Two Variable Regression Model

Extensions

1. Hypotheses Testing – General Case
   - Individual Test
   - Overall Test
2. Hypotheses Testing – Specific Cases

Model Evaluation:

Classical Normal Linear Regression Model

1. Probability Distribution & Normality Assumption of the Disturbance Term
2. Properties of OLS Estimators under Normality Assumption
3. Method of Maximum Likelihood (ML)
4. Hypothesis Testing
   - Individual Test – t-test
   - Overall Test – F-test
5. Coefficient of Determination R^2

Probability Distribution of Disturbances

Under CLRM, there is no assumption on distribution of disturbances u_i.

Objective of econometrics is not just only estimating the model, but also testing some statistical hypotheses concerning on population and sample regression model.

Since probability distributions of estimators are necessary to draw inferences about their population, assumption of probability distribution of u_i plays an important role in hypothesis testing.

Normality Assumption

Classical *normal* linear regression assumes that u_i is *normally* distributed with:

Mean: $E\left(u_i\right)=0$

Variance: $E\left(u_i^2\right)=\sigma^2$

cov(u_i, u_j): $E\left(u_i u_j\right)=0 \qquad \forall i \neq j$

and $u_i \sim N(0,\sigma^2)$

Since u_i and u_j are independent, u_i is also independently distributed.

$$u_i \sim NID(0,\sigma^2)$$

Why do we assume normal distribution?

Normality Assumption

Reasons of assuming normal distribution:

1. Central Limit Theorem:

   If there are a large number of independent and identically distributed random variables, the distribution of their sum tends to a normal distribution as the number of such variables increases indefinitely.

2. If the number of variables is not very large or if these variables are not strictly independent, their sum may still be normally distributed.

Normality Assumption

3. Probability distributions of OLS estimators can be easily derived.

   Because any linear function of normally distributed variables is itself normally distributed.

4. Normal distribution is a comparatively simple distribution.

Properties of OLS Estimators under Normality Assumption $\hat{\beta}_1,\ \hat{\beta}_2,\ and\ \hat{\sigma}^2$

1. Unbiased.
2. Efficient Estimators or Minimum Variance unbiased.
3. Consistency: As the sample size increases indefinitely, the estimators converge to their true population values.

Properties of OLS Estimators under Normality Assumption $\hat{\beta}_1,\ \hat{\beta}_2,\ and\ \hat{\sigma}^2$

4. $\hat{\beta}_1$ is normally distributed $\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\hat{\beta}_1}\right)$

   or $Z = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$

5. $\hat{\beta}_2$ is normally distributed $\hat{\beta}_2 \sim N\left(\beta_2, \sigma^2_{\hat{\beta}_2}\right)$

   or $Z = \dfrac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim N(0,1)$

6. $\dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(df=n-2)}$

Properties of OLS Estimators under Normality Assumption $\hat{\beta}_1,\ \hat{\beta}_2,\ and\ \hat{\sigma}^2$

7. $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ are distributed independently of $\hat{\sigma}^2$

8. $\left(\hat{\beta}_1, \hat{\beta}_2\right)$ have minimum variance in the entire class of unbiased estimators.

   Best Unbiased Estimators (BUE).

If $u_i \sim N(0, \sigma^2)$

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

$$\operatorname{var}(Y_i) = \sigma^2$$

Then $Y_i \sim N\left(\beta_1 + \beta_2 X_i, \sigma^2\right)$

Method of Maximum Likelihood

Method of point estimation with some stronger theoretical properties than OLS.

MLE Estimator of σ^2 is biased estimators but asymptotically (n increase indefinitely) unbiased.

Unrestriction from normality assumption.

Hypothesis Testing – Individual Test

General Comments

Null Hypothesis or H_0

$$H_o: \quad \beta_2 = 0$$

Alternative Hypothesis or H_a

$$H_a: \quad \beta_2 \neq 0$$

Two approaches for testing hypothesis:

- Confidence Interval
- Test of Significance

Hypothesis Testing – Individual Test

Test-of-Significance Approach

A test of significance is a procedure by which sample results are used to verify the truth or falsify of a null hypothesis.

