

Methodology of Econometrics

1. Statement of theory or hypothesis
2. Specification of mathematical model of the theory
3. Specification of econometric model of theory
4. Obtaining the data
5. Estimation of the parameters of the econometric model
6. Hypothesis testing
7. Forecasting or prediction
8. Using model for control or policy purposes

Statement of theory or hypothesis

Keynes:

Marginal propensity to consume (MPC), the rate of change of consumption for a unit change in income is greater than zero but less than one.

Specification of mathematical model of the theory

Keynesian consumption function:

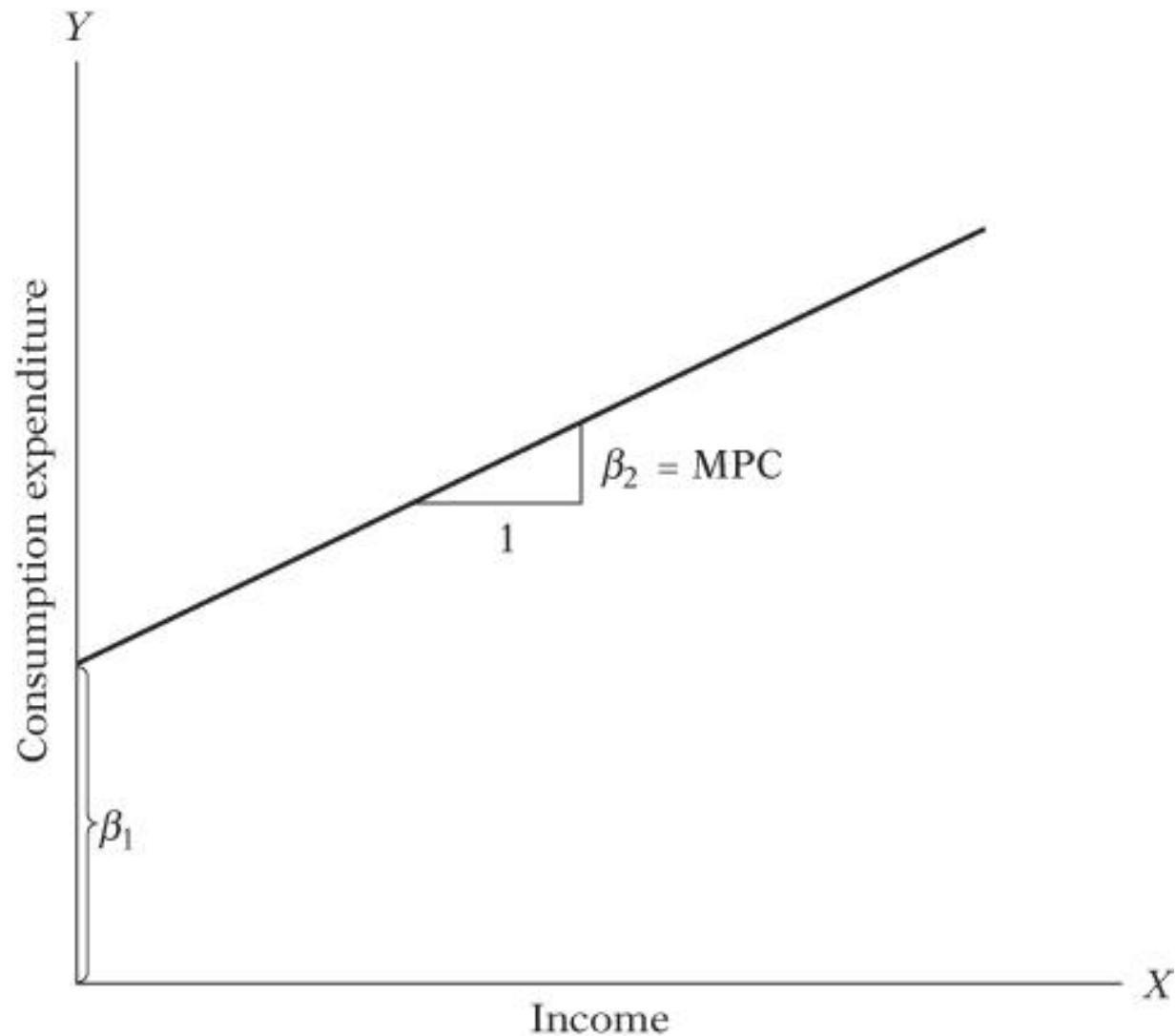
$$Y = \beta_1 + \beta_2 X \quad 0 < \beta_2 < 1$$

where Y = consumption expenditure

X = income

β_i = parameters $i = 1, 2$

Specification of mathematical model of the theory



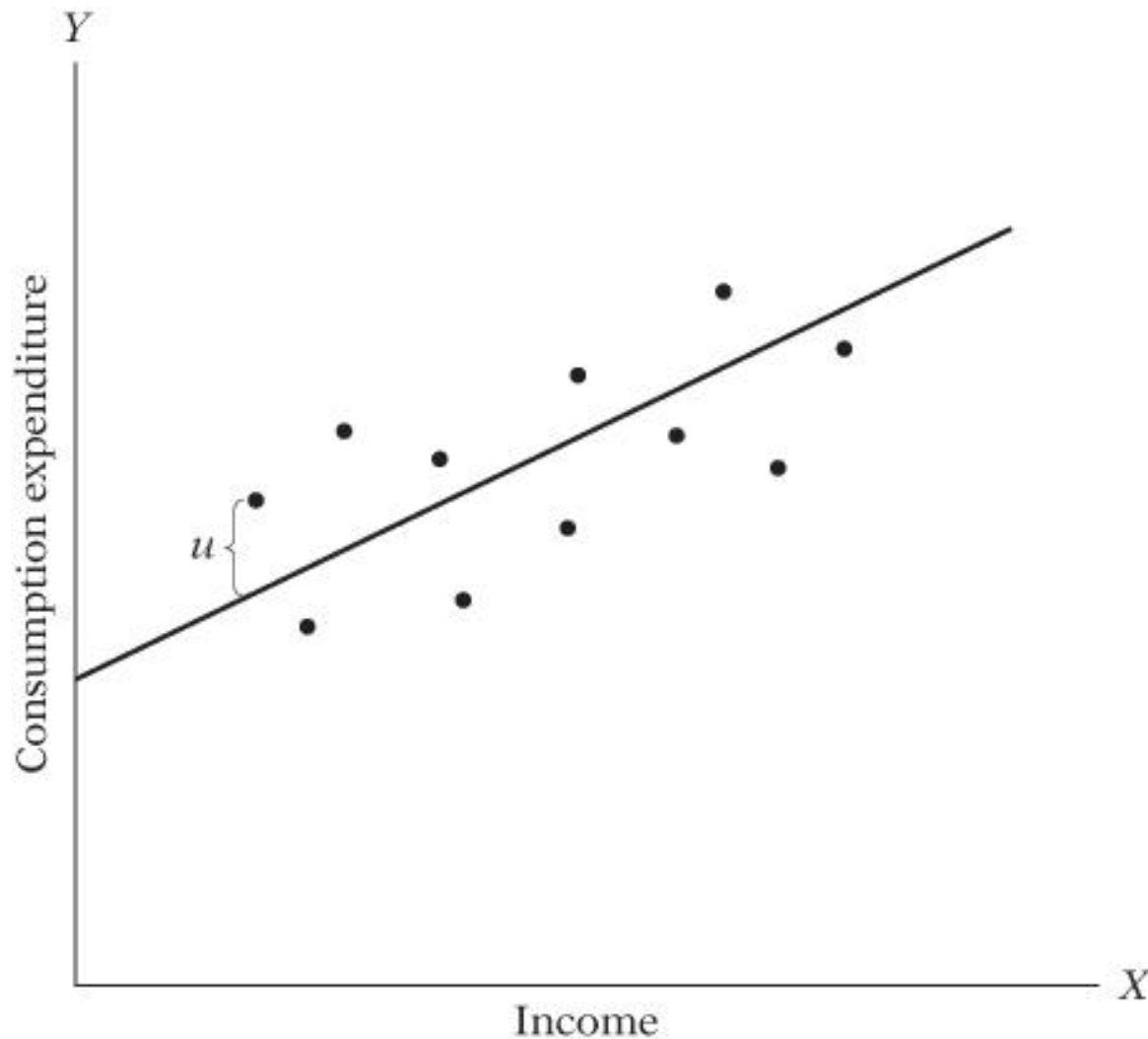
Specification of Econometric model of the theory

