Multiple Regression Analysis: Estimation. Wooldridge (2013), Chapter 3

- Ordinary Least Squares (OLS) Estimator
- Interpreting Multiple Regression
- A “Partialling Out” Interpretation of the OLS estimator - Frisch-Waugh (1933) Theorem
- Simple vs Multiple Regression Estimate
- The R-squared
- Unbiasedness of the OLS estimator
- Too Many or Too Few Variables
- Variance of the OLS Estimators
- The Gauss-Markov Theorem
- Variance of the OLS Estimators - Misspecified Models
- Estimating the Error Variance
- Incorporating Non-linearities

Multiple Regression Analysis

The Multiple Regression model takes the form

$$E\left(y|x_1, ..., x_k\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

or equivalently

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$

where $E\left(u|x_1, ..., x_k\right) = 0$.

Parallels with Simple Regression:

- y is the dependent variable (regressand).
- $x_1, ..., x_k$ are the k regressors.
- u is still the error term (or disturbance).
- β_0 is still the intercept.
- β_1 to β_k all called slope parameters.

Multiple Regression Analysis

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u,$$

where $E\left(u | x_1, ..., x_k\right) = 0$.

Examples:

- $y-$sales, the regressors are advertising expenditure, income, price relative to competitors.
- $y-$ personal consumption, the regressors are disposable income, wealth, interest rates.
- $y-$ Investment, the regressors are interest rates and profits (past and future).
- $y-$ Wages, the regressors are schooling, experience, ability and gender.

To estimate $\beta_0, \beta_1, \beta_2, ..., \beta_k$ we choose $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ that minimize

$$\mathcal{S}\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik}\right)^2$$

The first order conditions are

$$\begin{aligned} -\frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik}) &= 0 \\ -\frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik})x_{ij} &= 0 \\ j &= 1, ..., k \end{aligned}$$

- This is a system of equations with $k+1$ equations and $k+1$ variables: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$. The Ordinary Least Squares estimator is obtained by solving the system of equations for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$.

The first order conditions can be written as

$$-\frac{2}{n}\sum_{i=1}^{n}\hat{u}_i = 0, \quad (1)$$

$$-\frac{2}{n}\sum_{i=1}^{n}\hat{u}_i x_{ij} = 0, \quad (2)$$

$$j = 1, ..., k,$$

where $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - ... - \hat{\beta}_k x_{ik}$.(*residuals*)

Remarks:

- Beyond the two-variable case it is not possible to write out an explicit formula for the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ (without the use of matrix algebra), although a solution exists.
- Equation (1) implies that the sum and the mean of the residuals are zero.
- Equations (1) and (2) imply that the covariances between the residuals and each regressor are zero.

The OLS regression line (*fitted values*) is now defined as

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k.$$

Writing it in terms of changes we obtain

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + ... + \hat{\beta}_k \Delta x_k.$$

Holding x_i ,$i = 1, ...k$ and $i \neq j$ fixed implies that

$$\Delta\hat{y} = \hat{\beta}_j \Delta x_j,$$

$j = 1, ..., k$. Thus each β has a ceteris paribus interpretation.

- In most cases, we will indicate the estimation of a relationship through OLS by writing as

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k. \tag{3}$$

- Sometimes, for the sake of brevity, it is useful to indicate that an OLS regression has been run without actually writing out the equation.
- We will often indicate that equation (3) has been obtained by OLS in saying that we run the regression of y on $x_1, x_2, ..., x_k$

- Regression of Wages on years of Education and years of Work Experience:

Dependent variable: Wages

Estimation Method: Ordinary Least Squares

Regressors	Estimates
Intercept	−5.56732
Education	0.97685
Experience	0.10367

- Another year of Education is predicted to increase the mean of wages by \$0.97685, holding Experience fixed.
- Another year of Experience is predicted to increase the mean of wages by \$0.10367, holding Education fixed.