Under normal distribution:

$$t = \frac{\hat{\beta}_2 - \beta_2}{se\left(\hat{\beta}_2\right)}$$

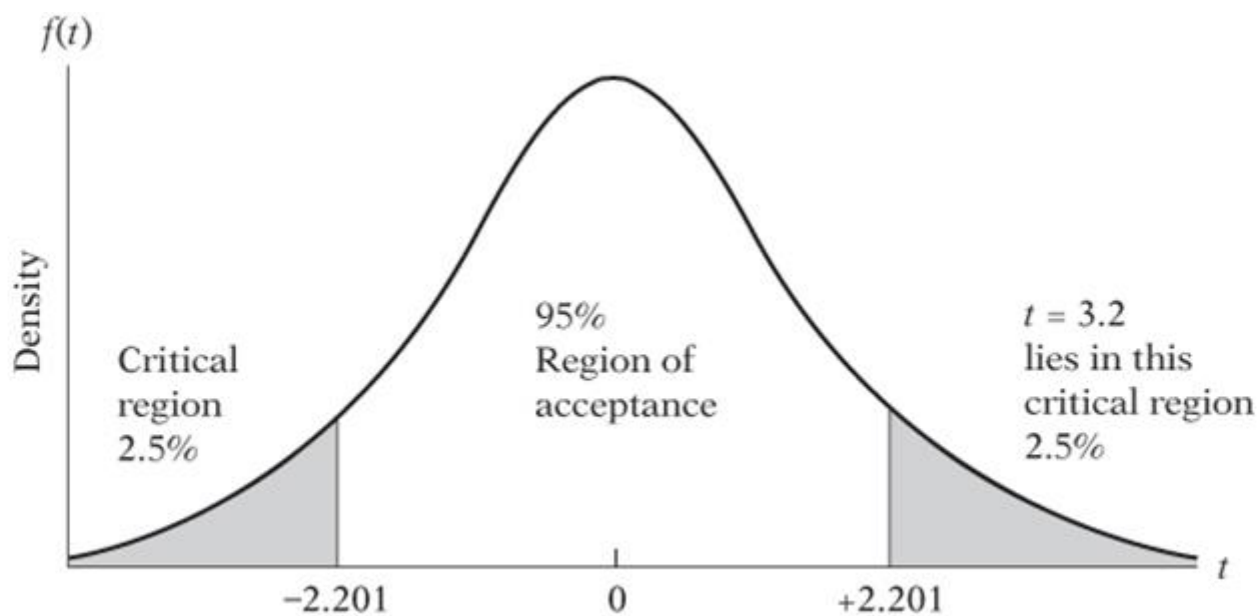
$$= \frac{\left(\hat{\beta}_2 - \beta_2\right)}{\hat{\sigma} \Big/ \sqrt{\sum \left(X_i - \bar{X}\right)^2}}$$

Hypothesis Testing – Individual Test

Some Practical Aspects

The Exact Level of Significance: The p Value

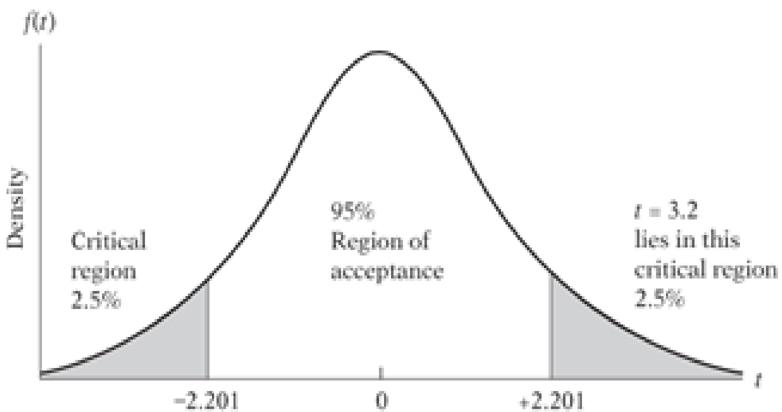
The p value is defined as the lowest significance level at which a null hypothesis can be rejected.