Transform consumption function to econometric regression model:

$$Y = \beta_1 + \beta_2 X + u$$

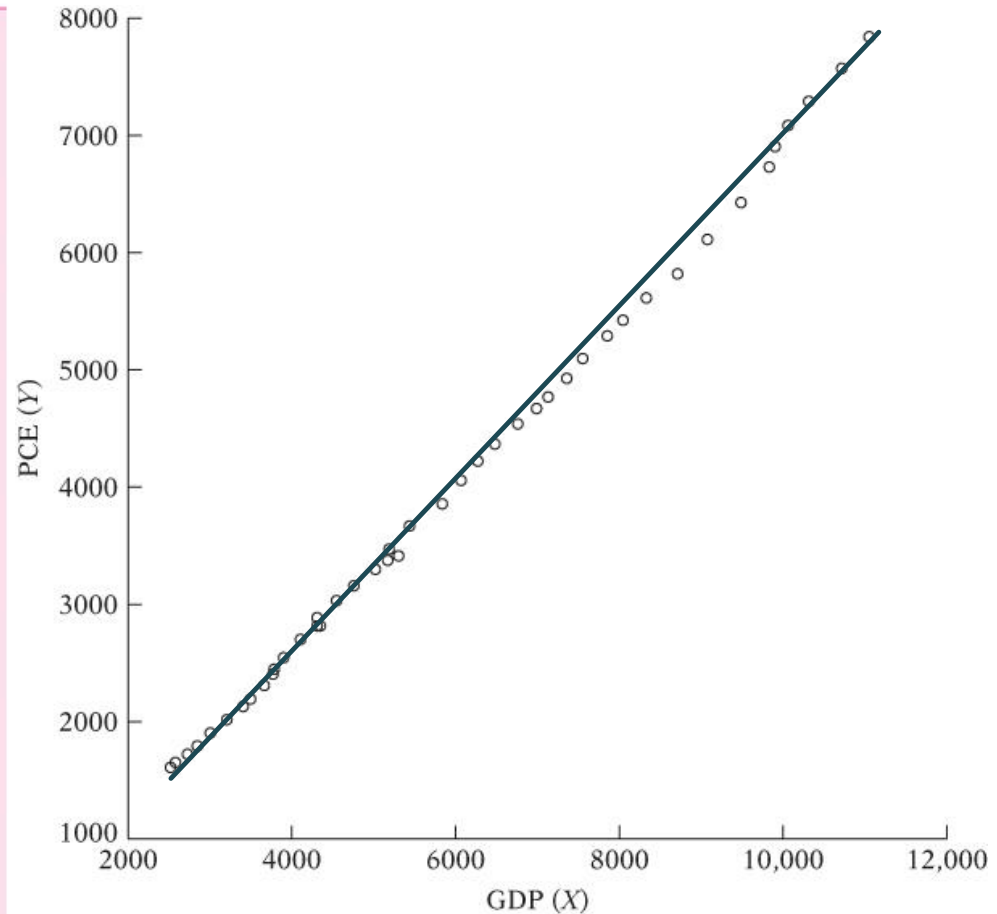
where u = disturbance or error term
a random (stochastic) variable

Specification of Econometric model of the theory



Obtaining the data

Year	PCE(Y)	GDP(X)
1960	1597.4	2501.8
1961	1630.3	2560.0
1962	1711.1	2715.2
1963	1781.6	2834.0
1964	1888.4	2998.6
1965	2007.7	3191.1
1966	2121.8	3399.1
1967	2185.0	3484.6
1968	2310.5	3652.7
1969	2396.4	3765.4
1970	2451.9	3771.9
1971	2545.5	3898.6
1972	2701.3	4105.0
1973	2833.8	4341.5
1974	2812.3	4319.6
1975	2876.9	4311.2
1976	3035.5	4540.9
1977	3164.1	4750.5
1978	3303.1	5015.0
1979	3383.4	5173.4
1980	3374.1	5161.7
1981	3422.2	5291.7
1982	3470.3	5189.3
1983	3668.6	5423.8
1984	3863.3	5813.6
1985	4064.0	6053.7
1986	4228.9	6263.6
1987	4369.8	6475.1
1988	4546.9	6742.7
1989	4675.0	6981.4
1990	4770.3	7112.5
1991	4778.4	7100.5
1992	4934.8	7336.6
1993	5099.8	7532.7
1994	5290.7	7835.5
1995	5433.5	8031.7
1996	5619.4	8328.9
1997	5831.8	8703.5
1998	6125.8	9066.9
1999	6438.6	9470.3
2000	6739.4	9817.0
2001	6910.4	9890.7
2002	7099.3	10048.8
2003	7295.3	10301.0
2004	7577.1	10703.5
2005	7841.2	11048.6



Estimation of the parameters of the econometric model

$$\hat{Y} = -299.5913 + 0.7218X$$

$$MPC = 0.7218$$

Hypothesis testing

Choose appropriated statistical method to test hypothesis, known as statistical inference.

Forecasting or Prediction

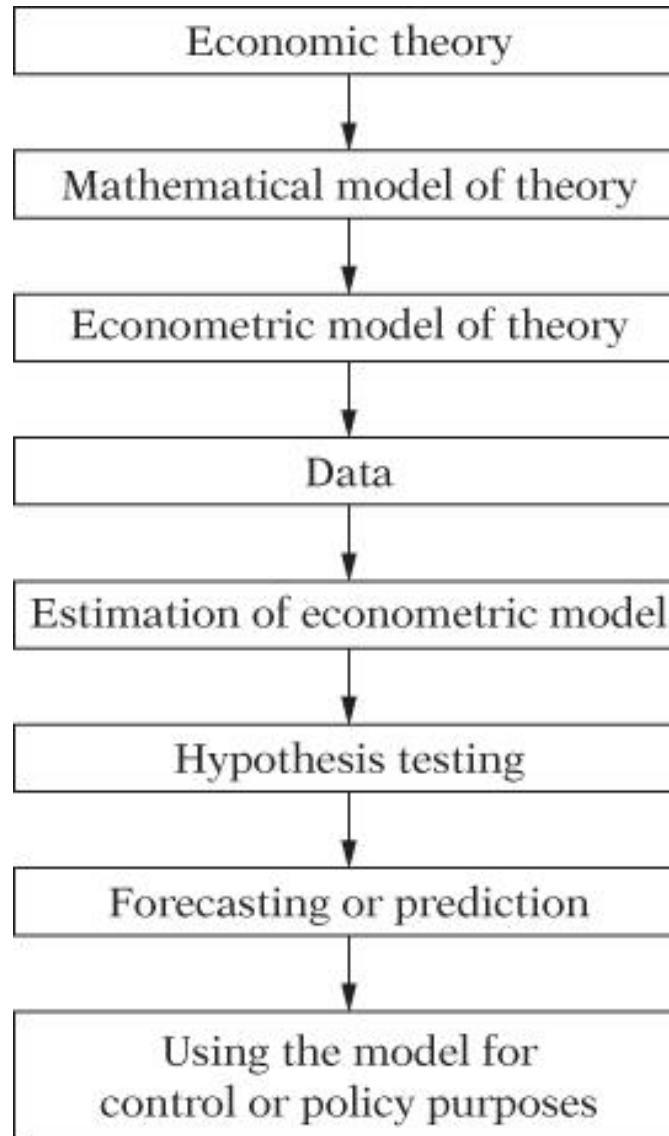
If income (or GDP) is expected to be
11,319.4 (billions)

Consumption should be

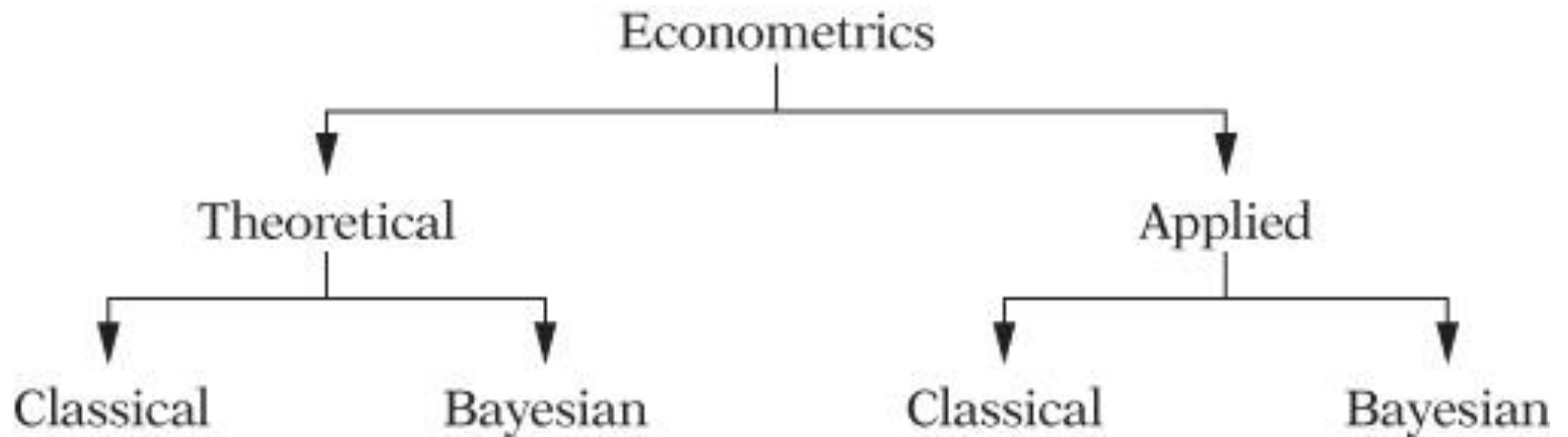
$$\begin{aligned}\hat{Y} &= -299.5913 + 0.7218(11319.4) \\ &= 7870.7516\end{aligned}$$

