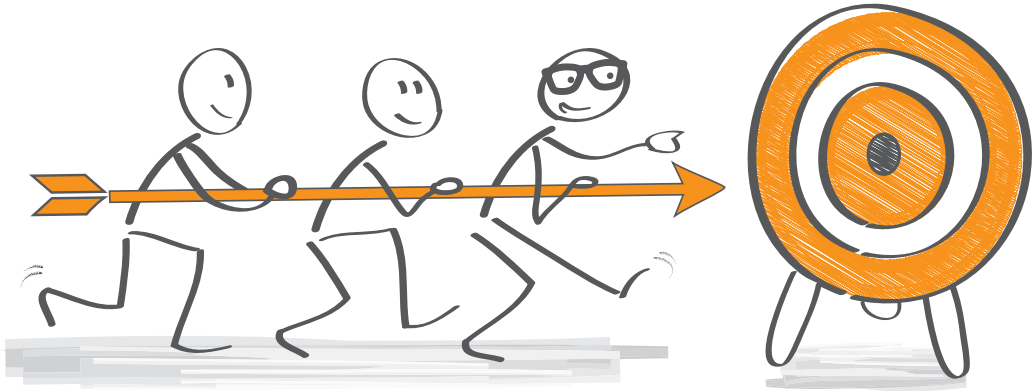
Regression Analysis
Other Regression Methods

Nicoleta Serban, Ph.D.
Associate Professor
Stewart School of Industrial and Systems Engineering

Time Series Regression

About this lesson



Multiple Linear Regression

Data: $\{(x_{11},.., x_{1p}),Y_1\},....,\{(x_{n1},.., x_{np}),Y_n\}$

What if uncorrelated errors assumption does not hold?
- **Degrees of freedom are not equal to the sample size**
- **Higher variability or uncertainty than estimated thus less reliable statistical inference**

- *Linearity/Mean Zero Assumption*: $E(\varepsilon_i) = 0$
- *Constant Variance Assumption*: $Var(\varepsilon_i) = \sigma^2$
- *Independence Assumption*: $\{\varepsilon_1,..., \varepsilon_n\}$ are independent random variables
- *Normality Assumption*: $\varepsilon_i \sim$ Normal