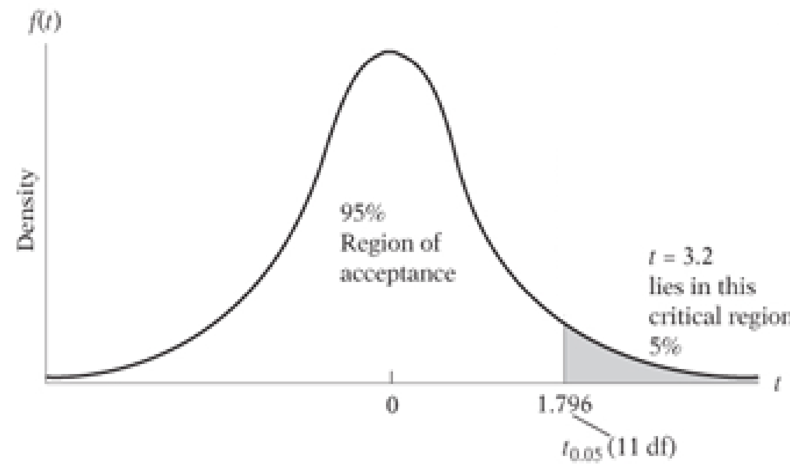


Hypothesis Testing – Individual Test

Summary

Type of Hypothesis	H_0: The Null Hypothesis	H_1: The Alternative Hypothesis	Decision Rule: Reject H_0 If
Two-tail	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{\alpha/2,\mathrm{df}}$
Right-tail	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\mathrm{df}}$
Left-tail	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\mathrm{df}}$





Regression Analysis & Analysis of Variance – Overall Test

$$F = \frac{MSS\ of\ ESS}{MSS\ of\ RSS}$$

$$= \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum \hat{u}_i^2 \big/ (n-2)}$$

$$= \frac{\hat{\beta}_2^2 \sum x_i^2}{\hat{\sigma}^2}$$

Regression Analysis & Analysis of Variance – Overall Test

Source of Variation	SS*	df	MSS†
Due to regression (ESS)	$\sum \hat{y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$	1	$\hat{\beta}_2^2 \sum x_i^2$
Due to residuals (RSS)	$\sum \hat{u}_i^2$	$n-2$	$\dfrac{\sum \hat{u}_i^2}{n-2} = \hat{\sigma}^2$
TSS	$\sum y_i^2$	$n-1$	

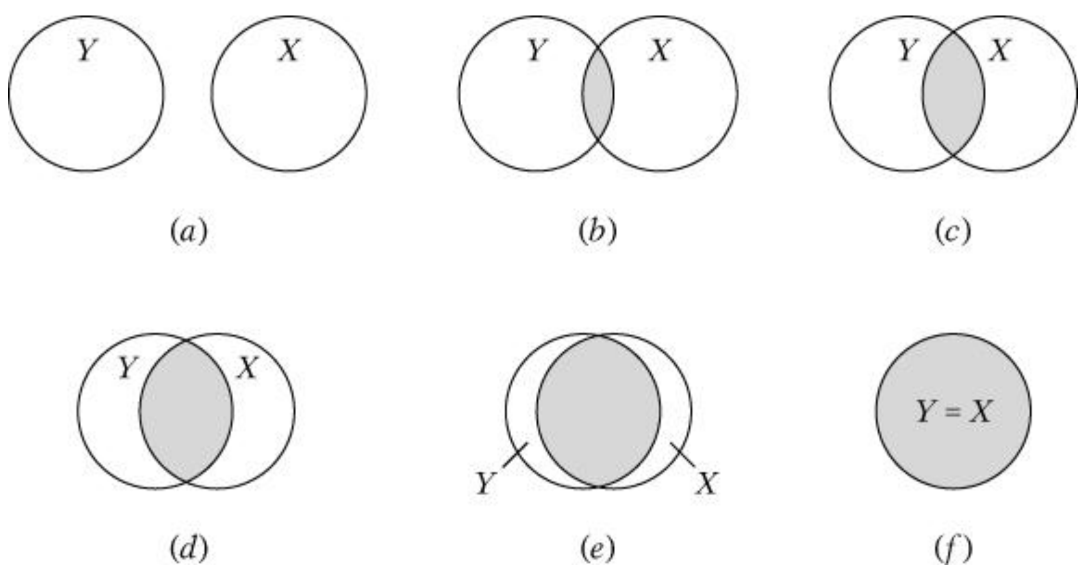
$$F = \frac{MSS\ of\ ESS}{MSS\ of\ RSS}$$

Source of Variation	SS	df	MSS	
Due to regression (ESS)	95.4255	1	95.4255	$F = \dfrac{95.4255}{0.8811}$
Due to residuals (RSS)	9.6928	11	0.8811	$= 108.3026$
TSS	105.1183	12		

Coefficient of Determination R^2

If all observations lie on the regression line, that sample regression line would *perfectly* fit the data.

Coefficient of determination (R^2) is a summary measure that tells how well the sample regression line fits the data.