Using model for policy purposes

Econometric Modeling



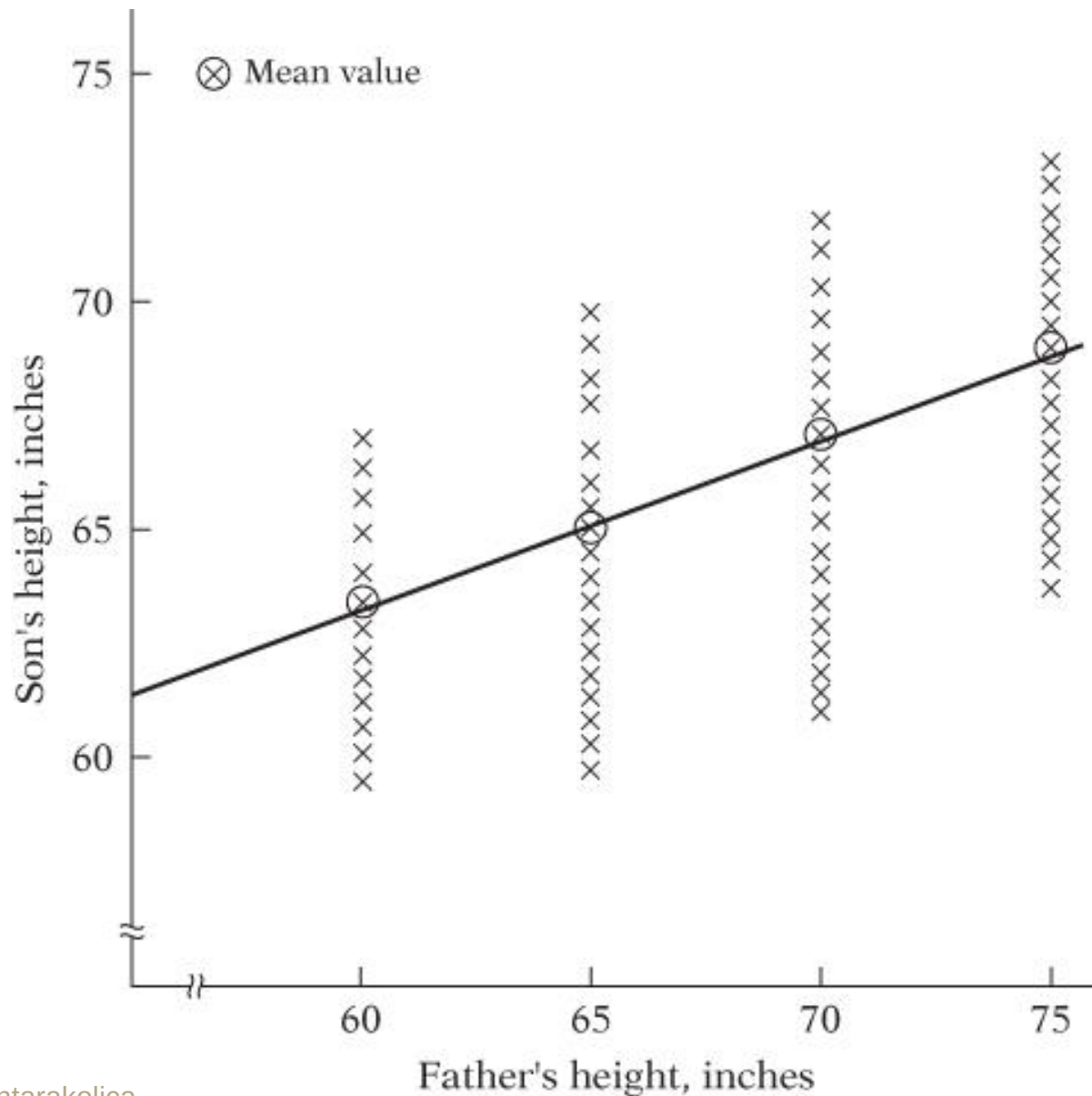
Types of Econometrics



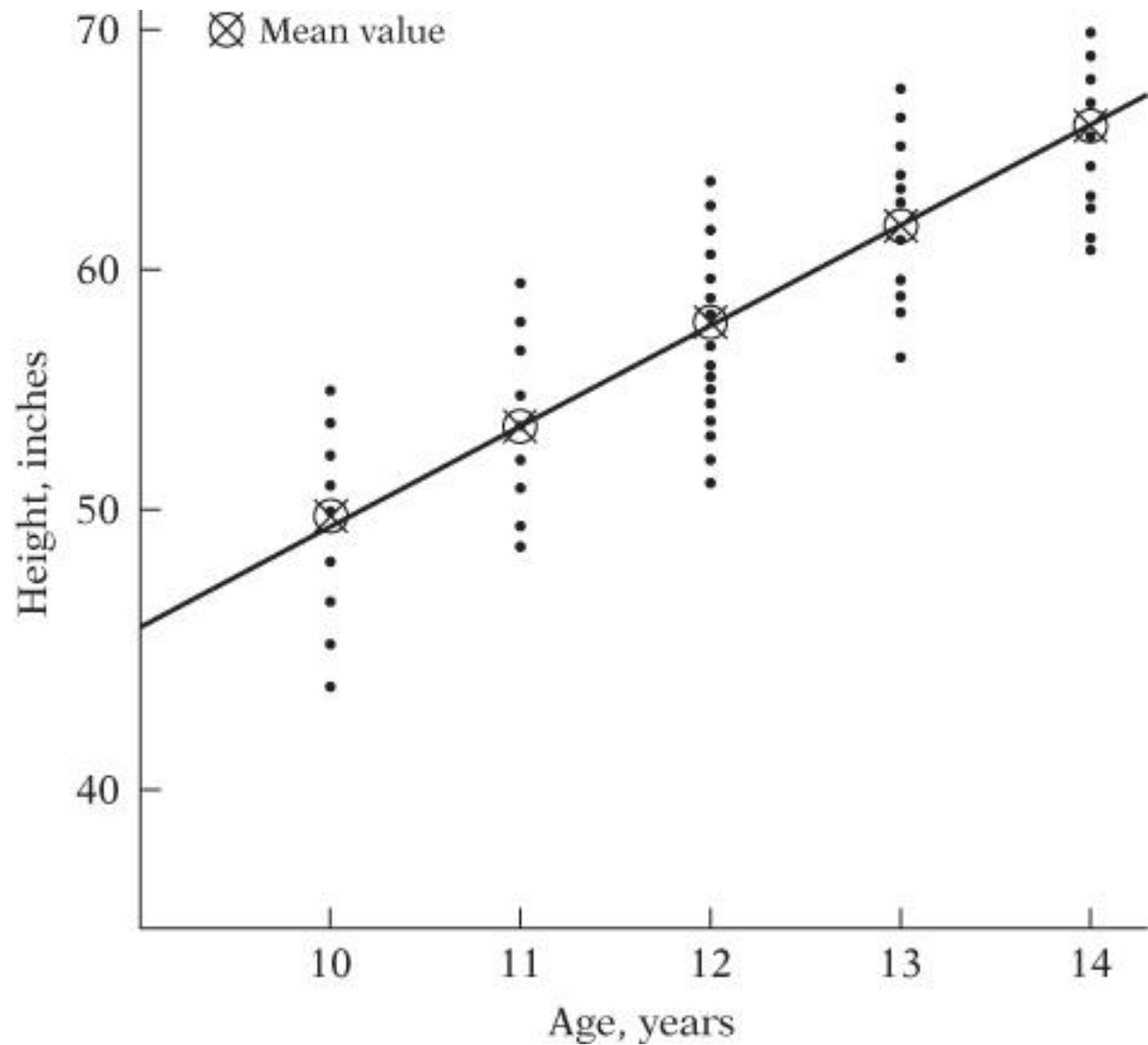
Nature of Regression Analysis

Regression analysis is concerned with the study of the dependence of one variable, the *dependent variable*, on one or more other variables, the *explanatory variables*, with a view to estimating and/or predicting the (population) mean of or average value of the former in terms of the known or fixed (in repeated sampling) values of the latter.