Consider the case $k = 2$, i.e.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

There is an interesting interpretation for $\hat{\beta}_1$:

- Let $\hat{r}_{i1}$ be the residuals from the regression of x_1 on x_2. The fitted values are $\hat{x}_1 = \hat{\gamma}_0 + \hat{\gamma}_2 x_2$.
- Notice that for $i = 1, ..., n$

$$x_{i1} = \underbrace{\hat{x}_{i1}}_{\substack{\text{part of } x_1 \text{ that can} \\ \text{be explained by } x_2}} + \underbrace{\hat{r}_{i1}}_{\substack{\text{part of } x_1 \text{ that cannot} \\ \text{be explained by } x_2}}$$

- It can be shown that the OLS estimator for β_1, $\hat{\beta}_1$, is equal to the estimator of the slope when we run a regression of y_i on $\hat{r}_{i1}$. That is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}.$$

- It can be shown that the OLS estimator for β_1, $\hat{\beta}_1$, is equal to the estimator of the slope when we run a regression of y_i on $\hat{r}_{i1}$. That is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$

- What is the interpretation of this?
- We're estimating the effect of x_1 on y after removing from x_1 the effect of x_2.

Compare the simple regression

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

with the multiple regression

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

Generally $\tilde{\beta}_1 \neq \hat{\beta}_1$ unless $\hat{\beta}_2 = 0$ (i.e. no partial effect of x_2) or x_1 and x_2 are uncorrelated in the sample.

Example:

- Regression of Wages on Education

Dependent valiable: Wages

Estimation Method: Ordinary Least Squares, sample size: 528

Regressors	Estimates
Intercept	-1.60468
Education	0.81395

- Regression of Wages on Education and Experience

Dependent valiable: Wages

Estimation Method: Ordinary Least Squares, sample size: 528

Regressors	Estimates
Intercept	-5.56732
Education	0.97685
Experience	0.10367

As in the simple regression model we can think of each observation as being made up of an explained part, and an unexplained part, $y_i = \hat{y}_i + \hat{u}_i$.
We then define the following:

- $\sum_{i=1}^{n} (y_i - \bar{y})^2$ is the *total sum of squares* (SST).
- $\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ is the *explained sum of squares* (SSE).
- $\sum_{i=1}^{n} \hat{u}_i^2$ is the *residual sum of squares* (SSR).

(Same definitions as in the linear regression model)
Then

$$SST = SSE + SSR.$$

Prove this result in the simple regression model!

- Can compute the fraction of the total sum of squares (SST) that is explained by the model, call this the *R-squared* of regression:

  $$R^2 = SSE/SST = 1 - SSR/SST,$$

  where

  $$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2, \quad SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2, \quad SSR = \sum_{i=1}^{n} \hat{u}_i^2.$$

- R^2 is a measure of *Goodness of fit*: proportion of the variance of the dependent variable that is explained by the model.
- The R^2 is called the *coefficient of determination*.
- $0 \leq R^2 \leq 1$.

It can be shown that R^2 is equal to the squares of the correlation between $\hat{y}$ and y

$$R^2 = \frac{\left[\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}}) (y_i - \bar{y})\right]^2}{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

(also valid for the simple regression model). It can also be shown that $\bar{\hat{y}} = \bar{y}$.

- R^2 can never decrease when another independent variable is added to a regression, and usually will increase.
- Because R^2 will usually increase with the number of independent variables, it is not a good way to compare models.
- An alternative measure usually reported by any statistical software is the *adjusted R-squared*:

$$
\begin{aligned}
\bar{R}^2 &= 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} \\
&= 1 - \frac{(n-1)}{(n-k-1)}(1-R^2).
\end{aligned}
$$

- $\bar{R}^2$ penalizes the number of regressors included.
- However, $\bar{R}^2$, is not not between 0 and 1. In fact, it can be negative.

Example: Regression of Wages on Education and Experience

Dependent valiable: Wages

Estimation Method: Ordinary Least Squares, sample size: 528

Regressors	Estimates
Intercept	-5.56732
Education	0.97685
Experience	0.10367

$$R^2 = 0.209,\ \bar{R}^2 = 0.206$$

- Population model is linear in parameters: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$.
- We can use a random sample of size $n, \{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : i = 1, 2, \ldots, n\}$, from the population model, so that the sample model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i.$$

- $E(u|x_1, x_2, \ldots\ x_k) = 0$, implying that all of the explanatory variables are exogenous.
- None of the x's is constant, and there are no exact linear relationships among them (no perfect *multicolinearity*).