Coefficient of Determination R^2

From $$Y_i = \hat{Y}_i + \hat{u}_i$$

In Deviation Form: where $y_i = Y_i - \bar{Y}$

$$y_i = \hat{y}_i + \hat{u}_i$$

Squaring and Summing over the Sample

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2$$

Coefficient of Determination R^2

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

TSS = ESS + RSS

Where:

TSS $= \sum y_i^2 = \sum \left(Y_i - \bar{Y}\right)^2$

= Total Sum of Squares (TSS)

= Total variation of the actual Y values about their sample mean

Coefficient of Determination R^2

$$\sum \hat{y}_i^2 = \sum \left(\hat{Y}_i - \bar{Y}\right)^2$$

= Explained Sum of Squares (ESS)

= Variation of the estimated Y values about their sample mean

$$\sum \hat{u}_i^2$$ = Residual Sum of Squares (RSS)

= Residual or unexplained variation of the Y values about regression line

Coefficient of Determination R^2

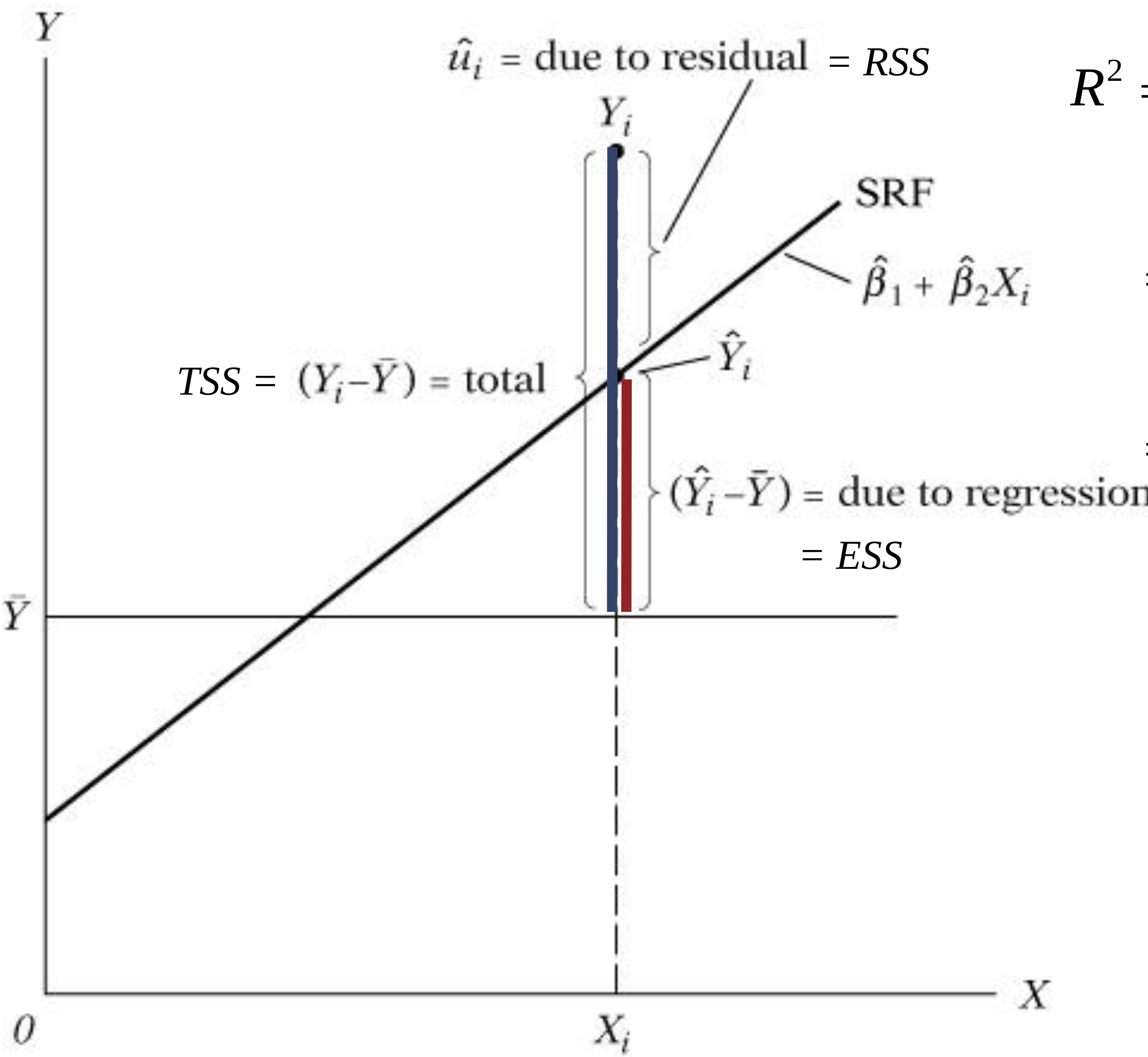
From $TSS = ESS + RSS$

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$

$$= \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} + \frac{\sum \hat{u}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

$$r^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Coefficient of Determination R^2



