Kyiv
School of
Economics

Introduction to Econometric Analysis

Ass. Prof. Andriy Stavytskyy

Types of Data and Notation

- Time series data
- Cross-sectional data
- Panel data, a combination of mentioned above types

Time series data

- The data may be
  - quantitative (e.g. exchange rates, stock prices, number of shares outstanding),
  - qualitative (e.g. day of the week).

- Examples of time series data

Series	Frequency
GNP or unemployment	monthly or quarterly
government budget deficit	annually
money supply	weekly
value of a stock	
market index	as transactions occur

Examples of Problems Using Time Series Regression

- How the value of a country's stock index has varied with that country's macroeconomic fundamentals.
- How the value of a company's stock price has varied when it announced the value of its dividend payment.
- The effect on country's currency of an increase in its interest rate.

Cross-sectional data

Cross-sectional data is data on one or more variables collected at a single point in time, e.g.

- A poll of usage of internet stock broking services
- Cross-section of stock returns on the New York Stock Exchange
- A sample of bond credit ratings for UK banks

Examples of Problems Using a Cross-Sectional Regression

- The relationship between company size and the return to investing in its shares
- The relationship between a country's GDP level and the probability that the government will default on its sovereign debt.

Panel Data

- Panel Data has the dimensions of both time series and cross-sections, e.g. the daily prices of number of blue chip stocks over two years.

- It is common to denote that each observation by the letter t and the total number of observations by T for time series data, and to denote each observation by the letter i and the total number of observations by N for cross-sectional data.