Example: Time Series

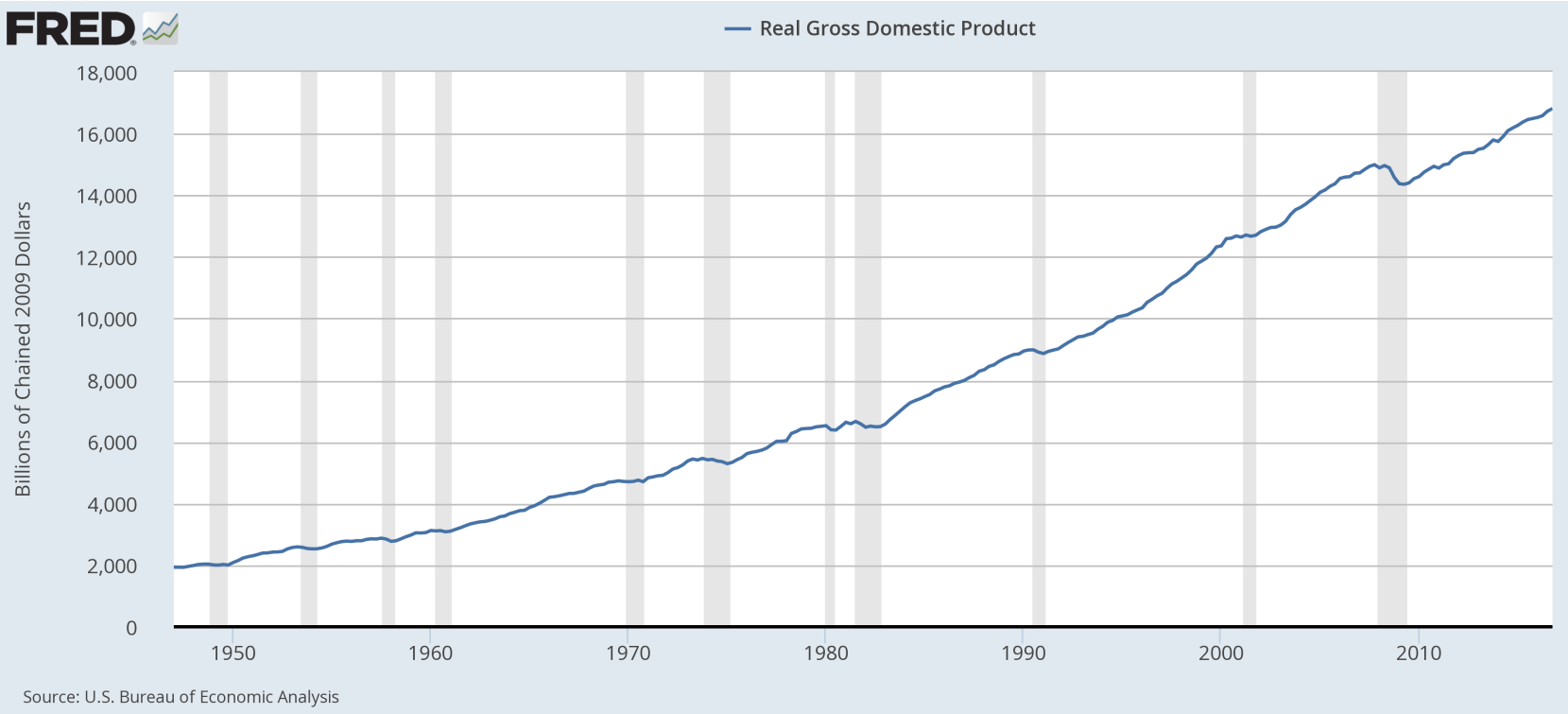
Correlation in time:

- US yearly GDP
- Monthly sales of Australian red wine
- Monthly accidental deaths in the U.S.
- Monthly interest rates in the U.S.
- Daily Average Temperature from La Harpe station in Hancock County, Illinois
- Daily stock price of IBM stock
- 1-minute intraday S&P500 return