$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = \frac{ESS}{TSS}$$

$$= 1 - \frac{RSS}{TSS}$$

$$= 1 - \frac{\sum \hat{u}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

Coefficient of Determination r^2

Properties of R^2

1. Nonnegative quantity
2. Value range between 0 and 1

Sample Correlation Coefficient (r)

$$r = \pm\sqrt{r^2}$$

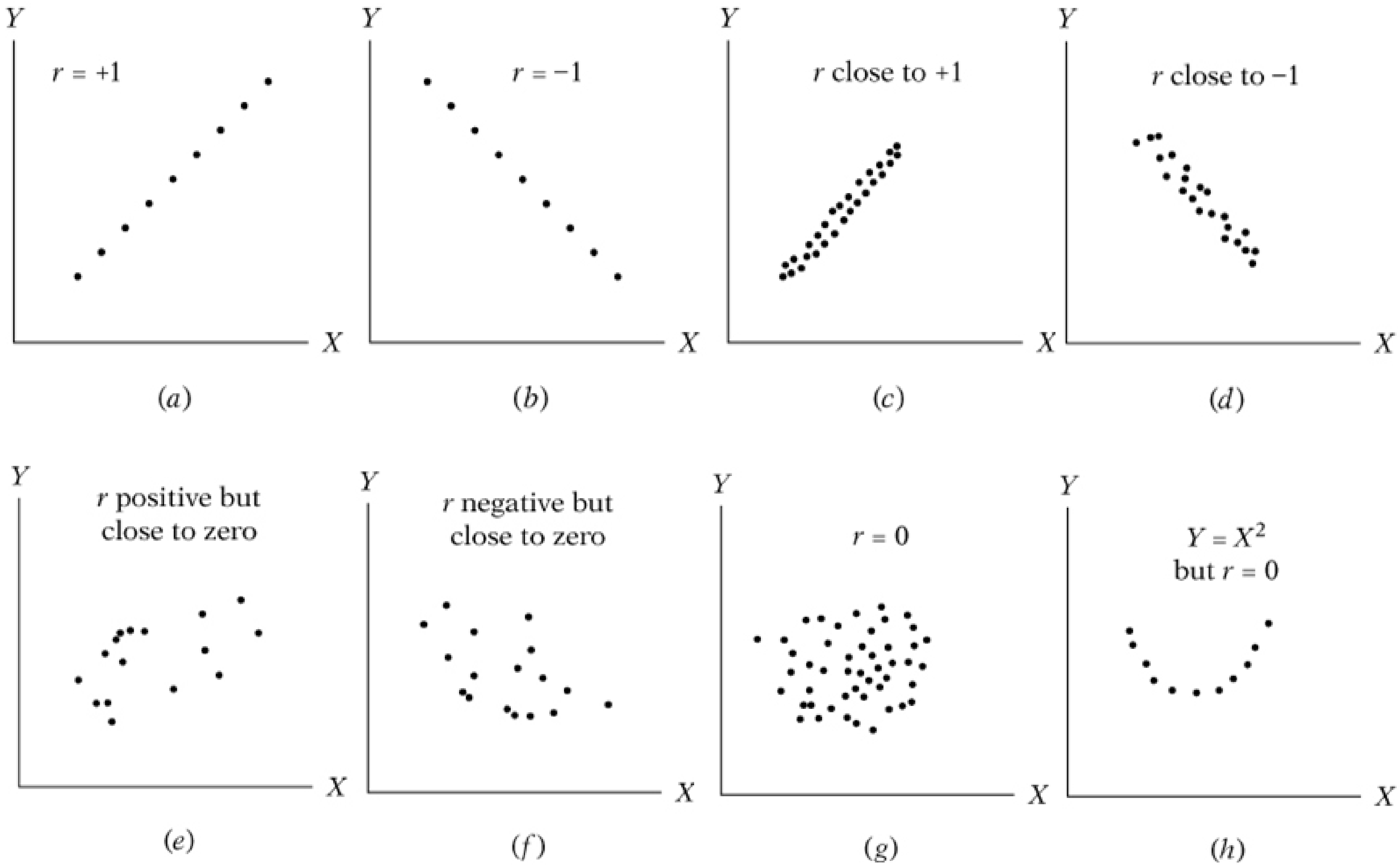
$$r = \frac{\sum x_i y_i}{\sqrt{\left(\sum x_i^2\right)\left(\sum y_i^2\right)}}$$

Correlation Coefficient r

Properties of r

1. Can be positive and negative.
2. Lies between -1 and $+1$.
3. Symmetric: $r_{XY} = r_{YX}$
4. Independent of the origin and scale.
5. If X and Y are independent, $r = 0$. But if $r=0$, it does not mean independent.
6. Measure only linear dependence.
7. Does not necessary imply cause-and-effect relationship.