Examples



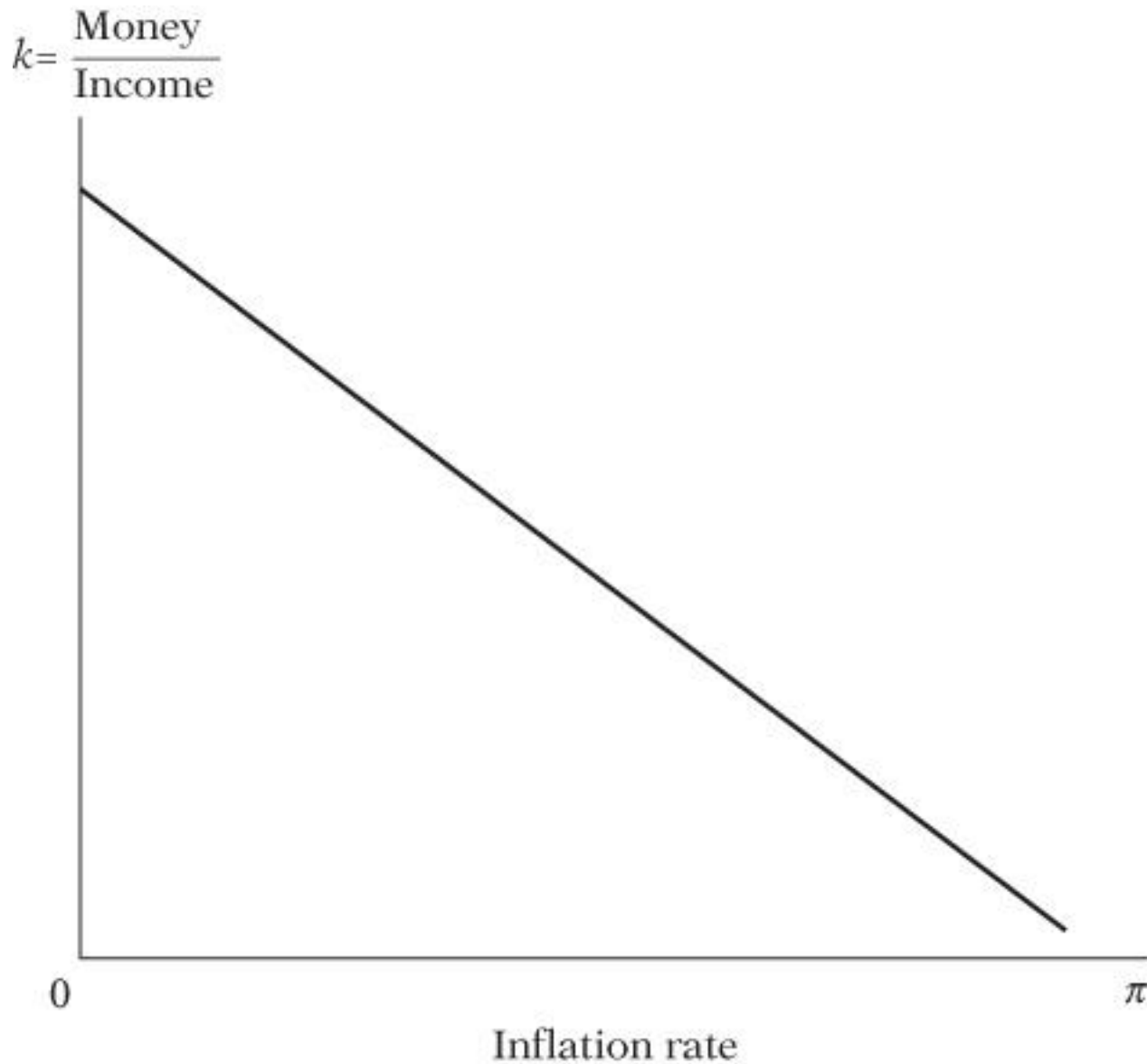
Examples



Examples



Examples



Statistical vs Deterministic Relationship

Regression vs Causation

Regression vs Correlation

Terminology and Notation

Y

Dependent Variable

Explained Variable

Predictand

Regressand

Response

Endogenous

X

Explanatory Variable

Independent Variable

Predictor

Regressor

Stimulus or Control Variable

Exogenous

Types of Data

Qualitative Data:

Nonnumerical Data

Quantitative Data:

Numerical Data

- Time Series Data
- Cross-sectional Data
- Pooled Data
- Panel Data

Accuracy of Data

Selectivity Bias

Two Variable Regression Analysis

Some Basic Ideas

Hypothetical Example

Concept of Population Regression Function

Meaning of Term “Linear”

Stochastic Specification of PRF

Significance of the Stochastic Disturbance
Term

Sample Regression Function (SRF)

Hypothetical Example

A hypothetical Country:

Total population of 60 families.

Y = Weekly family consumption expenditure

X = Disposable family income

Divide 60 families into 10 groups of same income

Hypothetical Example

$Y \downarrow \quad X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y , \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y , $E(Y X)$	65	77	89	101	113	125	137	149	161	173

Conditional Probability: $\Pr(Y = 55 | X = 80) = \frac{1}{5} = 0.2$

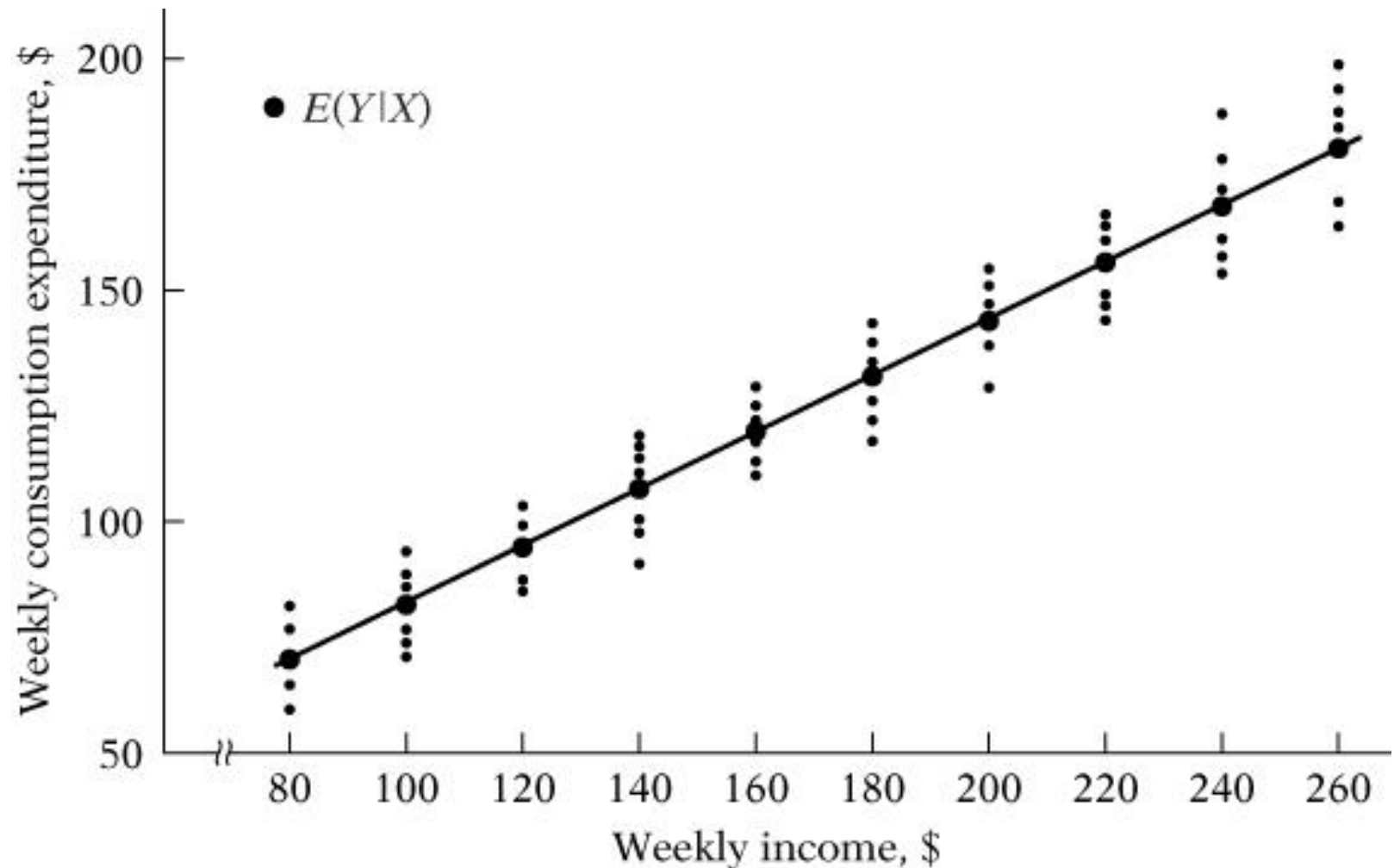
Conditional Mean:

$$E(Y | X = 80) = 55(0.2) + 60(0.2) + 65(0.2) + 70(0.2) + 75(0.2) = 65$$

Hypothetical Example

$X \rightarrow$ $p(Y X_i)$ ↓	80	100	120	140	160	180	200	220	240	260
Conditional probabilities $p(Y X_i)$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	—	$\frac{1}{6}$	—	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	—	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	—	—	—	$\frac{1}{7}$	—	—	—	$\frac{1}{7}$	—	$\frac{1}{7}$
Conditional means of Y	65	77	89	101	113	125	137	149	161	173