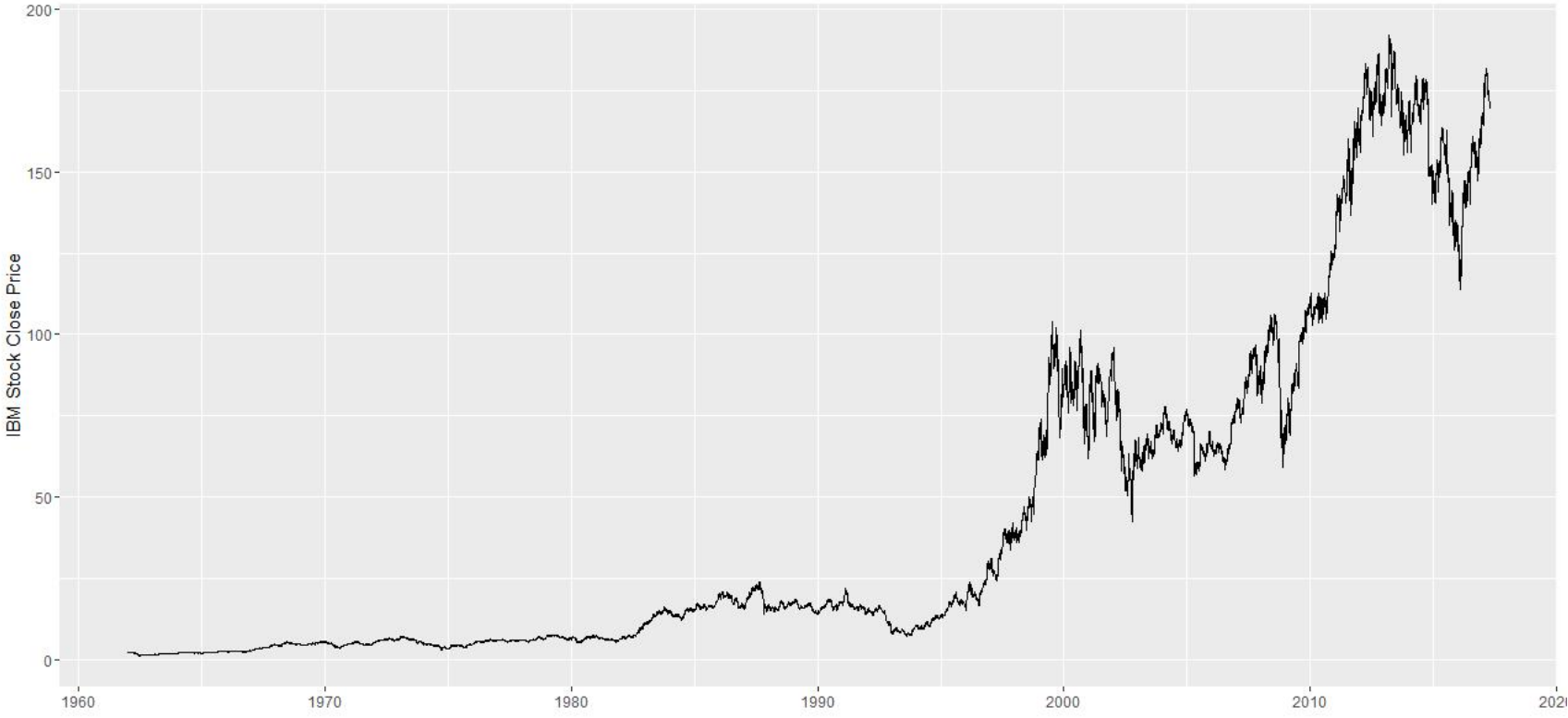
Time Series: Characteristics

- **Trend**: long-term increase or decrease in the data over time
- **Seasonality**: influenced by seasonal factors (e.g. quarter of the year, month, or day of the week)
- **Periodicity**: exact repetition in regular pattern (seasonal series often called periodic, although they do not exactly repeat themselves)
- **Cyclical trend**: data exhibit rises and falls that are not of a fixed period
- **Heteroskedasticity**: varying variance with time
- **Correlation**: positive (successive observations are similar) or negative (successive observations are dissimilar)