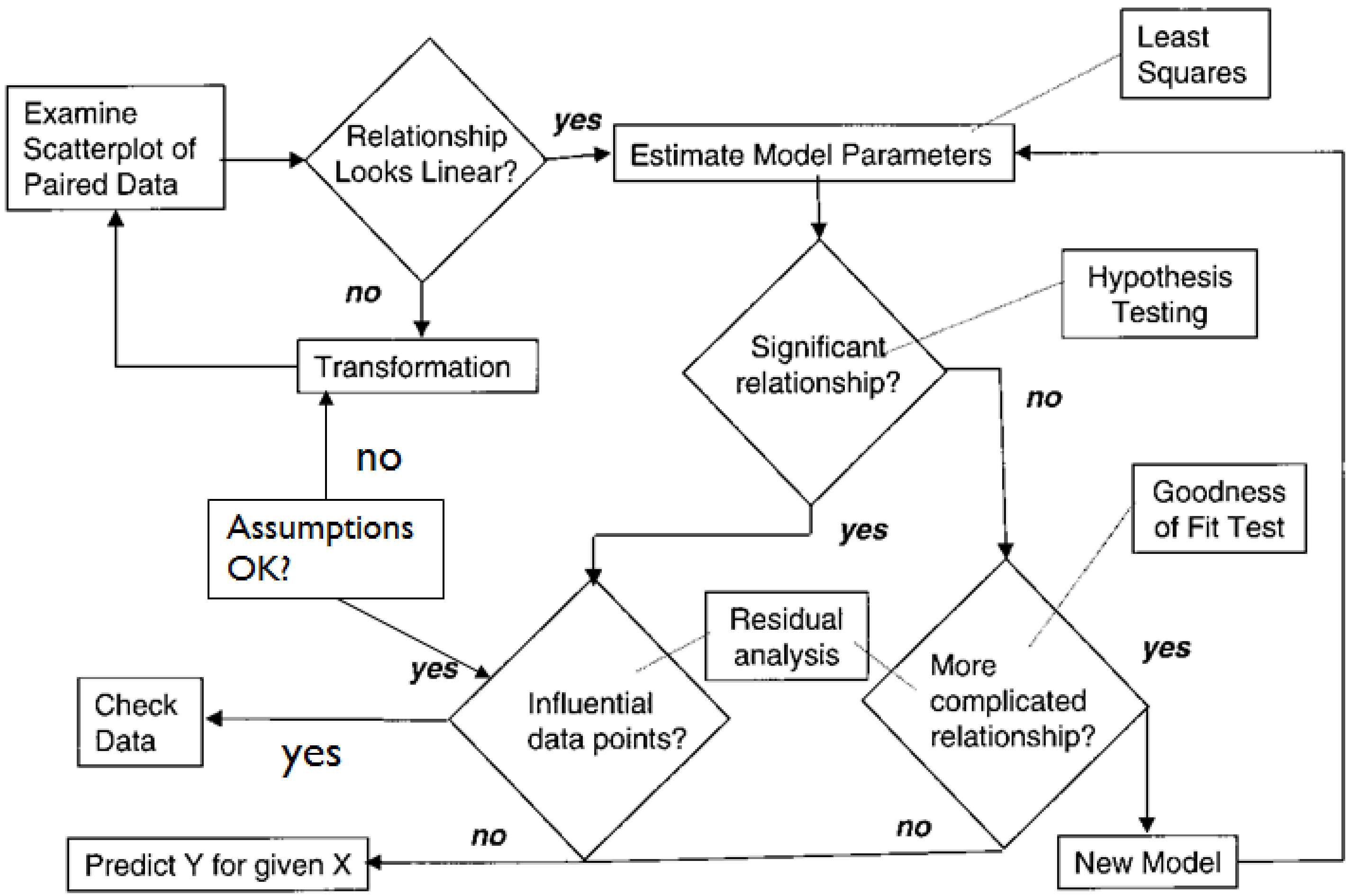
Goal

Develop a statistical model that can predict the values of a dependent (response) variable based upon the values of the independent (explanatory) variables.

Regression Modeling Steps

- Define a problem or question
- Specify model
- Collect data
- Do descriptive data analysis
- Estimate unknown parameters and evaluate the model
- Test the model
- Use model for prediction

How is a Linear Regression Analysis done?



Software

- MS Excel
- EViews
- Mathematica
- SPSS
- Statistica
- R/R-Studio
- Gretl
- MathLab
- …

Linear regression

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1, n}$$

y_t - dependent variable;

$x_{1t}, x_{2t}, \ldots, x_{k-1t}$ - independent variables;

ε_t - residuals.

