

# Econometrics

	<b>Course Calendar</b>
<b>Week</b>	Main Content
<b>Week 7</b>	Extension of Simple Regression: Functional Forms I
<b>Week 8</b>	Extension of Simple Regression: Functional Forms II
<b>Week 9</b>	Extension of Simple Regression: Functional Forms III
<b>Week 10</b>	Multiple Regression
<b>Week 11</b>	Multiple Regression: Estimation and Interpretation
<b>Week 12</b>	Multiple Regression: Hands-on-Exercise
<b>Week 13</b>	Introduction to Dummy Variables
<b>Week 14</b>	Introduction to Dummy Variables and Regression Methods
<b>Week 15</b>	Regression with Dummy Variables: Hands-on-Exercise
<b>Week 16</b>	Application of Regression

# Econometrics

## Lecture 7. Functional Forms of Regression-I

Geetha Rani Prakasam, Ph.D.

Professor

# Recap

- Introduction to Simple Regression: PRF and SRF
- Simple Regression Estimation: sample parameters, var and SE of the sample parameters,  $r^2$  & Hands-on-Exercise
- Central Limit Theorem, Probability and Probability Density Function (PDF)
- Hypothesis Testing: Basics
- Simple Regression: Testing of Hypothesis

# Outline

- Units of measurement
- Regression on Standardized Variables
- Functional forms of the linear regression model:- Regression through origin
- **Regression through origin with an example**
- **Introduction to Different functional forms**
- **summary**

# Units of Measurement

- How the Y and X variables are measured and
- Whether a change in the units of measurement affects the regression results or not.
- To grasp the ideas on units of measurement, consider the data given in Table 6.2 (next slide), which refers to U.S. gross private domestic investment (GPDI) and gross domestic product (GDP), in billions as well as millions of (chained) 1992 dollars.

**TABLE 6.2** GROSS PRIVATE DOMESTIC INVESTMENT AND GDP, UNITED STATES, 1988–1997

Observation	GPDIBL	GPDIM	GDPB	GDPM
1988	828.2000	828200.0	5865.200	5865200
1989	863.5000	863500.0	6062.000	6062000
1990	815.0000	815000.0	6136.300	6136300
1991	738.1000	738100.0	6079.400	6079400
1992	790.4000	790400.0	6244.400	6244400
1993	863.6000	863600.0	6389.600	6389600
1994	975.7000	975700.0	6610.700	6610700
1995	996.1000	996100.0	6761.600	6761600
1996	1084.1000	1084100.0	6994.800	6994800
1997	1206.4000	1206400.0	7269.800	7269800

*Note:* GPDIBL = gross private domestic investment, billions of 1992 dollars.

GPDIM = gross private domestic investments, millions of 1992 dollars.

GDPB = gross domestic product, billions of 1992 dollars.

GDPM = gross domestic product, millions of 1992 dollars.

*Source:* *Economic Report of the President*, 1999, Table B-2, p. 328.

# Units of Measurement

- Suppose in the regression of GPD on GDP one researcher uses data in billions of dollars but another expresses data in millions of dollars. Will the regression results be the same in both cases?
- If not, which results should one use?
- Do the units in which the regressand and regressor(s) are measured make any difference in the regression results?
- If so, what is the appropriate course to follow in choosing units of measurement for regression analysis?
- To answer these questions, let us look at the example systematically.

## 6.2 Scaling and units of measurement

- **Let**  $Y_i = \beta^{\wedge}_1 + \beta^{\wedge}_2 X_i + u^{\wedge}_i$  (6.2.1)

where  $Y = \text{GDPDI}$  and  $X = \text{GDP}$ .

- **Define**  $Y^*_i = w_1 Y_i$  (6.2.2)

- **and**  $X^*_i = w_2 X_i$  (6.2.3)

where  $w_1$  and  $w_2$  are constants, called the scale factors;  $w_1$  may equal  $w_2$  or be different.

- From (6.2.2) and (6.2.3) it is clear that  $Y^*_i$  and  $X^*_i$  are rescaled  $Y_i$  and  $X_i$  respectively.
- Thus, if  $Y_i$  and  $X_i$  are measured in billions of dollars and one wants to express them in millions of dollars, we will have  $Y^*_i = 1000 Y_i$  and  $X^*_i = 1000 X_i$ ; here  $w_1 = w_2 = 1000$ .

