\partial x} \cdot \frac{x}{y} \Leftrightarrow \hat{\beta}_1 = \frac{\partial y}{\partial x}$$

where $y^* = \ln y$ and $x^* = \ln x$.

TABLE 2.3

Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

2.5 Means & Variance of OLSE

- Now, we view $\hat{\beta}_i$ as estimators for the parameters β_i that appears in the population, which means properties of the distributions of $\hat{\beta}_i$ over different random samples from the population.

Unbiasedness of OLS

- **Unbiased estimator:** An estimator whose expected value (or mean of its sampling distribution) equals the population value (regardless of the population value).

Special Lecture 25

Cont. Unbiasedness of OLS

Assumption for unbiasedness

1. Linear in parameters as $y = \beta_0 + \beta_1 x + u$
2. Random sampling $\{(x_i, y_i): i = 1, 2, \ldots, n\}$,
   Thus, $y_i = \beta_0 + \beta_1 x_i + u_i$
3. Sample variation in the x_i, thus
   $$\sum (x_i - \bar{x})^2 > 0$$
4. Zero conditional mean, $E(u|x) = 0$

Special Lecture 26

Cont. Unbiasedness of OLS

- In order to think about unbiasedness, we need to rewrite our estimator in terms of the population parameter.

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \beta_1 + \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2} \quad (2.49), (2.52)$$

$$\text{then } E\left(\hat{\beta}_1\right) = \beta_1 + \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \cdot E(u_i \mid x) = \beta_1 \quad (2.53)$$

* we can also get $E(\hat{\beta}_0) = \beta_0$ in the same way.

Special Lecture 27

Unbiasedness Summary

- The OLS estimates of β_1 and β_0 are unbiased.
- Proof of unbiasedness depends on our 4 assumptions – if any assumption fails, then OLS is not necessarily unbiased.
- Remember unbiasedness is a description of the ***estimator*** – in a given sample our ***estimate*** may be "near" or "far" from the true parameter.

Special Lecture 28

Variances of the OLS Estimators

- Now we know that the sampling distribution of our estimate is centered around the true parameter.
  - We want to think about how spread out this distribution is.
  - It is much easier to think about this variance under an additional assumption, so assume

5. $\text{Var}(u|x) = \sigma^2$ (Homoskedasticity)

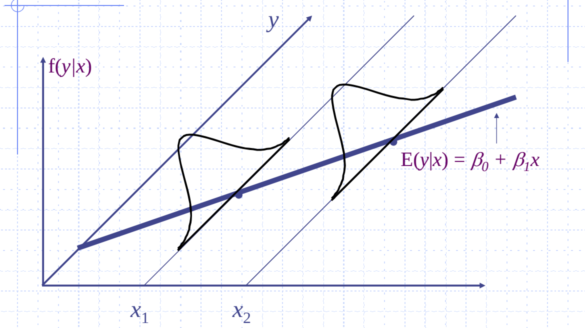
Special Lecture 29

Cont. Variance of OLSE

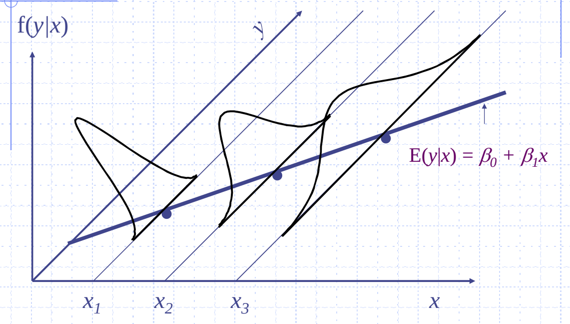
- σ^2 is also the unconditional variance, called the ***error variance***, since
  - $\text{Var}(u|x) = E(u^2|x) - [E(u|x)]^2$
  - $E(u|x) = 0$, so $\sigma^2 = E(u^2|x) = E(u^2) = \text{Var}(u)$
  - And σ, the square root of the error variance, is called the ***standard deviation of the error***.
- Then we can say
  $E(y|x) = \beta_0 + \beta_1 x$ and $\text{Var}(y|x) = \sigma^2$

Special Lecture 30

Homoskedastic Case



Heteroskedastic Case



Cont. Variance of OLSE

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \quad (2.57)$$

- The larger the error variance, σ^2, the larger the variance of the slope estimate.
- The larger the variability in the x_i, the smaller the variance of the slope estimate.
- As a result, a larger sample size should decrease the variance of the slope estimate.

Estimating the Error Variance

- We don't know what is the error variance, σ^2, because we don't observe the errors, u_i.
- What we observe are only the residuals, $\hat{u}_i$, not the errors, u_i.
- So we can use the residuals to form an estimate of the error variance.

Cont. Error Variance Estimate

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$
$$= (\beta_0 + \beta_1 x_i + u_i) - \hat{\beta}_0 - \hat{\beta}_1 x_i$$
$$= u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i$$

Then, an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{(n-2)} \sum \hat{u}_i^2 \quad (2.61)$$

Cont. Error Variance Estimate

$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ = Standard error of the regression

recall that s.d.$(\hat{\beta}) = \sqrt{Var(\hat{\beta})}$

If we substitute $\hat{\sigma}$ for σ, then we have

the standard error of $\hat{\beta}_1$,

$$\text{se}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}$$

2.6 Regression through the Origin

- Now, consider the model without a intercept:

$$\tilde{y} = \tilde{\beta}_1 x \quad (2.63).$$

- Solving the FOC to the minimization problem, OLS estimated slope is

$$\tilde{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} \quad (2.66).$$

* Recall that a intercept can always normalize E(u) to 0 in the model with β_0.

Special Lecture 37