Stat 481/581: Introduction to Time Series

Instructor: Yan Lu

Fall 2010

- Motivation

  Experimental data, been observed at different points in time

  Example: Daily stock market quotations

  Monthly unemployment figures

  Many statistical methods traditionally assume observations are iid

- Time Series (T.S.) A stochastic process or a sequence of random variables $\{X_t, t \in S\}$, where S is some set of indices.
- The value t usually represents time (hour, month, year). Time points $t_1, t_2, \cdots ; t_n$
- Typically $S = \{0; \pm 1, \pm 2, \cdots\}$,
  $S = \{1990, 1991, 1992, 1993\}$
- We are only going to deal with discrete time processes: S is finite or a countable set.

- Samples in T.S.: A realization of the process X_t denoted by $\{x_t; \in I\}$; where I is a finite set.

- Examples of I in a discrete time case:
  $I = \{1, 2, 3, 4, \cdots\}$
  $I = \{1980, 1981, 1985, 1986, \cdots 1995\}$

- Equally spaced time series are the most common in practice. This is the case of $\{I = t_1, t_2, t_3, \cdots t_n\}$ where $\triangle = t_i - t_{i-1}$ with $\triangle$ a constant.

Difference with traditional Statistical Inference

- Traditionally, data is assumed to be an i.i.d process (random sample).
  Example: $X_1, X_2, \cdots X_n$ are i.i.d. and $X_i \sim N(u; \sigma^2)$.
- In T.S. we are relaxing this assumption and wish to model the dependency among observations.

Main goals in Time Series

- Based on the data, we wish to characterize

  $E(X_t) = \mu_t$ (mean or trend)

  $V(X_t) = \sigma_t^2$ (variance or volatility)

  $Cov(X_t; X_s) = E(X_t - \mu_t)(X_s - \mu_s)$ (autocovariance)

- Determine the periodicity or cycles of the observed process (spectral/periodogram analysis).

- Decompose time series into latent processes.

  $X_t = a_t + S_t + v_t$ where a_t represents the trend; S_t represents the seasonality; v_t represents noise.

- Formulate and estimate a parametric model for X_t (need to propose methods of estimation and model diagnostics).
- This point is related to the estimation of autoregressive (AR) or ARIMA models. (Box and Jenkins methodology).
- Estimation of Missing values.

  Suppose we observe $x_1, x_2, \cdots, x_{200}$, 200 observations but x_{100} was not observed. We wish an estimate $\hat{x}_{100}$ for x_{100}.
- Prediction or Forecasting (what a future value is). Suppose our data is $x_1, x_2, \cdots x_{200}$, we wish to forecast the next 10 values, $x_{201}, x_{202}, \cdots x_{210}$.

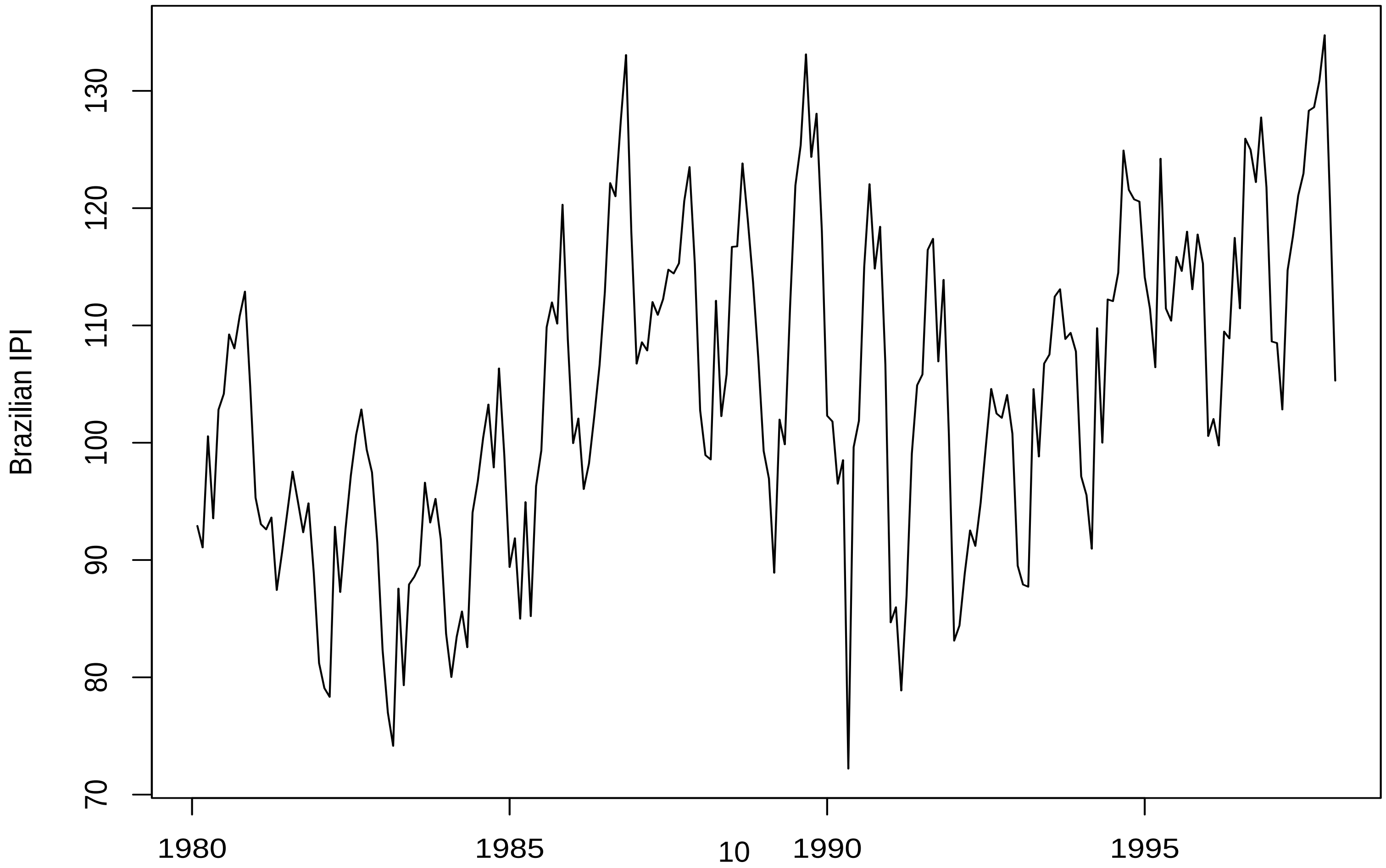
Time Series plot:

- The traditional display for data in time series is to plot each value x_t versus each time t.
- Need to be careful about the labels, scales and the pixels chosen to produce the graph.
- The plot allows to find stationarity or non-stationarity, cycles, trends, outliers or interventions.
- It will assist in the formulation of a parametric model.
- Many examples will be presented along the course.

Brazilian Industrial Production Index

- 215 monthly observations. The data starts in February 1980 and ends in December 1997.
- Data exhibits “ups” and “downs”.
- Data exhibits a periodic or cyclical pattern.
- The process generating the observations appears to be non-stationary.
- The behavior shown by this data is typical of econometric time series.

Figure 1: Brazilian Industrial Production Index



R Code for Brazilian IPI example

```
>y=read.table(“C:/braipi”,skip=1)

#reading data

>x=ts(y[,2],start=c(1980,2),frequency=12)

#plot data
> ts.plot(x,xlab=“time”,ylab=“Brazilian IPI”)
```