# Units of Measurement

Now consider the regression using  $Y^*_i$  and  $X^*_i$  variables:

$$Y^*_i = \beta^{*^1} + \beta^{*^2} X^*_i + u^{*^i} \quad (6.2.4)$$

We want to find out the relationships between the following pairs:

1.  $\beta^{*^2} = (w_1/w_2) \beta^2$
2.  $\beta^{*^1} = w_1 \beta^1$
3.  $\sigma^{*^2} = w_1^2 \sigma^2$
4.  $\text{Var}(\beta^{*^1}) = w_1^2 \text{Var}(\beta^1)$
5.  $\text{Var}(\beta^{*^2}) = (w_1/w_2)^2 \text{Var}(\beta^2)$
6.  $r^2_{xy} = r^2_{x^*y^*}$

# Units of Measurement

- From least-squares theory we know (see Lectures 1 &2) that

- $\beta^{\wedge}_1 = \bar{y} - \beta^{\wedge}_2 \bar{x}$  (6.2.5)

- $\beta^{\wedge}_2 = \sum x_i y_i / \sum x_i^2$  (6.2.6)

- $\text{var}(\beta^{\wedge}_1) = \sigma^2 \sum x_i^2 / n \sum x_i^2$  (6.2.7)

- $\text{var}(\beta^{\wedge}_2) = \sigma^2 / \sum x_i^2$  (6.2.8)

- $\sigma^{\wedge 2} = \sum u^{\wedge 2}_i / (n-k)$  (6.2.9)

Applying the OLS method to (6.2.4), we obtain similarly

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \quad (6.2.10)$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \quad (6.2.11)$$

$$\text{var}(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n \sum x_i^{*2}} \cdot \sigma^{*2} \quad (6.2.12)$$

$$\text{var}(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \quad (6.2.13)$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \quad (6.2.14)$$

# Units of Measurement

- From these results it is easy to establish relationships between the two sets of parameter estimates.
- All that one has to do is recall these definitional relationships:
- $Y^*_{i} = w_1 Y_i$  (or  $y^*_i = w_1 y_i$ );
- $X^*_{i} = w_2 X_i$  (or  $x^*_i = w_2 x_i$ );
- $\hat{u}^*_i = w_1 \hat{u}_i$ ;  $Y^{-*} = w_1 Y^{-}$  and
- $X^{-*} = w_2 X^{-}$ .
- Making use of these definitions, then we can easily verify that

# Units of Measurement

$$\hat{\beta}_2^* = \left( \frac{w_1}{w_2} \right) \hat{\beta}_2 \quad (6.2.15)$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \quad (6.2.16)$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \quad (6.2.17)$$

$$\text{var}(\hat{\beta}_1^*) = w_1^2 \text{var}(\hat{\beta}_1) \quad (6.2.18)$$

$$\text{var}(\hat{\beta}_2^*) = \left( \frac{w_1}{w_2} \right)^2 \text{var}(\hat{\beta}_2) \quad (6.2.19)$$

$$r_{xy}^2 = r_{x^*y^*}^2 \quad (6.2.20)$$

# Units of Measurement

- From these results, it is made clear that, given the regression results based on one scale of measurement, one can derive the results based on another scale of measurement once the scaling factors, the  $w$ 's, are known.
- If  $w_1 = w_2$ , that is, the scaling factors are identical, the slope coefficient and its standard error remain unaffected in going from the  $(Y_i, X_i)$  to the  $(Y^*_i, X^*_i)$  scale, which should be intuitively clear.
- In practice, though, one should choose the units of measurement sensibly; there is little point in carrying all those zeros in expressing numbers in millions or billions of dollars.

# Units of Measurement

- From the results given in (6.2.15) through (6.2.20) one can easily derive some special cases.
- However, the intercept and its standard error are both multiplied by  $w_1$ .
- But if the X scale is not changed (i.e.,  $w_2 = 1$ ) and the Y scale is changed by the factor  $w_1$ , the slope as well as the intercept coefficients and their respective standard errors are all multiplied by the same  $w_1$  factor.

# Units of Measurement

- Finally, if the  $Y$  scale remains unchanged (i.e.,  $w_1 = 1$ ) but the  $X$  scale is changed by the factor  $w_2$ , the slope coefficient and its standard error are multiplied by the factor  $(1/w_2)$  but the intercept coefficient and its standard error remain unaffected.
- It should, however, be noted that the transformation from the  $(Y, X)$  to the  $(Y^*, X^*)$  scale does not affect the properties of the OLS estimators discussed in the preceding lectures.

## 6-2. Scaling and units of measurement

- A numerical example: The Relationship Between GPDI And GDP, United States, 1988–1997
- To substantiate the preceding theoretical results, let us return to the data given in Table 6.2 and examine the following results (numbers in parentheses are the estimated standard errors).
- **Both GPDI and GDP in billions of dollars:**
- $$\text{GPDI}_t = -1026.498 + 0.3016 \text{ GDP}_t \quad (6.2.21)$$
- $$se = (257.5874) (0.0399) \quad r^2 = 0.8772$$

## 6-2. Scaling and units of measurement

- **Both GPDl and GDP in millions of dollars:**
- $\text{GPDl}_t = -1,026,498 + 0.3016 \text{ GDP}_t \quad (6.2.22)$
- $\text{se} = (257,587.4) \quad (0.0399) \quad r^2 = 0.8772$
- Notice that the intercept as well as its standard error is 1000 times the corresponding values in the regression (6.2.21) (note that  $w_1 = 1000$  in going from billions to millions of dollars), but the slope coefficient as well as its standard error is unchanged, in accordance with theory.