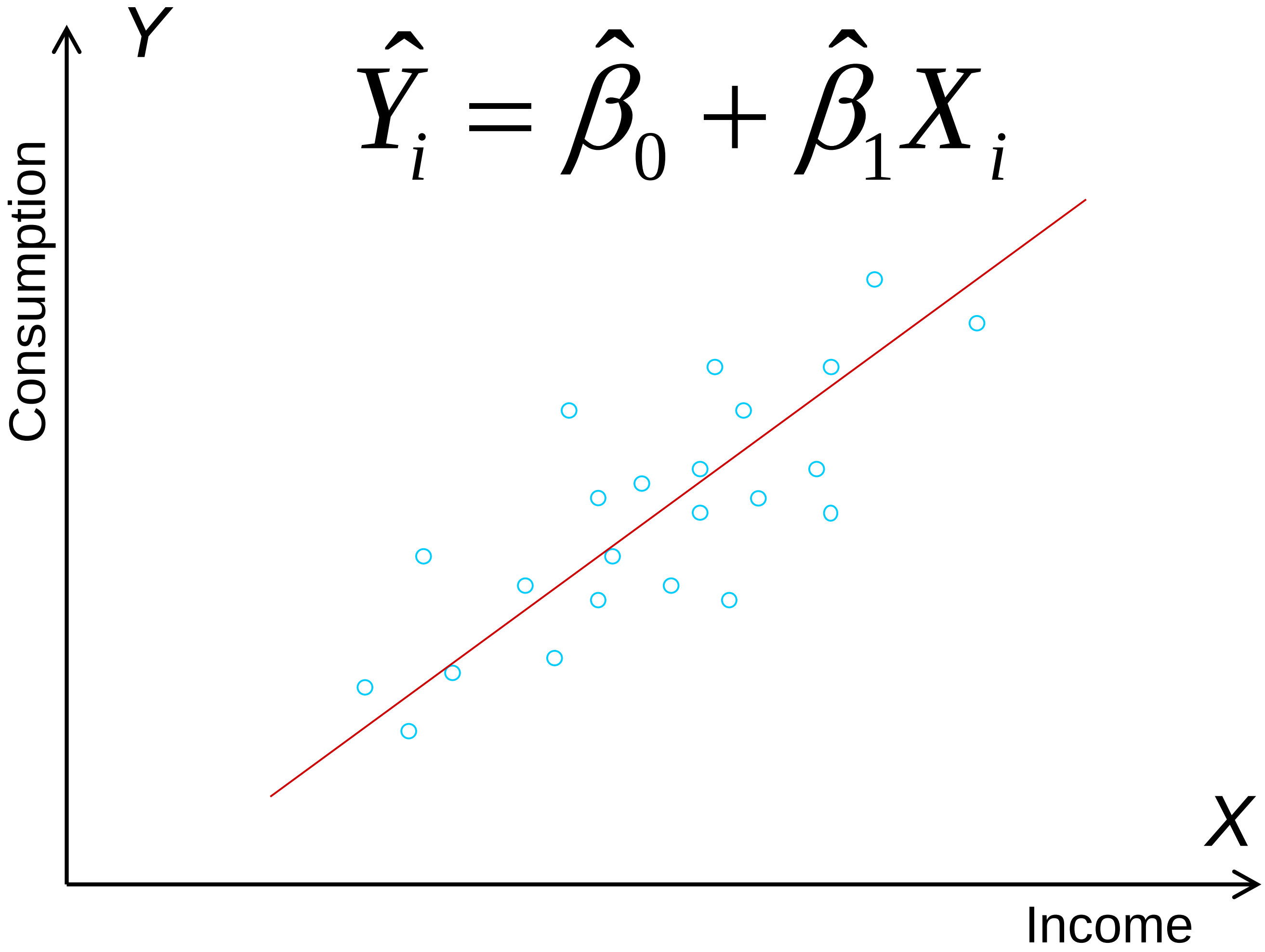
Example

Economic theory predicts that increase of incomes leads to increase of consumption.

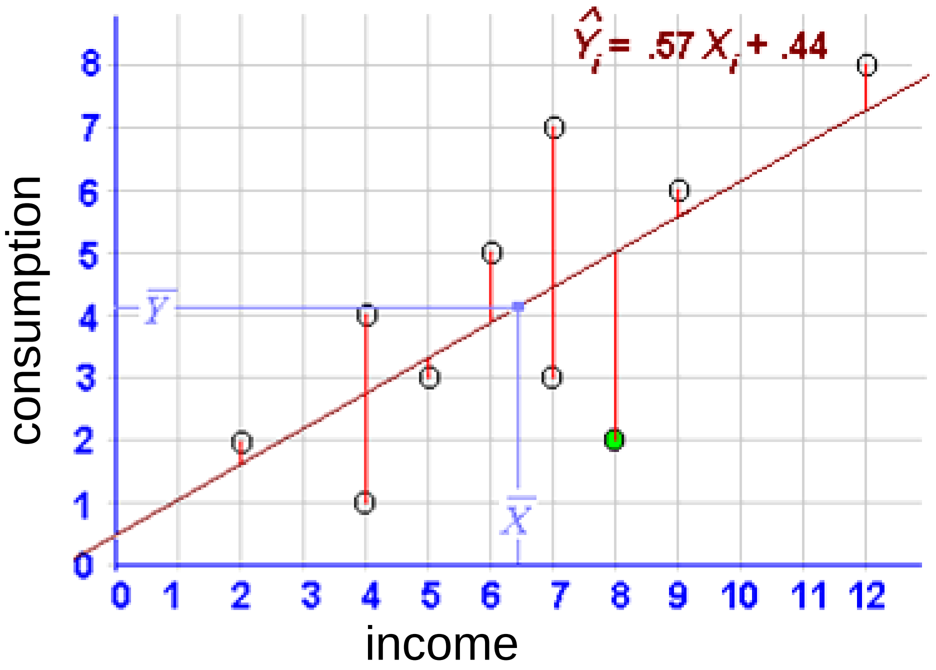
- Y – time series of consumption, USD
- X – time series of income, USD
- n – number of persons observed