Correlation Coefficient r



Extensions of Two-Variable Linear Regression Model

1. Regression Through the Origin
2. Scaling and Units of Measurement
3. Functional Forms of Regression Models
4. Distribution of Stochastic Error Term

Regression Through the Origin

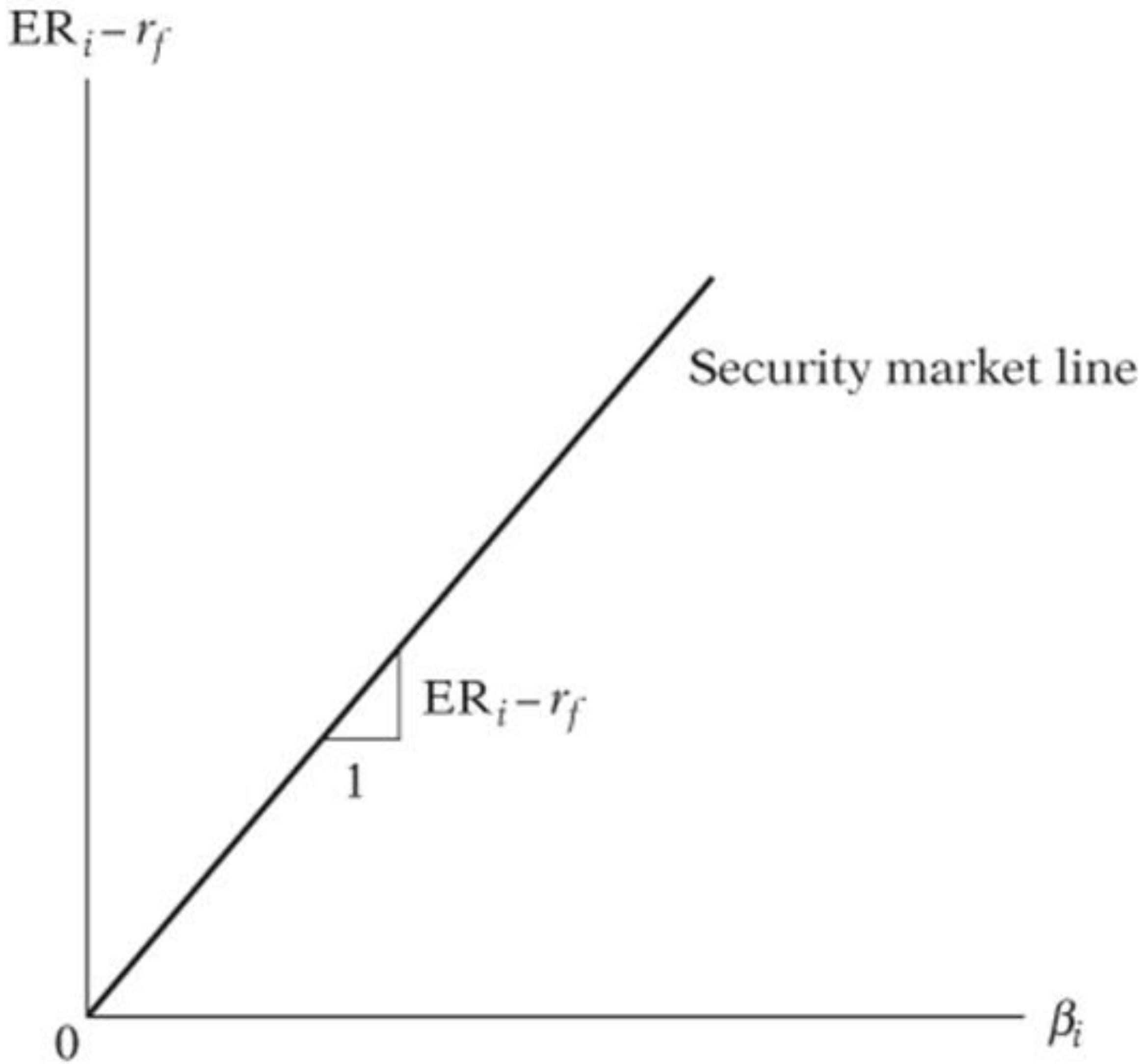
If theoretical framework suggests regression model with no intercept, the model will be *regression through the origin*.