## 6-2. Scaling and units of measurement

- GPDl in billions of dollars and GDP in millions of dollars:
- $$\text{GPDl}_t = -1026.498 + 0.000301 \text{ GDP}_t \quad (6.2.23)$$
- $$\text{se} = (257.5874) (0.0000399) \quad r^2 = 0.8772$$
- As expected, the slope coefficient as well as its standard error is 1/1000 its value in (6.2.21), since only the X, or GDP, scale is changed.

## 6-2. Scaling and units of measurement

- GPDl in millions of dollars and GDP in billions of dollars:
- $\text{GPDl}_t = -1,026,498 + 301.5826 \text{ GDP}_t$  (6.2.24)
- $se = (257,587.4) (39.89989) \quad r^2 = 0.8772$
- Again notice that both the intercept and the slope coefficients as well as their respective standard errors are 1000 times their values in (6.2.21), in accordance with our theoretical results.

## 6-2. Scaling and units of measurement

- Notice that in all the regressions presented above the  $r^2$  value remains the same, which is not surprising because the  $r^2$  value is invariant to changes in the unit of measurement, as it is a pure, or dimensionless, number.
- Since the slope coefficient  $\beta_2$  is simply the rate of change, it is measured in the units of the ratio.
- Units of the dependent variable / Units of the explanatory variable.

## 6-2. Scaling and units of measurement

- Thus in regression (6.2.21) the interpretation of the slope coefficient 0.3016 is that if GDP changes by a unit, which is 1 billion dollars, GPDI on the average changes by 0.3016 billion dollars.
- In regression (6.2.23) a unit change in GDP, which is 1 million dollars, leads on average to a 0.000302 billion dollar change in GPDI.
- The two results are of course identical in the effects of GDP on GPDI; they are simply expressed in different units of measurement.

## 6.3 Regression on Standardized Variables

- We saw in the previous section that the units in which the regressand and regressor(s) are expressed affect the interpretation of the regression coefficients.
- This can be avoided if we are willing to express the regressand and regressor(s) as standardized variables.
- A variable is said to be standardized if we subtract the mean value of the variable from its individual values and divide the difference by the standard deviation of that variable.

## 6.3 Regression on Standardized Variables

- Thus, in the regression of Y and X, if we redefine these variables as

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \quad (6.3.1)$$

$$X_i^* = \frac{X_i - \bar{X}}{S_X} \quad (6.3.2)$$

- where  $\bar{y}$  = sample mean of Y,  $S_Y$  = sample standard deviation of Y,  $\bar{x}$  = sample mean of X, and  $S_X$  is the sample standard deviation of X; the variables  $Y^* i$  and  $X^* i$  are called standardized variables.
- An interesting property of a standardized variable is that its mean value is always zero and its standard deviation is always 1.

## 6.3 Regression on Standardized Variables

- As a result, it does not matter in what unit the regressand and regressor(s) are measured.
- Therefore, instead of running the standard (bivariate) regression:
- $Y_i = \beta_1 + \beta_2 X_i + u_i$  (6.3.3)
- we could run regression on the standardized variables as

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + u_i^* \quad (6.3.4)$$

$$= \beta_2^* X_i^* + u_i^* \quad (6.3.5)$$

## 6.3 Regression on Standardized Variables

- Since it is easy to show that, in the regression involving standardized regressand and regressor(s), the intercept term is always zero.
- Because for the standardized variables the mean values of the dependent variable and the regressor are zero. Hence the intercept value is zero.
- The regression coefficients of the standardized variables, denoted by  $\beta^*1$  and  $\beta^*2$ , are known in the literature as the beta coefficients.
- Incidentally, notice that (6.3.5) is a regression through the origin. How do we interpret the beta coefficients?
- The interpretation is that if the (standardized) regressor increases by one standard deviation, on average, the (standardized) regressand increases by  $\beta^*2$  standard deviation units.

## 6.3 Regression on Standardized Variables

- Thus, unlike the traditional model (6.3.3), we measure the effect not in terms of the original units in which Y and X are expressed, but in standard deviation units.
- To show the difference between (6.3.3) and (6.3.5), let us return to the GPDl and GDP example discussed in the previous section.
- The results of (6.2.21) discussed previously are reproduced here for convenience.
- $$\text{GPDl } t = -1,026,498 + 0.3016 \text{ GDPl } t \quad (6.2.21)$$
- $$se = (257,587.4) (0.0399) \quad r^2 = 0.8772$$

where GPDl and GDPl are measured in billions of dollars.