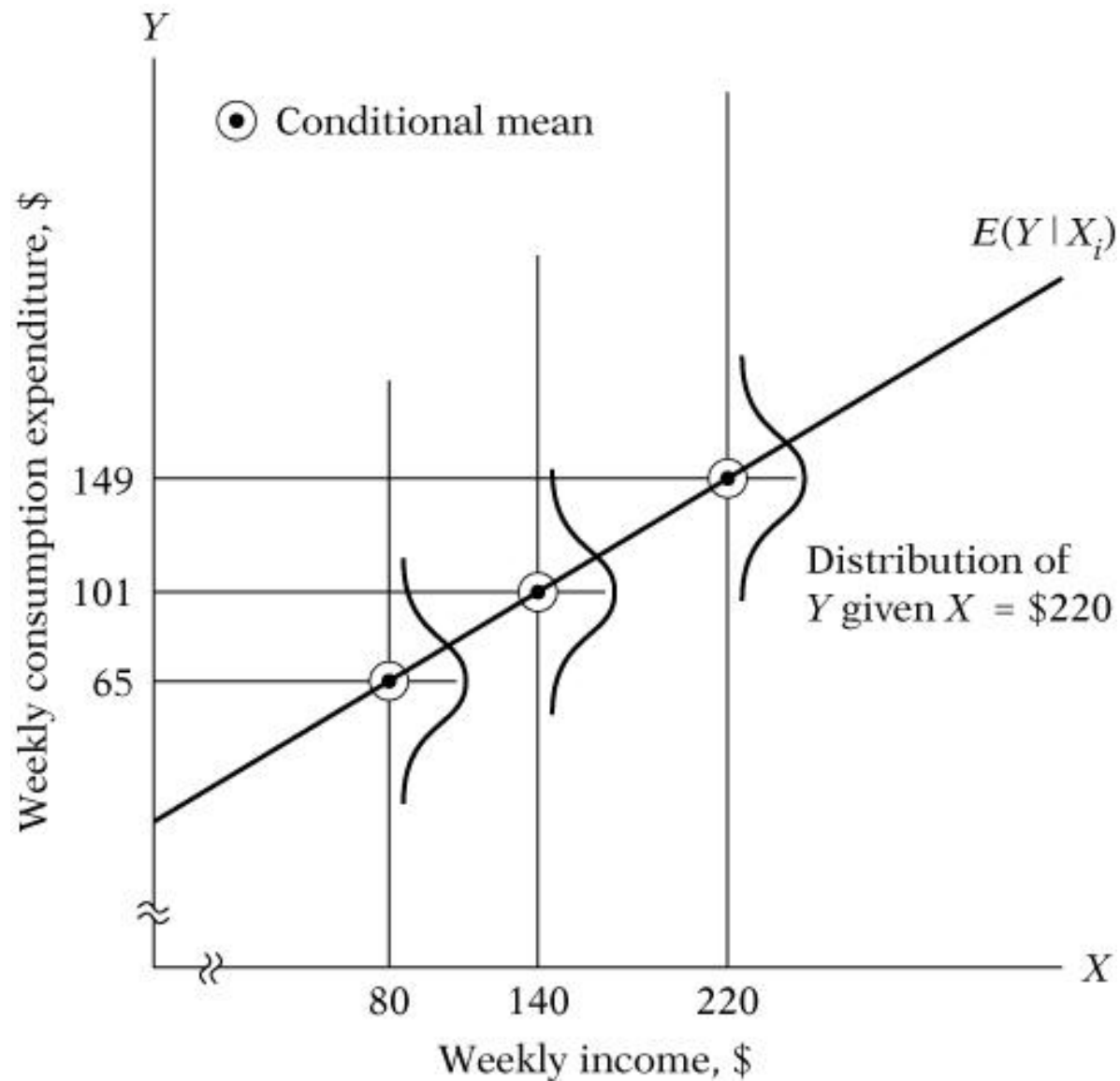
Hypothetical Example



Hypothetical Example

$X \rightarrow$ $p(Y X_i)$ $\downarrow$	80	100	120	140	160	180	200	220	240	260
Conditional probabilities $p(Y X_i)$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	—	$\frac{1}{6}$	—	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	—	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	—	—	—	$\frac{1}{7}$	—	—	—	$\frac{1}{7}$	—	$\frac{1}{7}$
Conditional means of Y	65	77	89	101	113	125	137	149	161	173

Hypothetical Example



Concept of Population Regression Function

Conditional Mean $E(Y|X_i)$ is function of X_i

$$E(Y|X_i) = f(X_i)$$

Population Regression Function $E(Y|X_i)$ is linear function of X_i

$$E(Y|X_i) = \beta_1 + \beta_2 X_i$$

where β_1 and β_2 are unknown but fixed parameters or regression coefficient.

Meaning of Term “Linear”

Linearity in Variables

Linear function: $E(Y | X_i) = \beta_1 + \beta_2 X_i$

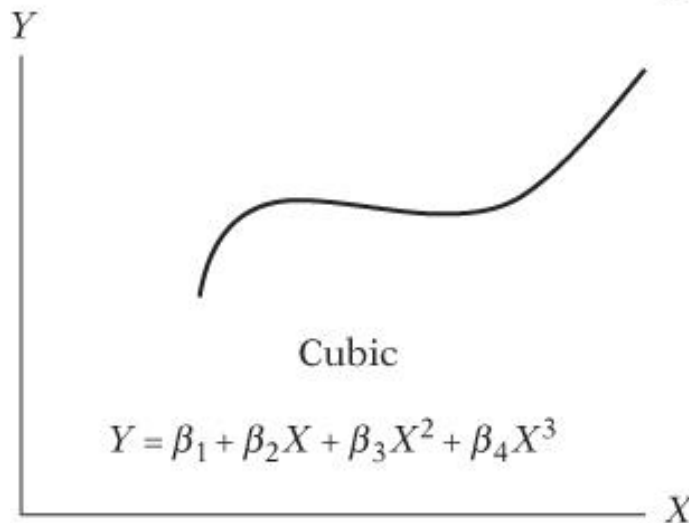
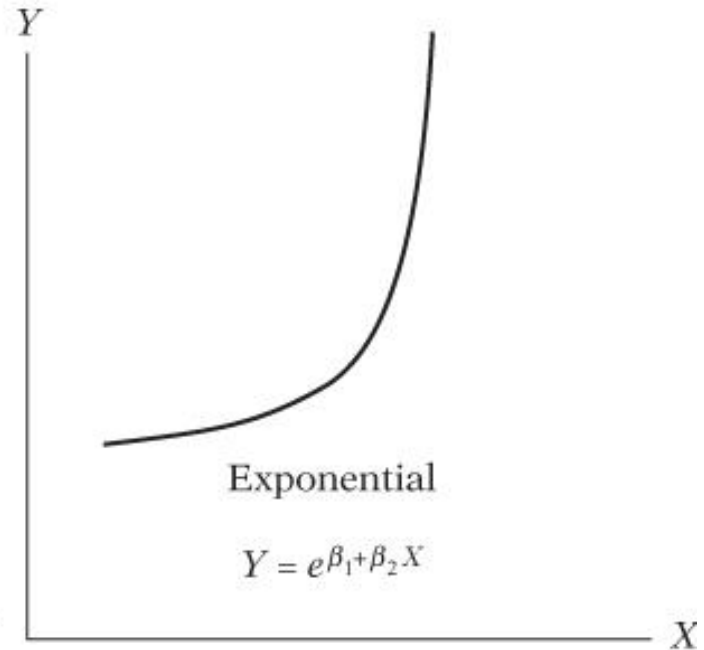
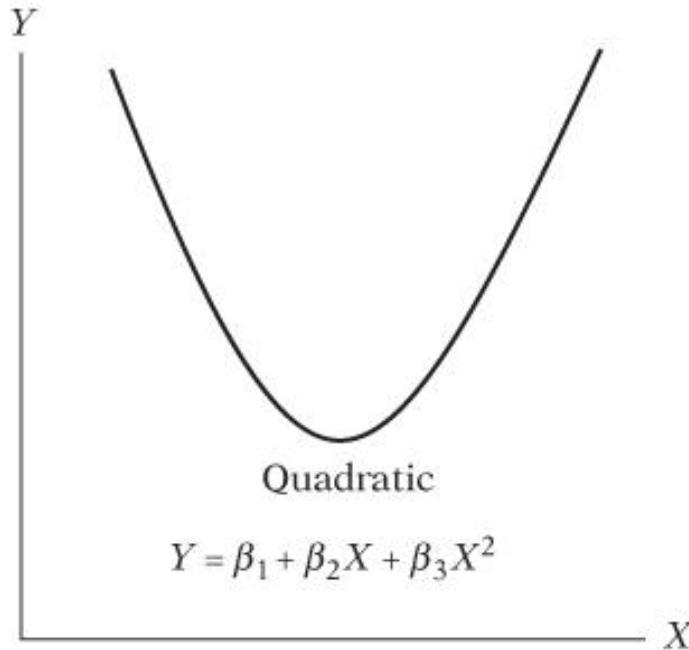
Nonlinear function: $E(Y | X_i) = \beta_1 + \beta_2 X_i^2$

Linearity in Parameters

Linear Regression: $E(Y | Z_i = X_i^2) = \beta_1 + \beta_2 Z_i$

Nonlinear in parameter: $E(Y | X_i) = \beta_1 + \sqrt{\beta_2} X_i$

Examples: Linear-in-Parameters



Examples:

$Y \downarrow \quad X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y , \$	55 60 65 70 75 – –	65 70 74 80 85 88 –	79 84 90 94 98 – –	80 93 95 103 108 113 115	102 107 110 116 118 125 –	110 115 120 130 135 140 –	120 136 140 144 145 – –	135 137 140 152 157 160 162	137 145 155 165 175 189 –	150 152 175 178 180 185 191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y , $E(Y X)$	65	77	89	101	113	125	137	149	161	173

$$Y_1 = 55 = \beta_1 + \beta_2(80) + u_1$$

$$Y_2 = 60 = \beta_1 + \beta_2(80) + u_2$$

$$Y_3 = 65 = \beta_1 + \beta_2(80) + u_3$$

$$Y_4 = 70 = \beta_1 + \beta_2(80) + u_4$$

$$Y_5 = 75 = \beta_1 + \beta_2(80) + u_5$$

Deviation of Y from its conditional mean

$$Y_i = 55, E(Y|X_i = 80) = 65$$

$$u_i = Y_i - E(Y|X_i)$$

Stochastic Specification of PRF

$$u_i = Y_i - E(Y|X_i)$$

or

$$Y_i = E(Y|X_i) + u_i$$

where:

Deviation u_i is unobservable random variable
or stochastic disturbance

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

where: $E(u_i|X_i) = 0$



Significance of the Stochastic Disturbance Term

1. Vagueness of theory
2. Unavailability of Data
3. Core variables vs. Control variables
4. Intrinsic randomness in human behavior
5. Poor proxy variables
6. Principle of Parsimony
7. Wrong functional form

Sample Regression Function (SRF)

Observing sample from population, we might get sample that does not represent population.