Proposition

Under the above assumptions the OLS estimators for β_0, β_1, ...β_k are unbiased, that is

$$E\left(\hat{\beta}_j\right) = \beta_j,\ j = 1, ..., k.$$

(prove this result in the simple regression model).

- What happens if we include variables in our specification that don't belong?
- There is no effect on our parameter estimate, and OLS remains unbiased.
- What if we exclude a variable from our specification that does belong?
- OLS will usually be biased.

Suppose that we know that the model is

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

where $E(u|x_1, x_2) = 0$ but we estimate $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$.

- As it was shown before

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1) y_i}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2}.$$

- Then **conditional on the regressors**

$$\begin{aligned} E\left(\tilde{\beta}_1\right) &= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1) x_{i2}}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \beta_2 \frac{S_{x_1,x_2}}{S_{x_1}^2}, \end{aligned}$$

where S_{x_1,x_2} is the sample covariance between x_1 and x_2 and $S_{x_1}^2$ is the sample variance of x_1.

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2 \frac{S_{x_1,x_2}}{S_{x_1}^2}$$
$$= \beta_1 + \beta_2 Corr(x_1, x_2) \frac{S_{x_2}}{S_{x_1}}.$$

Summary of Direction of Bias

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive Bias	Negative Bias
$\beta_2 < 0$	Negative Bias	Positive Bias

- Two cases where bias is equal to zero:
  - $\beta_2 = 0$, that is x_2 doesn't really belong in model.
  - x_1 and x_2 are uncorrelated in the sample.
- If $corr(x_2, x_1)$ and β_2 have the same sign, bias will be positive.
- If $corr(x_2, x_1)$ and β_2 have the opposite sign, bias will be negative.
- The More General Case: Technically, can only obtain the sign of the bias for the more general case if all of the included x's are uncorrelated.

The Variance-covariance matrix of the OLS estimator $(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ has the form:

$$\begin{bmatrix} Var\,(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) & ... & Cov(\hat{\beta}_0, \hat{\beta}_k) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var\,(\hat{\beta}_1) & ... & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(\hat{\beta}_0, \hat{\beta}_k) & Cov(\hat{\beta}_1, \hat{\beta}_k) & ... & Var\,(\hat{\beta}_k) \end{bmatrix}$$

- Let **x** stand for $(x_1, x_2, \ldots x_k)$.
- Assume $Var(u|\mathbf{x}) = \sigma^2$ (Homoskedasticity).
- Assuming that $Var(u|\mathbf{x}) = \sigma^2$ also implies that $Var(y|\mathbf{x}) = \sigma^2$.
- The 4 assumptions for unbiasedness, plus this homoskedasticity assumption are known as the Gauss-Markov assumptions.

Given the Gauss-Markov Assumptions

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)},$$

where the $SST_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from the regressing x_j on all other $x's$.

Components of OLS Variances:

- The error variance: a larger σ^2 implies a larger variance for the OLS estimators.
- The total sample variation: a larger SST_j implies a smaller variance for the estimators.
- Linear relationships among the independent variables: a larger R_j^2 implies a larger variance for the estimators.

The Variances of the OLS Estimator conditional on the sample values $\{x_i : i = 1, 2, \ldots, n\}$ are given by

$$
\begin{aligned}
Var\left(\hat{\beta}_0\right) &= \frac{\sigma^2}{(n-1)\, n\, S_x^2} \sum_{i=1}^{n} x_i^2, \\
Var\left(\hat{\beta}_1\right) &= \frac{\sigma^2}{(n-1)\, S_x^2},
\end{aligned}
$$

where $S_x^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Given our 5 Gauss-Markov Assumptions it can be shown that OLS is "BLUE":

- Best (have minimum variance, such that $Var(\hat{\beta}_j) \leq Var(\hat{\beta}_j^*)$, $j = 1, ..., k$ where $\hat{\beta}_j^*$ is any alternative estimator.
- Linear - weighted sum of the dependent variable.
- Unbiased- $E(\hat{\beta}_j) = \beta_j, E(\hat{\beta}_j^*) = \beta_j, j = 1, ..., k$.
- Estimator.