## 6.3 Regression on Standardized Variables

- The results corresponding to (6.3.5) are as follows, where the starred variables are standardized variables:

$$\widehat{\text{GPD}}_t^* = 0.9387 \text{GDP}_t^* \quad (6.3.7)$$

se = (0.1149)

- We know how to interpret (6.3.6): If GDP goes up by a dollar, on average GPD goes up by about 30 cents.
- How about (6.3.7)?
- Here the interpretation is that if the (standardized) GDP increases by one standard deviation, on average, the (standardized) GPD increases by about 0.94 standard deviations.

## 6.3 Regression on Standardized Variables

- What is the advantage of the standardized regression model over the traditional model?
- The advantage becomes more apparent if there is more than one regressor, a topic we will discuss in Lecture 8.
- By standardizing all regressors, we put them on equal basis and therefore can compare them directly.
- If the coefficient of a standardized regressor is larger than that of another standardized regressor appearing in that model, then the latter contributes more relatively to the explanation of the regressand than the latter.
- In other words, we can use the beta coefficients as a measure of relative strength of the various regressors.
- We discuss more on this in the next two lectures.

## 6.3 Regression on Standardized Variables

- Before we move to the next topic, two points may be noted.
- First, for the standardized regression (6.3.7) we have not given the  $r^2$  value because this is a regression through the origin for which the usual  $r^2$  is not applicable, as will be discussed in the next sub-topic.
- Second, there is an interesting relationship between the  $\beta$  coefficients of the conventional model and the beta coefficients.

## 6.3 Regression on Standardized Variables

- For the bivariate case, the relationship is as follows:

$$\hat{\beta}_2^* = \hat{\beta}_2 \left( \frac{S_x}{S_y} \right) \quad (6.3.8)$$

- where  $S_x$  = the sample standard deviation of the X regressor and  $S_y$  = the sample standard deviation of the regressand.
- Therefore, one can crisscross between the  $\beta$  and beta coefficients if we know the (sample) standard deviation of the regressor and regressand.
- We will see later that this relationship holds true in the multiple regression in Lectures 10 and 11.

# Regression through the Origin

# What is regression through the origin?

- Regression through the origin is a technique used in some disciplines when theory suggests that the regression line must run through the origin, i.e., the point 0,0.
- Example: We have a dataset that has standardized test scores for writing and reading ability.
- The tests are normed to have a mean of 50 and standard deviation of 10.
- The OLS regression for predicting the writing score from the reading score is given as:
  - $F(1, 198) = 109.52$      $P = 0.0000$      $R\text{-squared} = 0.36$

# Regression through the origin: Example

WRITE	COEFF.	STD. ERROR	T	P> T
READ	.5517051	.0527178	10.47	0.000
CONSTANT	23.95944	2.805744	8.54	0.000

- This model is interpreted as:
- The coefficient of .55 for read indicates that for each one point increase in reading the expected writing score should increase by .55.
- The value of the constant is approximately 24 and means that a person with a reading score of zero would have a predicted writing score of 24.

# Regression through the origin: Example

- An educator argues that the model doesn't make sense because a people with zero reading ability should not be able have a score of nearly 24 on writing ability.
- In fact the person should have a score of zero on writing.
- The argument is the regression line should go through the origin, i.e., the regression model should be run without a constant.
- Here is what that model would look like.
  - $F(1, 199) = 7064.58$     $P = 0.0000$     $R\text{-squared} = 0.97$