$$Y_i = \beta_2 X_i + u_i$$

An example is Capital Asset Pricing Model (CAPM) of modern portfolio theory.

$$\left(R_i - r_f\right) = \beta_i \left(R_m - r_f\right)$$

Regression Through the Origin



Regression Through the Origin

CAPM Regression Model

$$R_i - r_f = \beta_i \left(R_m - r_f \right) + u_i$$

Market Model $$R_i - r_f = \alpha_i + \beta_i \left(R_m - r_f \right) + u_i$$

If CAPM holds, α_i is expected to be zero.

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i$$

Regression Through the Origin

<u>Model with intercept</u>	<u>Model without intercept</u>
$\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$	$\hat{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$
$\operatorname{var}\left(\hat{\beta}_2\right) = \dfrac{\sigma^2}{\sum x_i^2}$	$\operatorname{var}\left(\hat{\beta}_2\right) = \dfrac{\sigma^2}{\sum X_i^2}$
$\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-2}$	$\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-1}$

Regression Through the Origin

r^2 for Regression-through-Origin Model

$$Raw\ r^2 = \frac{\sum (X_i Y_i)^2}{\sum X_i^2 \sum Y_i^2}$$

Scaling and Units of Measurement

Let $$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

Define $$Y_i^* = w_1 Y_i$$

$$X_i^* = w_2 X_i$$

Regression using Y^* and X^*

$$Y_i^* = \hat{\beta}_i^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^*$$

Transformation from Y and X to Y^* and X^* scale does not affect the properties of the OLS estimators.

Scaling and Units of Measurement

<u>Interpretation</u>

Since β_2 is the rate of change, it is measured in units of the ratio.

$$\frac{Units\ of\ the\ Dependent\ Variable\ Y}{Units\ of\ the\ Explanatory\ Variable\ X}$$

Functional Forms of Regression Models

Some commonly used regression models, which may be nonlinear in the variables but are linear in parameters.

1. Log-linear model
2. Semilog models
3. Reciprocal models

These models can be transformed to be a linear regression model.

Log-Linear Model: How to Measure Elasticity

Consider the following model, exponential regression model:

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$$

Linearize model by take logarithm:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$$