TABLE 2.4

A Random Sample from the Population of Table 2.1

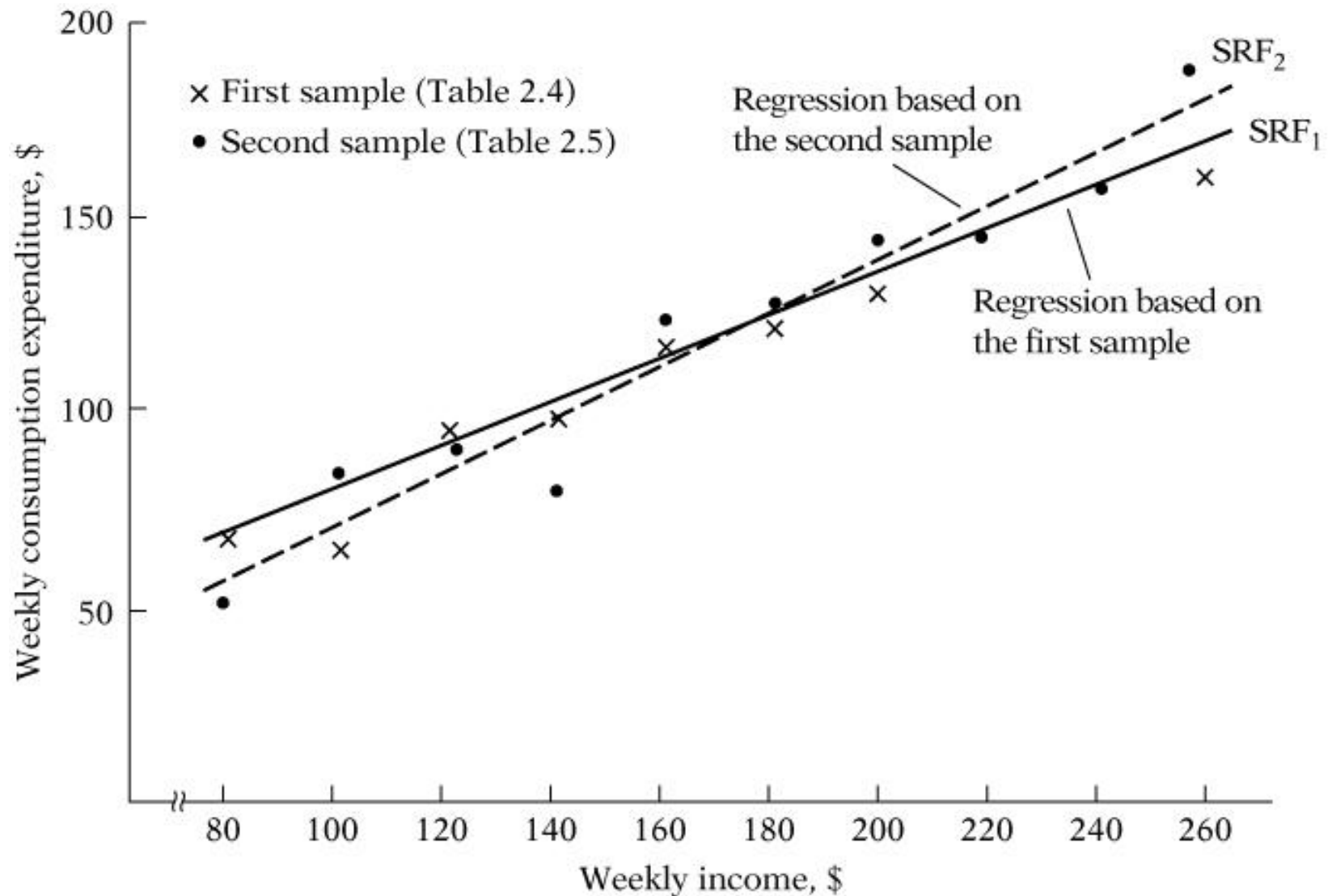
Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

TABLE 2.5

Another Random Sample from the Population of Table 2.1

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

Sample Regression Function (SRF)



Sample Regression Function (SRF)

Sample regression line or Sample regression function (SRF): $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

where $\hat{Y}_i$ = estimator of $E(Y|X_i)$

$\hat{\beta}_1$ = estimator of β_1

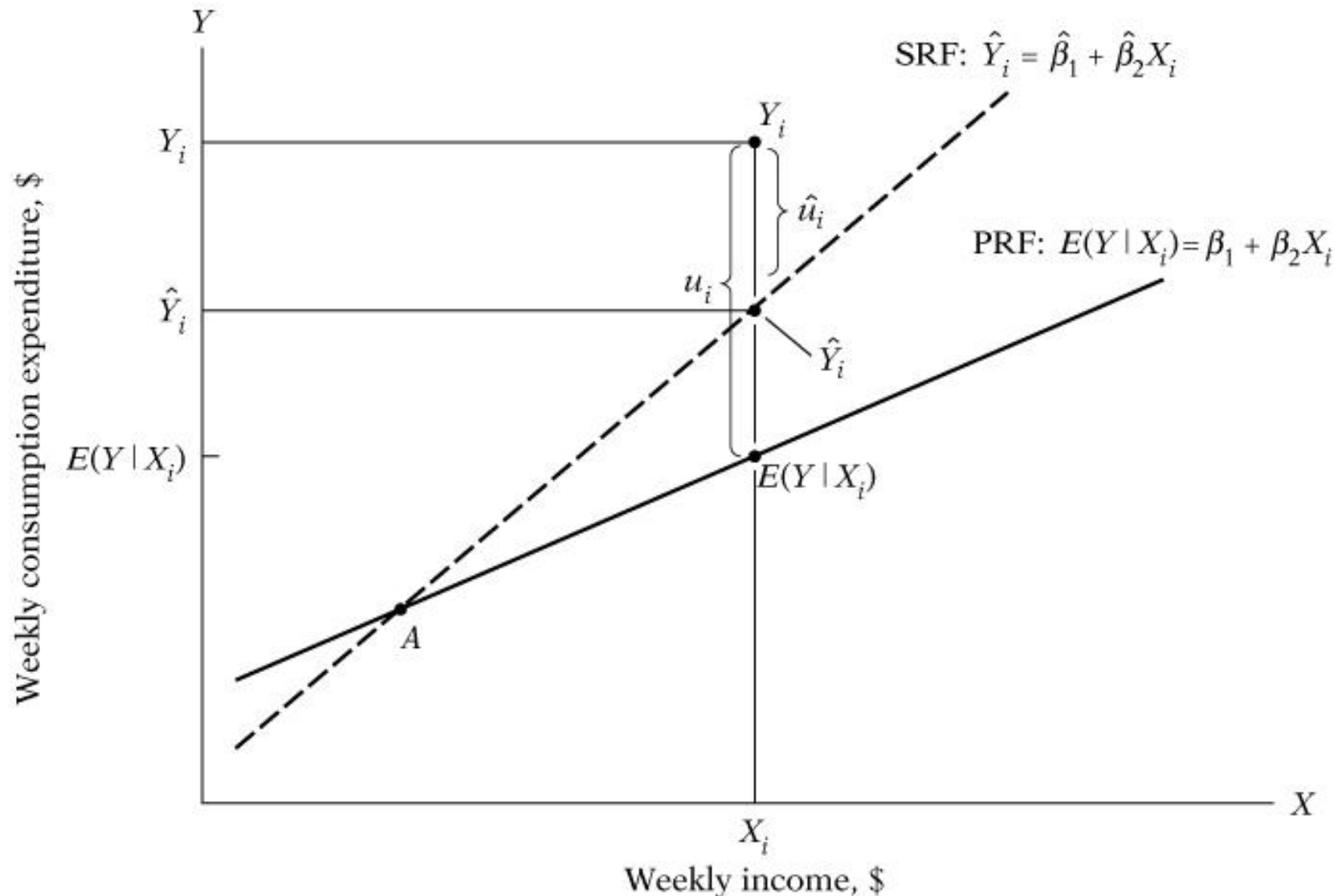
$\hat{\beta}_2$ = estimator of β_2

SRF: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$

PRF: $E(Y | X_i) = \beta_1 + \beta_2 X_i$

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

Sample Regression Function (SRF)



Two Variable Regression Model

Problems of Estimation

1. Method of Ordinary Least Squares
2. Classical Linear Regression Model (CLRM):
Assumptions Underlying the Method of Least Squares
3. Standard Errors of Least-Squares Estimates
4. Properties of Least-Squares Estimators:
Gauss-Markov Theorem