Thus, if the assumptions hold, use OLS.

We prove here the The Gauss-Markov Theorem in the case of the simple linear regression model for the estimator of the slope parameter.
An estimator is said to be linear if it can be written as a simple weighted sum of the dependent variable, where the weights do not depend on this variable.
Consider the estimator for the slope coefficient

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)y_i}{\sum_{i=1}^{n}x_i\left(x_i-\bar{x}\right)} \\ &= \sum_{i=1}^{n} w_i y_i,\end{aligned}$$

where

$$w_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}x_i\left(x_i-\bar{x}\right)}.$$

Outline of the proof:

1. Consider an alternative linear unbiased estimator.
2. Show that the new estimator can never have smaller variance than the OLS estimator.

Step 1: An alternative linear estimator for the slope coefficient will have the form

$$\bar{\beta}_1 = \sum_{i=1}^{n} k_i y_i.$$

where k_i is a function of the regressors. Unbiasedness means that $E\left(\bar{\beta}_1\right) = \beta_1$ and this requires that the weights should satisfy

$$\sum_{i=1}^{n} k_i = 0 \text{ and } \sum_{i=1}^{n} k_i x_i = 1.$$

Step 2: Notice that conditional on the sample values $\{x_i : i = 1, 2, \ldots, n\}$ we have

$$Var\left(\bar{\beta}_1\right) = \sigma^2 \sum_{i=1}^n k_i^2, \; Var\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(x_i - \bar{x}\right)^2}.$$

Hence

$$\begin{aligned}
Var\left(\bar{\beta}_1\right) - Var\left(\hat{\beta}_1\right) &= \sigma^2 \sum_{i=1}^n k_i^2 - \frac{\sigma^2}{\sum_{i=1}^n \left(x_i - \bar{x}\right)^2} \\
&= \sigma^2 \left(\sum_{i=1}^n k_i^2\right) [1 - \frac{1}{\sum_{i=1}^n k_i^2 \sum_{i=1}^n \left(x_i - \bar{x}\right)^2}] \\
&= \sigma^2 \left(\sum_{i=1}^n k_i^2\right) [1 - \frac{\left(\sum_{i=1}^n k_i x_i\right)^2}{\sum_{i=1}^n k_i^2 \sum_{i=1}^n \left(x_i - \bar{x}\right)^2}] \\
&= \sigma^2 \left(\sum_{i=1}^n k_i^2\right) [1 - correlation(x_i, k_i)^2] \geq 0
\end{aligned}$$

Suppose that we know that the model is

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

where $E(u|x_1, x_2) = 0$.

- Consider again $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$ so that

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{(n-1)\, S_{x_1}^2}.$$

  where $S_{x_1}^2$ is the sample variance of x_1.
- Recall that

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{(n-1)\, S_{x_1}^2 \,(1 - R_1^2)}.$$

- Thus $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$ unless x_1 and x_2 are uncorrelated, then they are the same.
- While the variance of the estimator is smaller for the misspecified model, unless $\beta_2 = 0$ the misspecified model is biased.
- As the sample size grows, the variance of each estimator shrinks to zero, making the variance difference less important.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)},$$

where the $SST_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from the regressing x_j on all other $x's$.

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

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

thus

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

- $se\left(\hat{\beta}_j\right)$ is called the <u>standard error of $\hat{\beta}_j$</u>.
- The square root of $\hat{\sigma}^2$ is called the <u>regression standard error</u>, or <u>standard error of the regression</u>
- $df = n - (k+1)$, or $df = n - k - 1$. df (i.e. degrees of freedom) is the (number of observations) – (number of estimated parameters). Therefore $\hat{\sigma}^2 = SSR/df$.

So far we have focussed on linear relationships between the dependent and independent variable, however in applied work in Economics we often encounter regression equations where the dependent variable appears in logarithmic form.
Why is this done?
Recall the Wages-Education example:

$$Wages = \beta_0 + \beta_1 Education + u,$$

$E\left[u|Education\right] = 0$. The sample regression function obtained was

$$\widehat{Wages} = -1.60468 + 0.81395 \times Education.$$

Notice that this implies that:

- For a person with 6 years of Education, an additional year will increase the hourly wages by \$0.81395.
- For a person with 15 years of Education, an additional year will increase the hourly wages by \$0.81395.