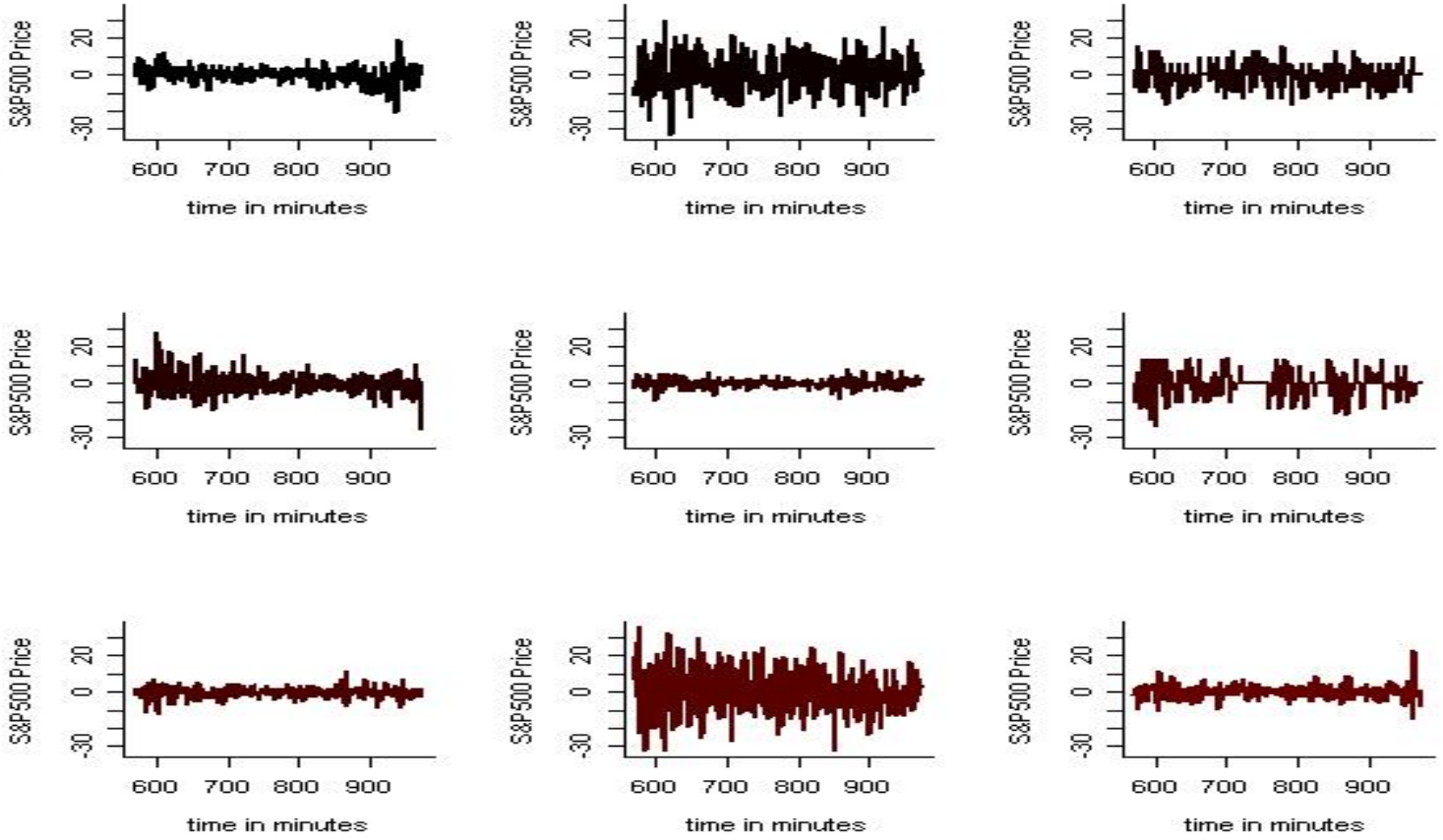
Example: GDP


FRED
Real Gross Domestic Product
Billions of Chained 2009 Dollars
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
1950
1960
1970
1980
1990
2000
2010
Source: U.S. Bureau of Economic Analysis

Example: Daily IBM Stock Price



Example: S&P500 Intraday



Is Time Series Analysis Necessary?

Time Series ⇒ Dependence

- Data redundancy: number of degrees of freedom is smaller than T (T is the number of observations)
- Data sampling: $Y_t, t = 1, .., T$ concentrated about a small part of the probability space

Ignoring dependence leads to

- Inefficient estimates of regression parameters
- Poor predictions
- Standard errors unrealistically small (too narrow CI ⇒ improper inferences)

Time Series: Basics

Data: Y_t, where *t* indexes time, e.g. minute, hour, day, month

Model: $Y_t = m_t + s_t + X_t$

- m_t is a trend component;
- s_t is a seasonality component with known periodicity *d* ($s_t = s_{t+d}$) such that $\sum_{j=1}^{d} s_j = 0$
- X_t is a stationary component, i.e. its probability distribution does not change when shifted in time

Estimation: m_t and s_t are first estimated and subtracted from Y_t to have left the stationary process X_t to be model using time series modeling approaches.

Summary