Method of Ordinary Least Squares

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

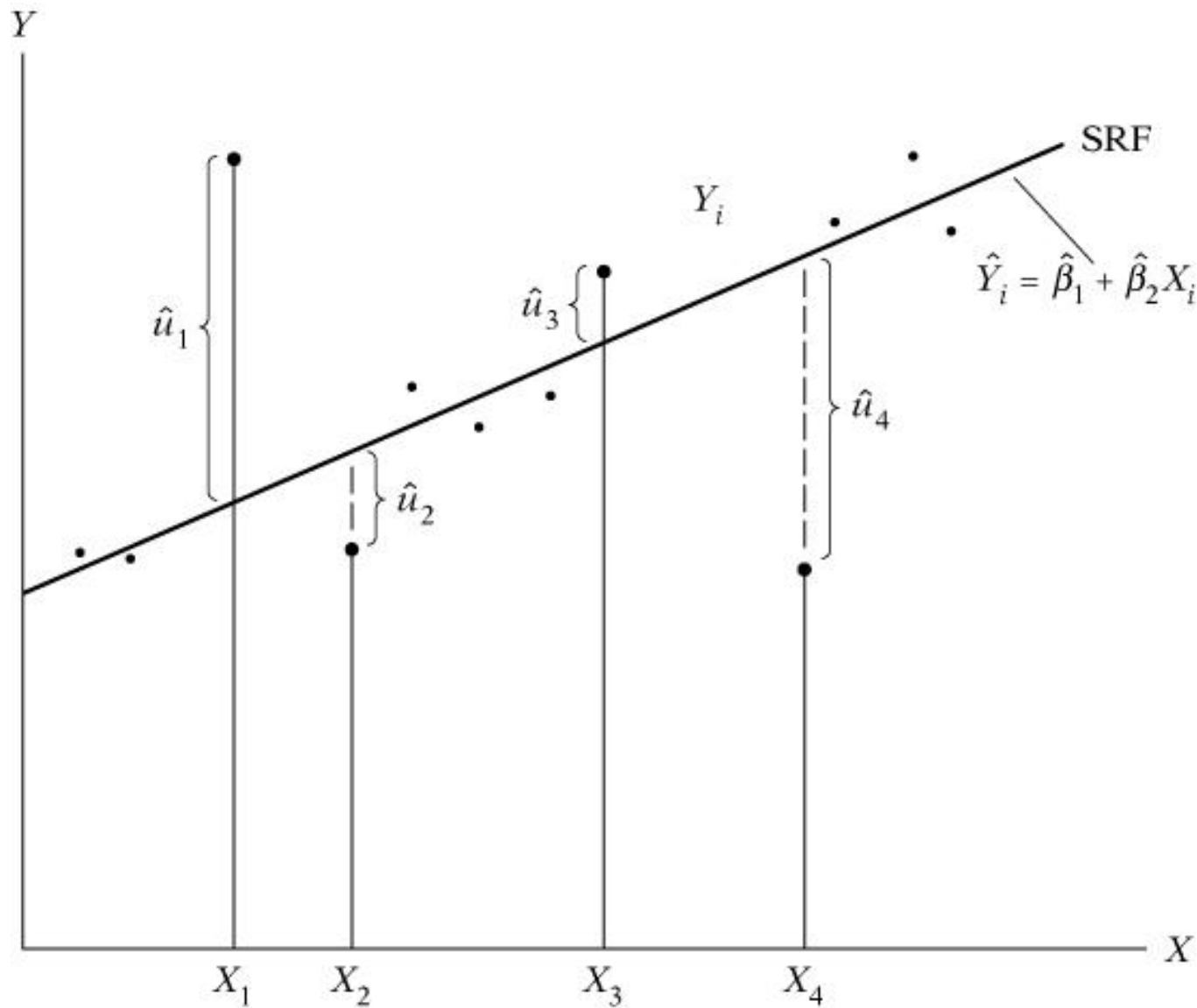
Then $\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$

OLS try to minimize sum of squares residuals

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 \rightarrow 0$$

If $\sum \hat{u}_i^2 = 0$ then $Y_i = \hat{Y}_i$

Method of Ordinary Least Squares



Method of Ordinary Least Squares

$$\text{Set } \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 0 \quad \text{then} \quad \sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$$

$$\text{Set } \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 0 \quad \text{then} \quad \sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$$

Solve these normal equations:

$$\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

Method of Ordinary Least Squares

$$\hat{\beta}_2 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum x_i y_i}{\sum x_i^2}$$

Method of Ordinary Least Squares

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \bar{Y} - \hat{\beta}_2 \bar{X}$$

Method of Ordinary Least Squares

Statistical Properties of OLS Estimators

1. Easily to compute
2. Point Estimators

Properties of Regression Line

1. Pass through the sample mean of Y and X
2. Mean of estimated $\hat{Y}_i$ equals to mean of actual Y_i
3. Mean value of residual is zero
4. Residuals are uncorrelated with predicted $\hat{Y}_i$
5. Residuals are uncorrelated with X_i

Assumptions of Least Squares

1. Linear Regression Model
2. X_i values are fixed in repeated sampling
3. Zero mean value of disturbance u_i
4. Homoscedasticity or equal variance of u_i
5. No autocorrelation between the disturbance
6. Zero covariance between u_i and X_i
7. Number of observations must be greater than number of parameters to be estimated
8. Variability in X_i values
9. Regression model is correctly specified
10. No perfect multicollinearity

Assumptions of Least Squares

1. Linear Regression Model

- Linear in parameter

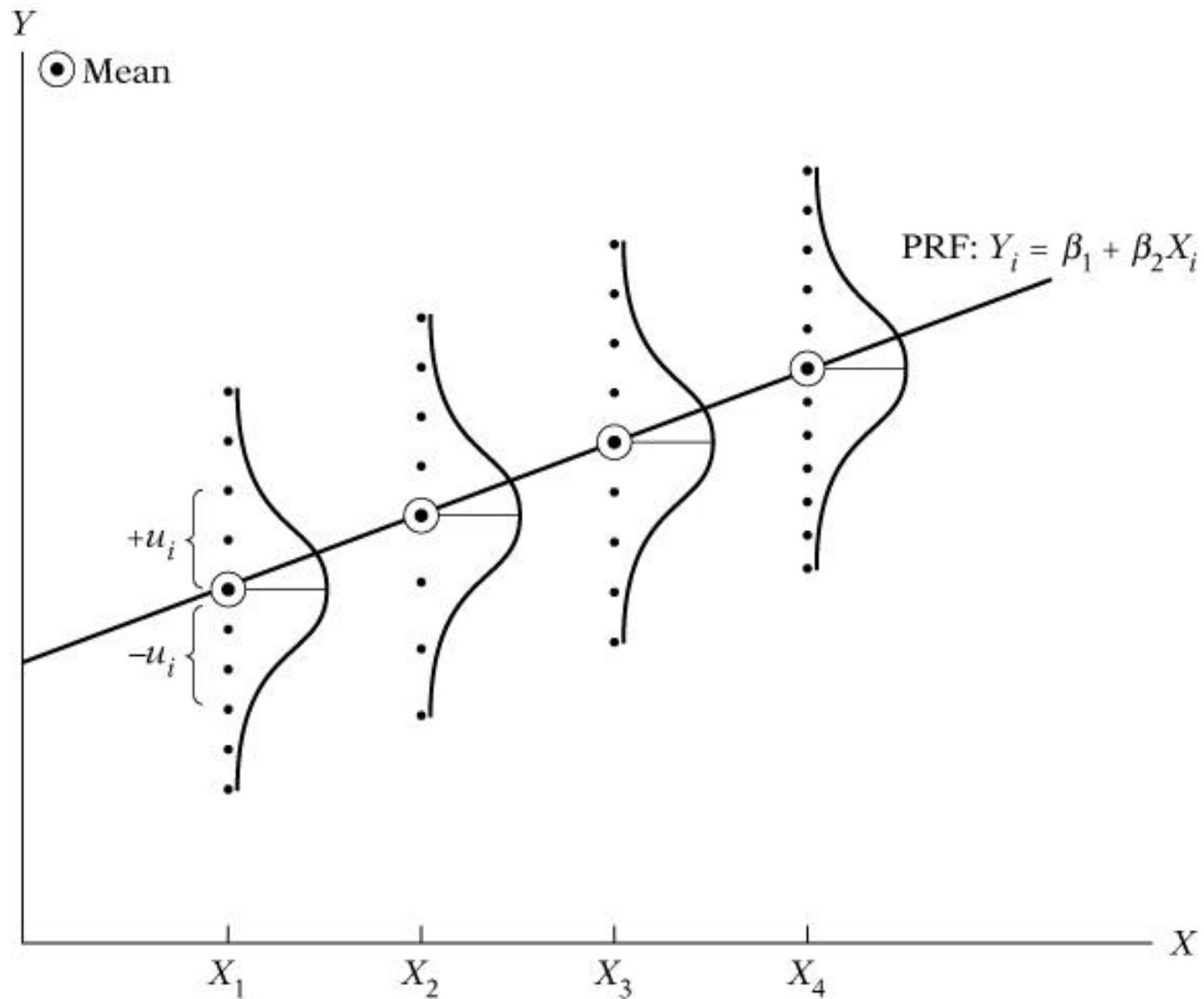
2. X values are fixed in repeated sampling