# Regression through the origin: Example

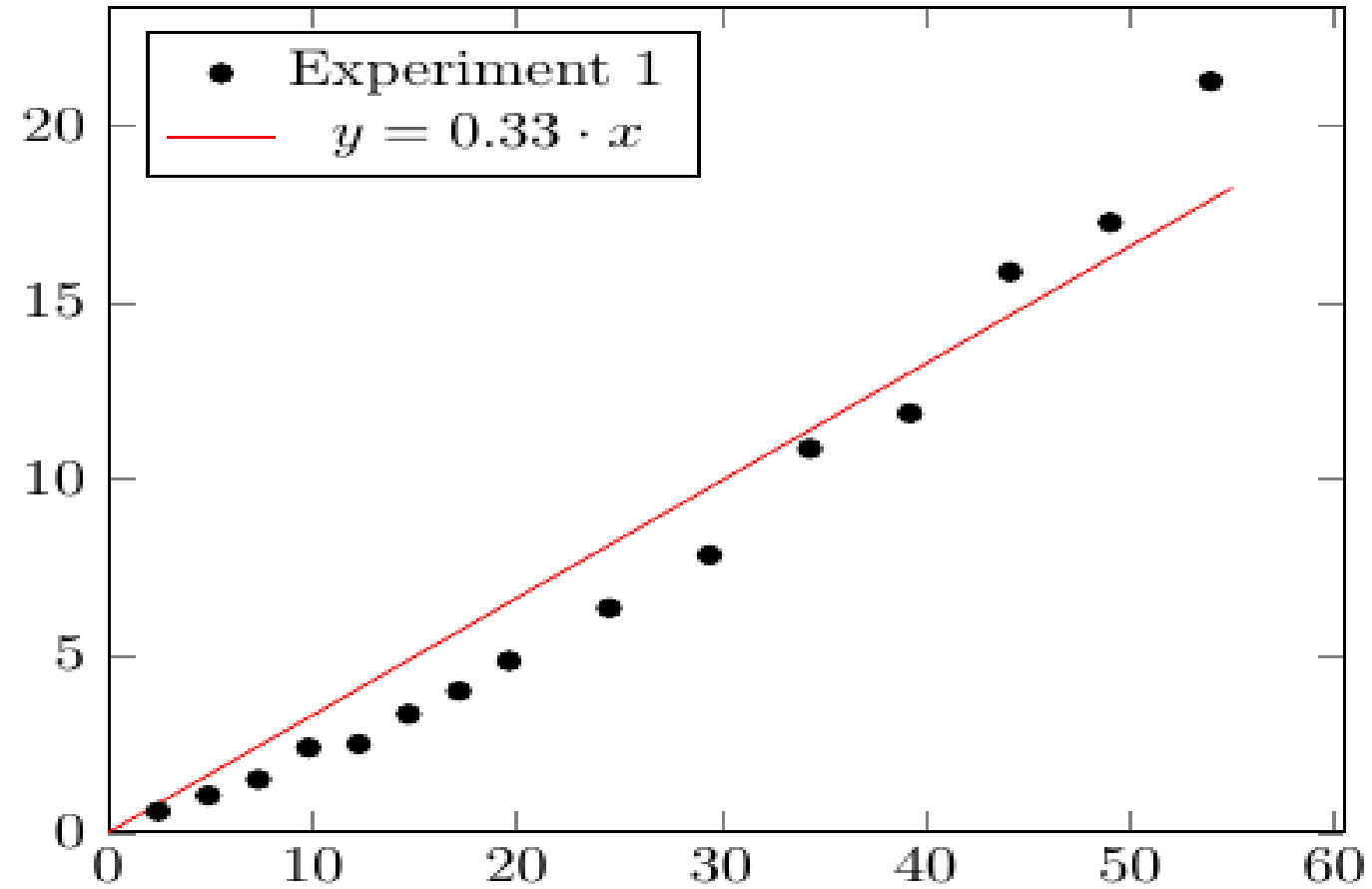
WRITE	COEFF.	STD. ERROR	T	P> T
READ	.9934953	.0118201	84.05	0.000

- This model is interpreted to mean that for every one point increase in reading ability there is a .99 (nearly one) point increase in writing ability and that when the reading score is zero the predicted writing score is also zero.
- In fact almost everything about this model looks great. The F-ratio is huge and the R-squared is .97.
- Most writers on the topic suggest that you test to see whether the constant is zero (test  $H_0: \beta_0 = 0$ ). Clearly in this example the intercept is significantly different from zero.

## 6-1. Regression through the origin

- As explained in the previous slides, there are occasions when the two-variable SRF assumes the following form:
- The SRF form of regression:
- $Y_i = \beta^{\wedge}_2 X_i + u^{\wedge}_i$  (6.1.5)
- In this model the intercept term is absent or zero, hence the name regression through the origin.
- Comparison of two types of regressions:
- Regression through-origin model and Regression with intercept

# Line fit: Regression through Origin



Source: <https://i.stack.imgur.com/zbQcl.png>

## 6-1. Regression through the origin

- Now applying the OLS method to (6.1.5), we obtain the following formulas for  $\hat{\beta}_2$  and its variance
- $\hat{\beta}_2 = \sum X_i Y_i / \sum X_i^2$  (6.1.6)
- $\text{var}(\hat{\beta}_2) = \sigma^2 / \sum X_i^2$  (6.1.7)
- where  $\sigma^2$  is estimated by
- $\hat{\sigma}^2 = \sum (\hat{u}_i)^2 / (n-1)$  (6.1.8)
- It is interesting to compare these formulas with those obtained when the intercept term is included in the model:

## 6-1. Comparison two types of regressions

$$\beta^{\wedge}_2 = \Sigma X_i Y_i / \Sigma X_i^2 \quad (6.1.6) \quad \text{Origin}$$

$$\beta^{\wedge}_2 = \Sigma x_i y_i / \Sigma x_i^2 \quad (3.1.6) \quad \text{Intercept}$$

$$\text{var}(\beta^{\wedge}_2) = \sigma^2 / \Sigma X_i^2 \quad (6.1.7) \quad \text{Origin}$$

$$\text{var}(\beta^{\wedge}_2) = \sigma^2 / \Sigma x_i^2 \quad (3.3.1) \quad \text{Intercept}$$

$$\sigma^{\wedge 2} = \Sigma (u^{\wedge}_i)^2 / (n-1) \quad (6.1.8) \quad \text{Origin}$$

$$\sigma^{\wedge 2} = \Sigma (u^{\wedge}_i)^2 / (n-2) \quad (3.3.5) \quad \text{Intercept}$$


## Difference between the two sets of formulas:

- Without intercept (O) – we use raw cross products
- With intercept (I) – we use adjusted (from mean) cross products.
- What are the consequences of using raw data?
- $\sum(\hat{u}_i)^2$  need not be zero.
- $R^2$  can be negative.
- Because of these special feature of this model, one should stick to the conventional, intercept-present model unless there is a very strong a priori expectation.

## Difference between the two sets of formulas:

- Although the interceptless or zero intercept model may be appropriate on occasions, there are some features of this model that need to be noted.
- First,  $\sum \hat{u}_i$ , which is always zero for the model with the intercept term (the conventional model), need not be zero when that term is absent.
- In short,  $\sum \hat{u}_i$  need not be zero for the regression through the origin.
- Second,  $r^2$ , the coefficient of determination introduced in Lecture 3, which is always nonnegative for the conventional model, can on occasions turn out to be negative for the interceptless model!

## Difference between the two sets of formulas:

- This deviation from normal (anomalous) result arises because the  $r^2$  introduced in Lecture 3 explicitly assumes that the intercept is included in the model.
- Therefore, the conventionally computed  $r^2$  may not be appropriate for regression-through-the-origin models.
- But one can compute what is known as the raw  $r^2$  for such models, which is defined as:
- Raw  $r^2 = (\sum X_i Y_i)^2 / \sum X_i^2 \sum Y_i^2$  (6.1.9)   $r^2$  for regression through-origin model.

## 6-1. Regression through the origin

- These are raw (i.e., not mean-corrected) sums of squares and cross products.
- Although this raw  $r^2$  satisfies the relation  $0 < r^2 < 1$ , it is not directly comparable to the conventional  $r^2$  value.
- For this reason some authors do not report the  $r^2$  value for zero intercept regression models.
- Because of these special features of this model, one needs to exercise great caution in using the zero intercept regression model.
- Unless there is very strong a priori expectation, one would be well advised to stick to the conventional, intercept-present model.

## 6-1. Regression through the origin

- This has a dual advantage. First, if the intercept term is included in the model but it turns out to be statistically insignificant (i.e., statistically equal to zero), for all practical purposes we have a regression through the origin.
- Without very strong a priori expectation, well advise is sticking to the conventional, intercept-present model.
- If intercept equals to zero statistically, for practical purposes we have a regression through the origin.
- If in fact there is an intercept in the model but we insist on fitting a regression through the origin, we would be committing a specification error

# An Illustrative Example: The Characteristic Line of Portfolio Theory

- Table 6.1 gives data on the annual rates of return (%) on Afuture Fund, a mutual fund whose primary investment objective is maximum capital gain, and on the market portfolio, as measured by the Fisher Index, for the period 1971–1980.
- The characteristic line of investment analysis, which can be written as
- $Y_i = \alpha_i + \beta_i X_i + u_i$  (6.1.10)
- where  $Y_i$  = annual rate of return (%) on Afuture Fund
- $X_i$  = annual rate of return (%) on the market portfolio
- $\beta_i$  = slope coefficient, also known as the Beta coefficient in portfolio theory, and
- $\alpha_i$  = the intercept

Table 6.1 Annual Rates of Return on Afuture Fund and on the Fisher Index (Market Portfolio), 1971–1980

Year	RETURN ON AFUTURE FUND (in %) Y	RETURN ON THE FISHER INDEX (in%) (X)
1971	67.5	19.5
1972	19.2	8.5
1973	-35.2	-29.3
1974	-42.0	-26.5
1975	63.7	61.9
1976	19.3	45.5
1977	3.6	9.5
1978	20.0	14.0
1979	40.3	35.3
1980	37.5	31.0

Data taken from Haim Levy and Marshall Sarnat, Portfolio and Investment Selection: Theory and Practice, Prentice-Hall International, Englewood Cliffs, N.J., 1984, pp. 730 and 738. These data were obtained by the authors from Weisenberg Investment Service, Investment Companies, 1981 edition:

As reported in Source: Basic Econometrics by Damodar Gujarati, Page.168.