Conclusion: This may not be reasonable.

- In empirical research it has been found that a better characterization of how the wages change is to assume that each year of education increases wages by a constant percentage.
- A model that gives (approximately) a constant percentage effect is:

$$\log(Wages) = \beta_0 + \beta_1 Education + u,$$

  Why?
- The key reason lies in the following fact: If $\Delta y / y$ is close to zero:

$$\log(y + \Delta y) - \log(y) \cong \frac{\Delta y}{y}$$

  that is the difference between the natural logarithm of $y + \Delta y$ and the natural logarithm of y is the percentage change divided by 100.
- Consider the linear regression model that where the dependent variable is in the logarithm form (known as *log-linear model*)

$$\log(y) = \beta_0 + \beta_1 x + u.$$

- Let us drop the error term u for simplicity

$$\log(y) = \beta_0 + \beta_1 x$$

and denote Δy be the change in y when x changes by Δx.

- One can show that in this model

$$\frac{\Delta y}{y} \cong \beta_1 \Delta x$$

In words: a unit change in x is associated with a $100 \times \beta_1\%$ expected change in y.

Running the regression of $\log(Wages)$ on *Education* we obtain:

Dependent variable: $\log(Wages)$

Estimation Method: Ordinary Least Squares

Regressors	Estimates
Intercept	0.98237
Education	0.08262

- Hence, an additional year of education is expected to increase the hourly wages by 8.262%.
- For a person with 6 years of Education, an additional year is expected to increase the hourly wages by $0.08262 \times wages_6$ dollars, where $wages_6$ are the wages of that person.
- For a person with 15 years of Education, an additional year is expected to increase the hourly wages by $0.08262 \times wages_{15}$ dollars, where $wages_{15}$ are the wages of that person.

Other cases: *Linear-Log model*

- x is in logarithms and y is not, that is

$$y = \beta_0 + \beta_1 \log(x) + u.$$

  Denote Δy be the change in y when x changes by Δx.

- *Ignoring the error term* one can show that

$$\Delta y \cong \beta_1 \Delta x / x.$$

In words: a 1% change in x is associated with a $0.01 \times \beta_1$ expected change in y.

Other cases: *Log-Log model* or *constant elasticity model*

- Both x and y are in logarithms, that is

$$\log(y) = \beta_0 + \beta_1 \log(x) + u.$$

  Denote Δy be the change in y when x changes by Δx.

- *Ignoring the error term* one can show that

$$\Delta y / y \cong \beta_1 \Delta x / x.$$

In words: a 1% change in x is associated with a β_1% expected change in y (β_1 is the elasticity of y with respect to x).

Example: Economists often fit models that take logs of variables such as:

$$\log(Output) = \beta_0 + \beta_1 \log(Labour) + \beta_2 \log(Capital) + u,$$

Ignoring the error term u, this model corresponds to the *Cobb-Douglas production function*. That is, it is equivalent to

$$\begin{aligned} Output &= A \times Labour^{\beta_1} \times Capital^{\beta_2}, \\ A &= \exp(\beta_0). \end{aligned}$$

Thus:

-
    - β_1 is the elasticity of *Output* with respect to *Labour*.
    - β_2 is the elasticity of *Output* with respect to *Capital*.

- Log models are invariant to the scale of the variables.
- They give a direct estimate of elasticity (if both the dependent variable and regressors are in logarithms).
- For models with $y > 0$, the conditional distribution is often heteroskedastic or asymmetric, while $\log(y)$ is much less so.
- The distribution of $\log(y)$ is more narrow, limiting the effect of outliers.

- What types of variables are often used in log form?
  - Values measured in a currency that must be positive.
  - Very large variables, such as population.
- What types of variables are often used in level form?
  - Variables measured in years. **Example:** Education, Experience, tenure, age, etc.
  - Variables that are a proportion or percent.