- X_i is assumed to be *nonstochastic*

3. Zero mean value of disturbance u_i

- $E(u_i | X_i) = 0$

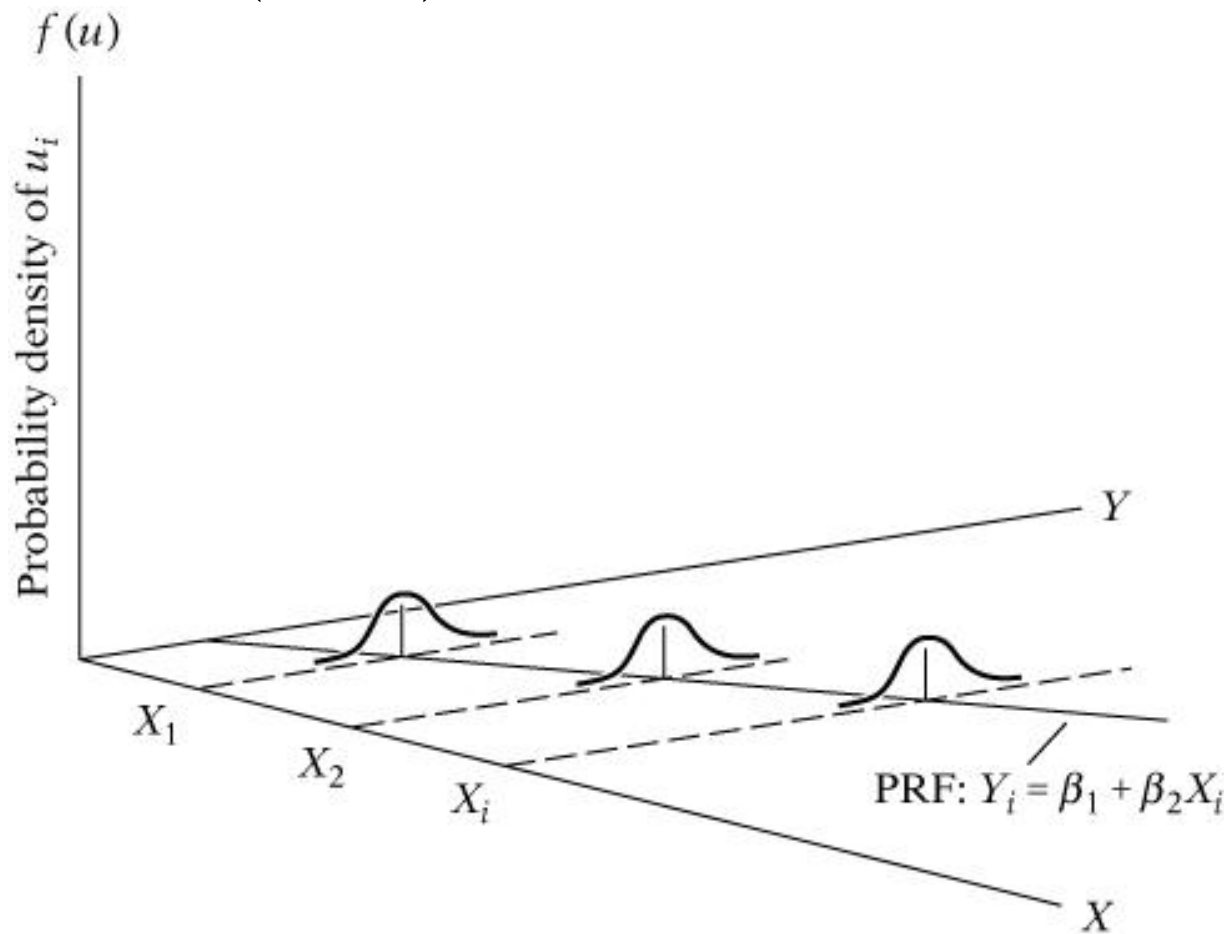
Assumptions of Least Squares



Assumptions of Least Squares

4. Homoscedasticity or equal variance of u_i

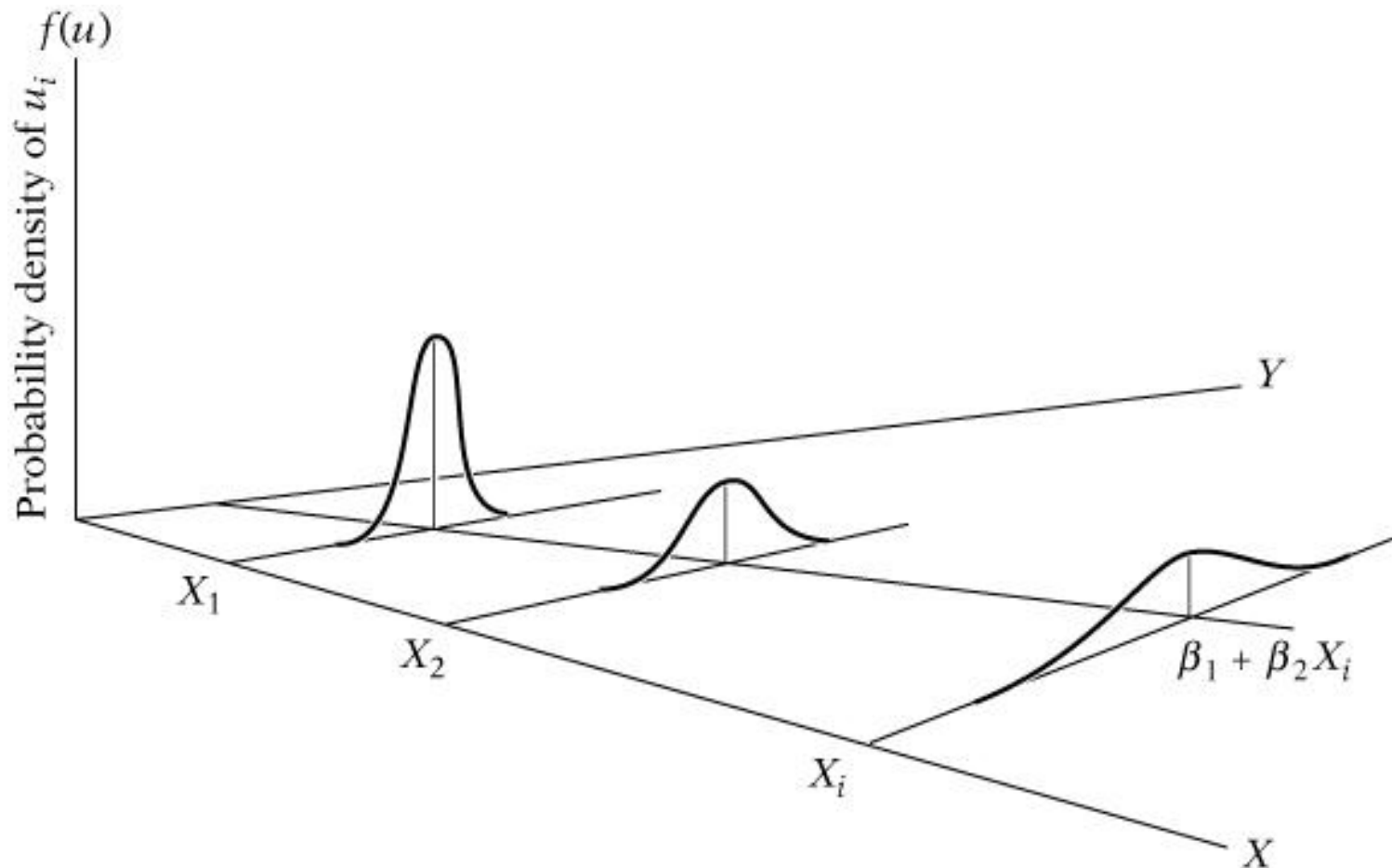
- $\text{var}(u_i | X_i) = \sigma^2$ or Constant variance



Assumptions of Least Squares

In case of violation or Heteroscedasticity

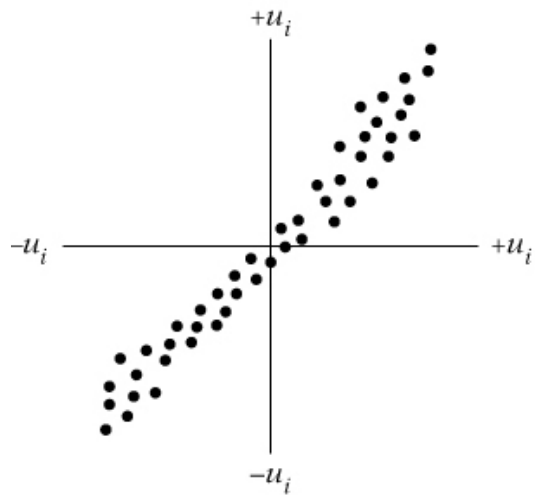
- $\text{var}(u_i | X_i) = \sigma_i^2$



Assumptions of Least Squares

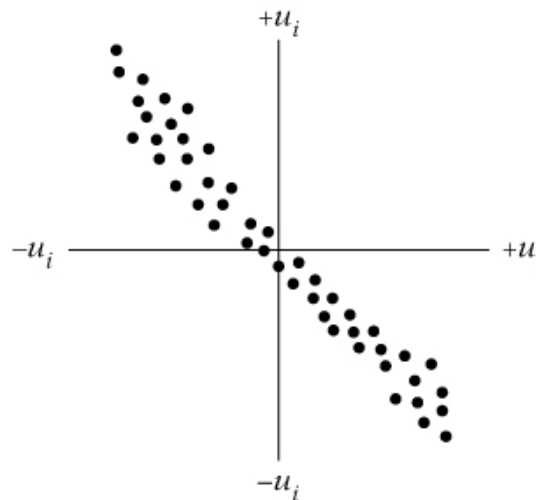
5. No autocorrelation between the disturbances

$$\text{cov}(u_i u_j | X_i, X_j) = 0$$



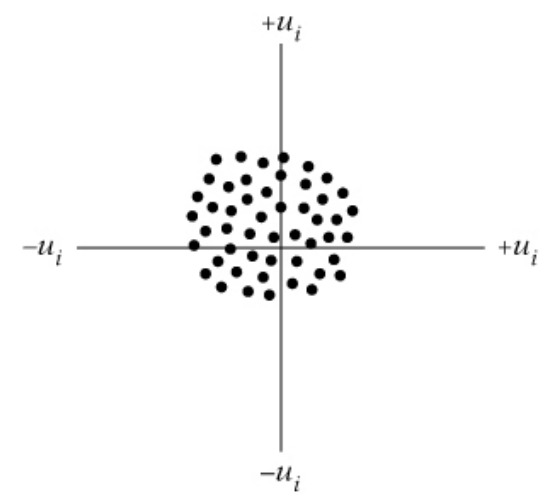
(a)

Positive Autocorrelation



(b)

Negative Autocorrelation



(c)

No Autocorrelation

Assumptions of Least Squares

6. Zero covariance between u_i and X_i

- $E(u_i X_i) = 0$ then $\text{cov}(u_i X_i) = 0$

7. Number of observations must be greater than number of parameters to be estimated

Number of observations must be greater than number of explanatory variables

8. Variability in X_i values

X_i values in given sample must not all be the same

9. Regression model is correctly specified

No specification bias or error

Assumptions of Least Squares

10. No perfect multicollinearity

There are no perfect linear relationships among the explanatory variables.

Standard Errors of Least-Squares

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$$

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

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

Properties of Least-Squares Estimator

1. Linear
2. Unbiased
3. Efficient estimator

Gauss-Markov Theorem:

Given assumptions of CLRM, the least-squares estimators, in the class of unbiased linear estimators, have minimum variance -- they are Best Linear Unbiased Estimators (BLUE).