# An Illustrative Example: The Characteristic Line of Portfolio Theory

- In the literature there is no consensus about the prior value of  $\alpha_i$ .
- Some empirical results have shown it to be positive and statistically significant and some have shown it to be not statistically significantly different from zero; in the latter case we could write the model as
- $$Y_i = \beta_i X_i + u_i \quad (6.1.11)$$
- that is, a regression through the origin.
- If we decide to use model (6.1.11), we obtain the following regression results
- $\hat{Y}_i = 1.0899 X_i (0.1916)$
- raw  $r^2 = 0.7825$  (6.1.12)  $t = (5.6884)$  which shows that  $\beta_i$  is significantly greater than zero.

# An Illustrative Example: The Characteristic Line of Portfolio Theory

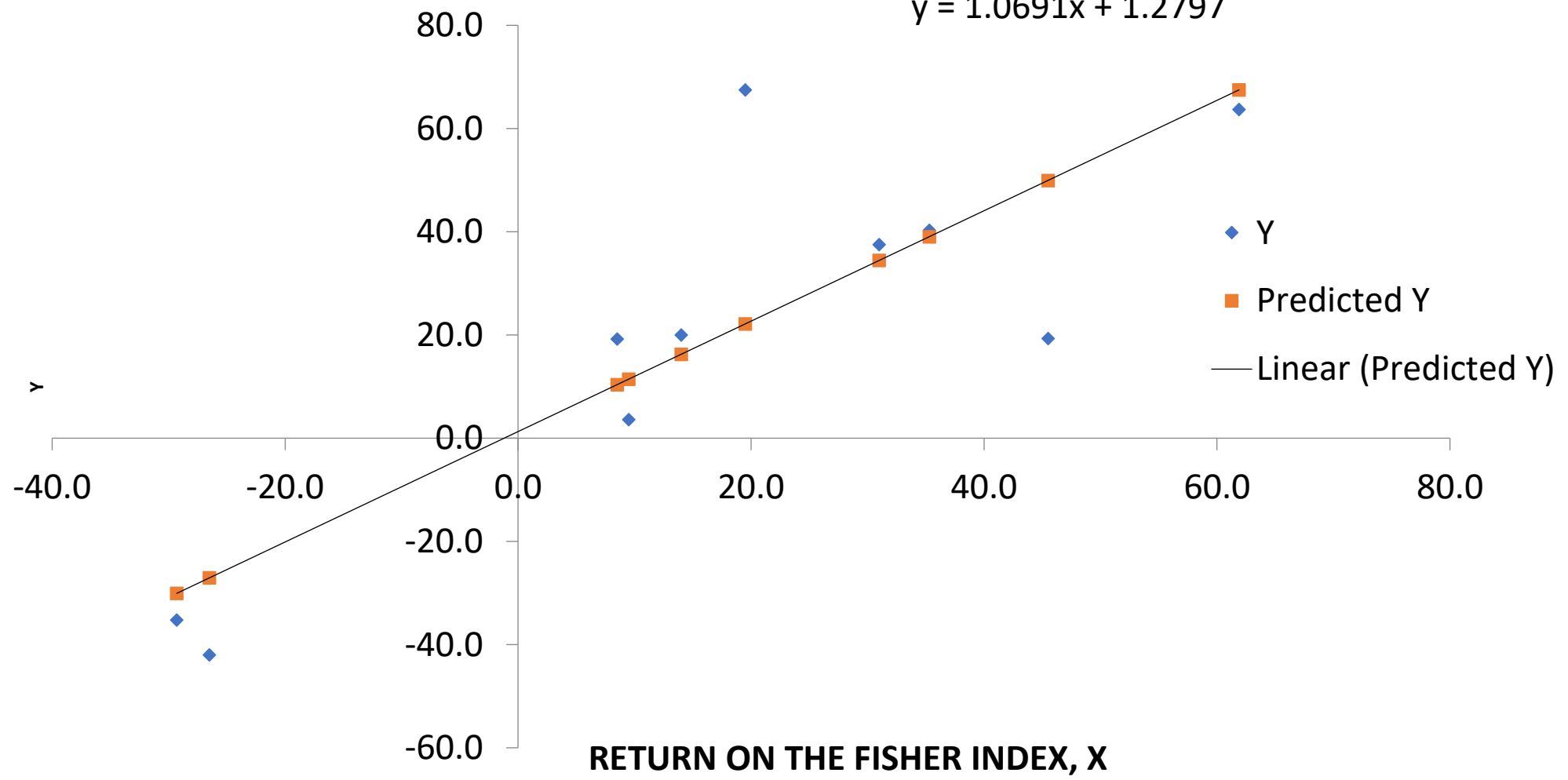
- The interpretation is that a 1 percent increase in the market rate of return leads on the average to about 1.09 percent increase in the rate of return on Afuture Fund.
- How can we be sure that model (6.1.11), not (6.1.10), is appropriate, especially in view of the fact that there is no strong a priori belief in the hypothesis that  $\alpha_i$  is in fact zero?
- This can be checked by running the regression (6.1.10).
- Using the data given in Table 6.1, we obtained the following results:

# An Illustrative Example: The Characteristic Line of Portfolio Theory

- $\hat{Y}_i = 1.2797 + 1.0691X_i$
- $(7.6886) (0.2383) \quad (6.1.13)$
- $t = (0.1664) (4.4860) \quad r^2 = 0.7155$
- Note: The  $r^2$  values of (6.1.12) and (6.1.13) are not directly comparable.
- From these results one cannot reject the hypothesis that the true intercept is equal to zero, thereby justifying the use of (6.1.1), that is, regression through the origin.