Example: regression line

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$



Example: residuals

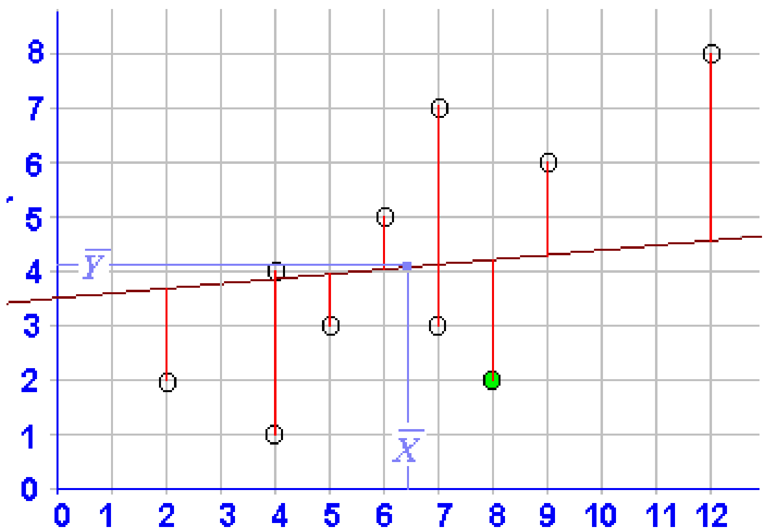


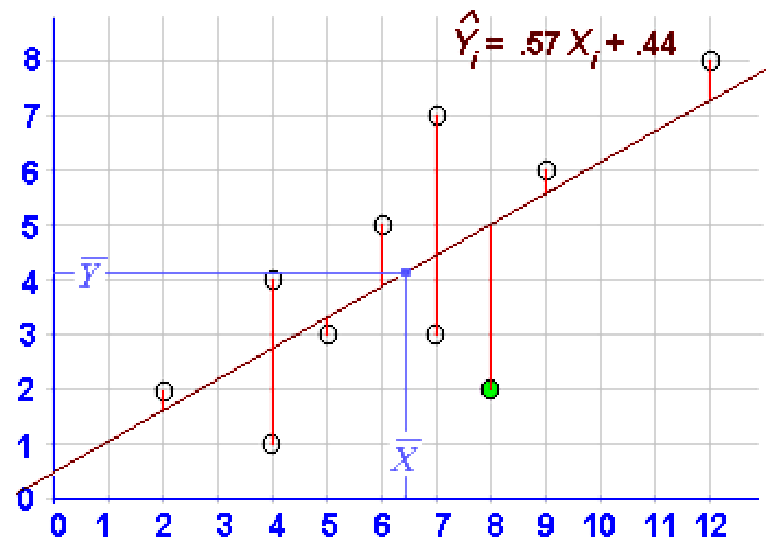
$$e_i = Y_i - \hat{Y}_i$$

Least squared method

$$\sum e_i = 0$$

$$\sum e_i^2 \quad - \text{ minimal}$$





Assumptions

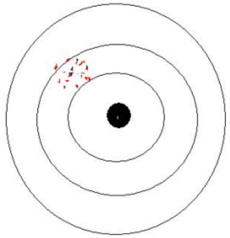
- Linearity - the Y variable is linearly related to the value of the X variable.
- Independence of Error - the error (residual) is independent for each value of X.
- Homoscedasticity - the variation around the line of regression be constant for all values of X.
- Normality - the values of Y be normally distributed at each value of X.

Method of Least Squares

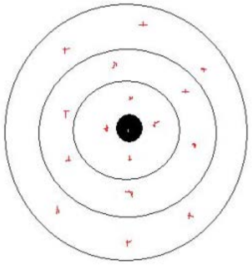
- The straight line that best fits the data.
- Determine the straight line for which the differences between the actual values (Y) and the values that would be predicted from the fitted line of regression (Y-hat) are as small as possible.

$$L = \sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \rightarrow \min$$

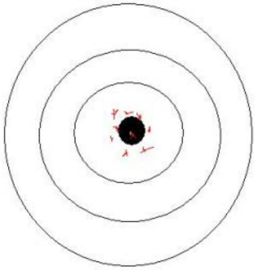
The Three Desirable Characteristics



- **Lack of bias** $E(\hat{\beta}) = \beta$



- **Efficiency**
- Standard error will be minimum
  - Remember:
    $$\operatorname{var}(\hat{\beta}) = \frac{1}{\sum x_i^2}\sigma^2 = \frac{\sigma^2}{\sum x_i^2}$$
  - OLS will minimize σ^2 (the error variance)



- **Consistency**
  - As N increases the standard error decreases
    - *Notice: as N increases so does* Σx_i^2

Inherently Linear Models

- Non-linear models that can be expressed in linear form
  - Can be estimated by least square in linear form
- Require data transformation

Dummy-Variable Regression Model

- Involves categorical X variable with two levels
  - e.g., female-male, employed-not employed, etc.
- Variable levels coded 0 & 1
- Assumes only intercept is different
  - Slopes are constant across categories

The most important factors

- Elasticity coefficients
- Standardized variables

Review

Linear regression

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1, n}$$

y_t - dependent variable;

$x_{1t}, x_{2t}, \ldots, x_{k-1t}$ - independent variables;

ε_t - residuals.

Assumptions

- Linearity - the Y variable is linearly related to the value of the X variable.
- Independence of Error - the error (residual) is independent for each value of X.
- Homoscedasticity - the variation around the line of regression be constant for all values of X.
- Normality - the values of Y be normally distributed at each value of X.

Regression Modeling Steps

- Define problem or question
- Specify model
- Collect data
- Do descriptive data analysis
- Estimate unknown parameters
- Evaluate model
- Test model
- Use model for prediction

Thank you for attention!

KSE