or $$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i$$

where $$\alpha = \ln \beta_1$$

Log-Linear Model: How to Measure Elasticity

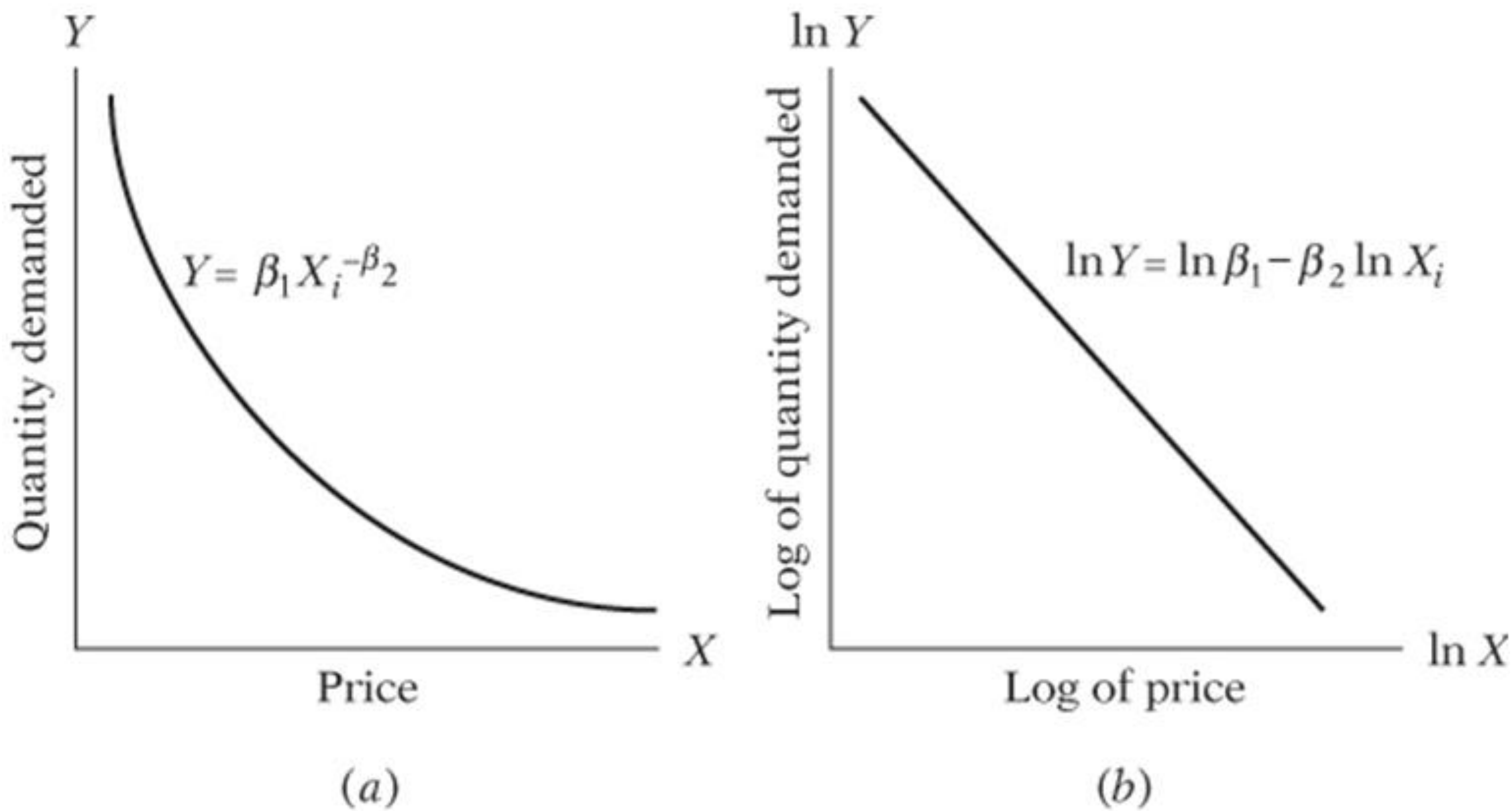
$$\beta_2 = \frac{Change\ in\ \ln Y}{Change\ in\ \ln X}$$

$$= \frac{Relative\ change\ in\ Y}{Relative\ change\ in\ X}$$

$$= \frac{\Delta Y / Y}{\Delta X / X} = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}$$

β_2 = *Elasticity of Y with respect to X*

Log-Linear Model: How to Measure Elasticity



Semilog Models: Log-Lin Models

Suppose we want to find the rate of growth of real GDP.

Let: Y_t = real GDP (RGDP) at time t

Y_o = the initial value of real GDP

$$Y_t = Y_0 (1+r)^t$$

Linearize this model: $\ln Y_t = \ln Y_0 + t \ln(1+r)$

Let $\beta_1 = \ln Y_0$ and $\beta_2 = \ln(1+r)$

$$\ln Y_t = \beta_1 + \beta_2 t$$

Semilog Models: Log-Lin Models

Regression model:

$$\ln Y_t = \beta_1 + \beta_2 t + u_t$$

$$\beta_2 = \frac{\textit{Relative change in regressand}}{\textit{Absolute change in regressor}}$$

$\beta_2 = \ln(1+r)$ = instantanesous rate of growth

$$anti\ln\beta_2 = (1+r)$$

r = compound rate of growth

Semilog Models: Lin-Log Models

Absolute change in Y for a percent change in X

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

$$\beta_2 = \frac{Absolute\ change\ in\ Y}{Relative\ change\ in\ X}$$

$$= \frac{\Delta Y}{\Delta X / X}$$

Nature of Stochastic Error Term

(1) $Y_i = \beta_1 X_i^{\beta_2} u_i \longrightarrow \ln Y_i = \alpha + \beta_2 \ln X_i + \ln u_i$

(2) $Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \longrightarrow \ln Y_i = \alpha + \beta_2 \ln X_i + u_i$

(3) $Y_i = \beta_1 X_i^{\beta_2} + u_i \longrightarrow \ln Y_i = \ln\left(\beta_1 X_i^{\beta_2} + u_i\right)$

where $\alpha = \ln \beta_1$

Equation (1) and (2) are linear-in-parameter.

Equation (3) is nonlinear-in-parameter.

(1) $\ln u_i \sim N\left(0, \sigma^2\right)$ Log-normal Distribution

(2) $u_i \sim N\left(0, \sigma^2\right)$ Normal Distribution