### Predicted Return on Afuture Fund

$$y = 1.0691x + 1.2797$$



# An Illustrative Example: The Characteristic Line of Portfolio Theory

- Note that there is not a great deal of difference in the results of (6.1.12) and (6.1.13), although the estimated standard error of  $\hat{\beta}$  is slightly lower for the regression-through-the-origin model, thus supporting Theil's argument (Henri Theil points out that if the intercept is in fact absent, the slope coefficient may be estimated with far greater precision than with the intercept term left in.) that if  $\alpha_i$  is in fact zero, the slope coefficient may be measured with greater precision:
- Using the data given in Table 6.1 and the regression results, we can easily verify that the 95% confidence interval for the slope coefficient of the regression-through-the-origin model is (0.6566, 1.5232) whereas for the model (6.1.13) it is (0.5195, 1.6186); that is, the former confidence interval is narrower than the latter.

# Extension of Simple Regression: Different Functional Forms

- How the two-variable models can deal with some interesting practical problems with variables that are non-linear.
- Parameters to be linear; not necessarily the variables.
- Using the above two points, in the lectures 8 and 9, we consider some commonly used regression models that may be nonlinear in the variables but are linear in the parameters or that can be made so by suitable transformations of the variables.
- Before that, why these different functional forms are important?

# Why learn about different functional forms?

- The reality or world is not linear.
- The assumption of the linearity of phenomena under consideration is highly arbitrary.
- It is usually necessary for research that encompasses a small number of observations because it facilitates parameter estimations.
- When we have a larger sample of observations, we may consider non-linear dependencies between dependent and independent variables.

# Why learn about different functional forms?

- To afford this, we may want to estimate a non-linear model.
- These kinds of models require more advanced estimation techniques and computation power.
- However, there is one vital alternative to this approach.
- Namely, we can approximate non-linear relations with the mean of linear models on transformed variables

# Different Functional Forms

- In particular, we discuss the following regression models:
- Semilog models – log-lin and lin-log models
- The log-linear model or double log model
- Reciprocal models
- The logarithmic reciprocal model
- We discuss the special features of each model, when they are appropriate, and how they are estimated and their interpretation.
- Each model is illustrated with suitable examples.

# Summary and conclusions

- The units and scale in which the regressand and the regressor(s) are expressed are very important because the interpretation of regression coefficients critically depends on them.
- In empirical research the researcher should not only quote the sources of data but also state explicitly how the variables are measured.
- Sometimes a regression model may not contain an explicit intercept term.
- Such models are known as **regression through the origin**.
- Although the algebra of estimating such models is simple, one should use such models with caution.

# Summary

- In such models the sum of the residuals  $\hat{u}_i$  is nonzero; additionally, the conventionally computed  $r^2$  may not be meaningful.
- Regression on Standardized Variables: Mean 0 and SD 1
- Basics of exponential functions
- Exponential growth and decay

# References

- Chapter 6: Extensions of the Two-variable Linear Regression Model in Basic Econometrics by Domodar Gujarati.
- <https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-what-is-regression-through-the-origin/>
- [https://math.libretexts.org/Bookshelves/Applied\\_Mathematics/Applied\\_Finite\\_Mathematics\\_\(Sekhon\\_and\\_Bloom\)/05%3A\\_Exponential\\_and\\_Logarithmic\\_Functions/5.03%3A\\_Graphs\\_and\\_Properties\\_of\\_Exponential\\_Growth\\_and\\_Decay\\_Functions](https://math.libretexts.org/Bookshelves/Applied_Mathematics/Applied_Finite_Mathematics_(Sekhon_and_Bloom)/05%3A_Exponential_and_Logarithmic_Functions/5.03%3A_Graphs_and_Properties_of_Exponential_Growth_and_Decay_Functions)
- [https://virtualllearningacademy.net/VLA/LessonDisplay/Lesson6156/MATHALGIIBU17Exponential\\_Decay.html](https://virtualllearningacademy.net/VLA/LessonDisplay/Lesson6156/MATHALGIIBU17Exponential_Decay.html)
- <https://sites.math.washington.edu/~conroy/general/expoanim/expoanim.htm>

# What next?

- Different functional forms but before that
- Basics of Logarithmic transformations of variables