**STA 581: Introduction to Time Series. Fall 2006**

Instructor: Gabriel Huerta, MWF 11-11:50am HUM 428.

- Definition of a time series.
- Difference between time series and other statistical approaches.
- Main goals of a time series analysis.
- Time series plots.

* Material relates to Shumway and Stoffer, sections 1.1-1.3

**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 $t_1, t_2, \ldots, t_n$
- Typically $S = \{0, \pm 1, \pm 2, \ldots\}$, $S = \{1990, 1991, 1992, 1993, \}$
- 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 $\{x_t; t \in I\}$ where $I$ is a finite set.
- Examples of $I$ in a discrete time case:

$$I = \{1, 2, 3, 4, \ldots, n\}$$

$$I = \{1980, 1981, 1985, 1986, \ldots, 1995\}$$

$$I = \{1/80, 2/80, \ldots, 12/80, 1/81, 2/81, \ldots, 12/81\}$$

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

**Difference with traditional Statistical Inference (STA 553)**

- The data is assumed to be an i.i.d process (random sample). Example: $X_1, X_2, \ldots, X_n$ are i.i.d. and $X_i \sim N(\mu, \sigma^2)$.
- In T.S. we are relaxing this assumption and wish to *model* the *dependency* among observations.
- For this purpose, we will discuss the concept of *autocorrelation*.

## Main goals in Time Series

- Based on the data, we wish to characterize

$$E(X_t) = \mu_t \quad (\textit{mean or trend})$$

$$V(X_t) = \sigma_t^2 \quad (\textit{variance or volatility})$$

$$Cov(X_t, X_s) = E(X_t - \mu_t)(X_s - \mu_s) \quad (\textit{autocovariance})$$

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

- Decompose time series into latent processes.

$$X_t = a_t + S_t + \nu_t$$

where $a_t$ represents the trend; $S_t$ represents the seasonality; $\nu_t$ represents noise.

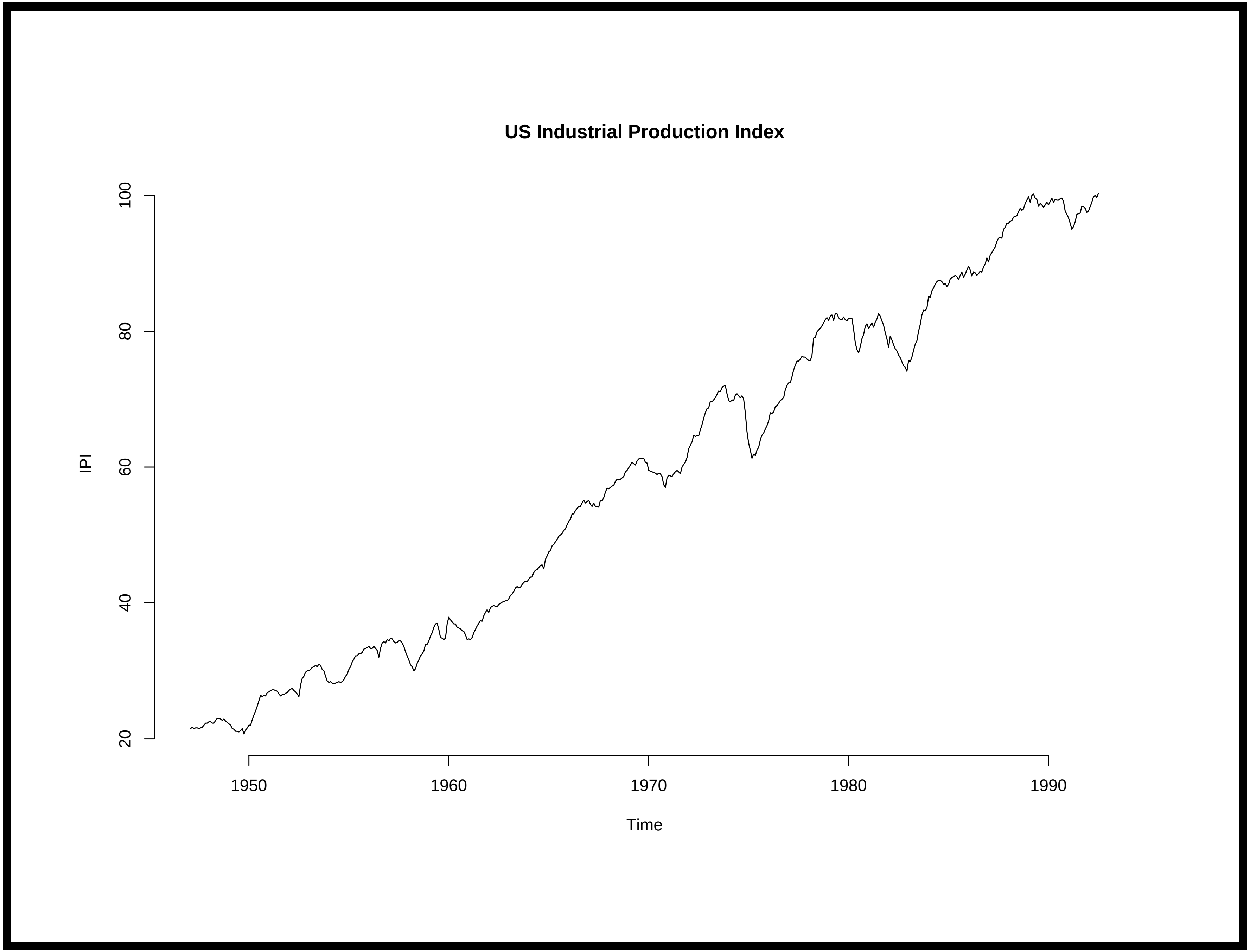
- Formulate and estimate a *parametric* model for $X_t$ (need to propose methods of estimation and model diagnostics).
- This point is related to the estimation of autoregressive (AR) or ARMA models. (Box and Jenkins methodology).
- Estimation of Missing values (fill "gaps"). Suppose we observe $x_1, x_2, \ldots, x_{200}$; 200 observations but $x_{100}$ was not observed. We wish an estimate $\hat{x_{100}}$ for $X_{100}$.
- Prediction or Forecasting ("would like to know what a future value is"). Suppose our data is $x_1, x_2, \ldots, x_{200}$, we wish to forecast the next 10 values, $x_{201}, x_{202}, \ldots, x_{210}$. In this case, our *forecasting horizon* is 10.

**Time Series plot:**

- The traditional display for data in time series is to plot each value $x_t$ versus each time $t$.
- The first step on any time series analysis.
- Need to be carefull 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.

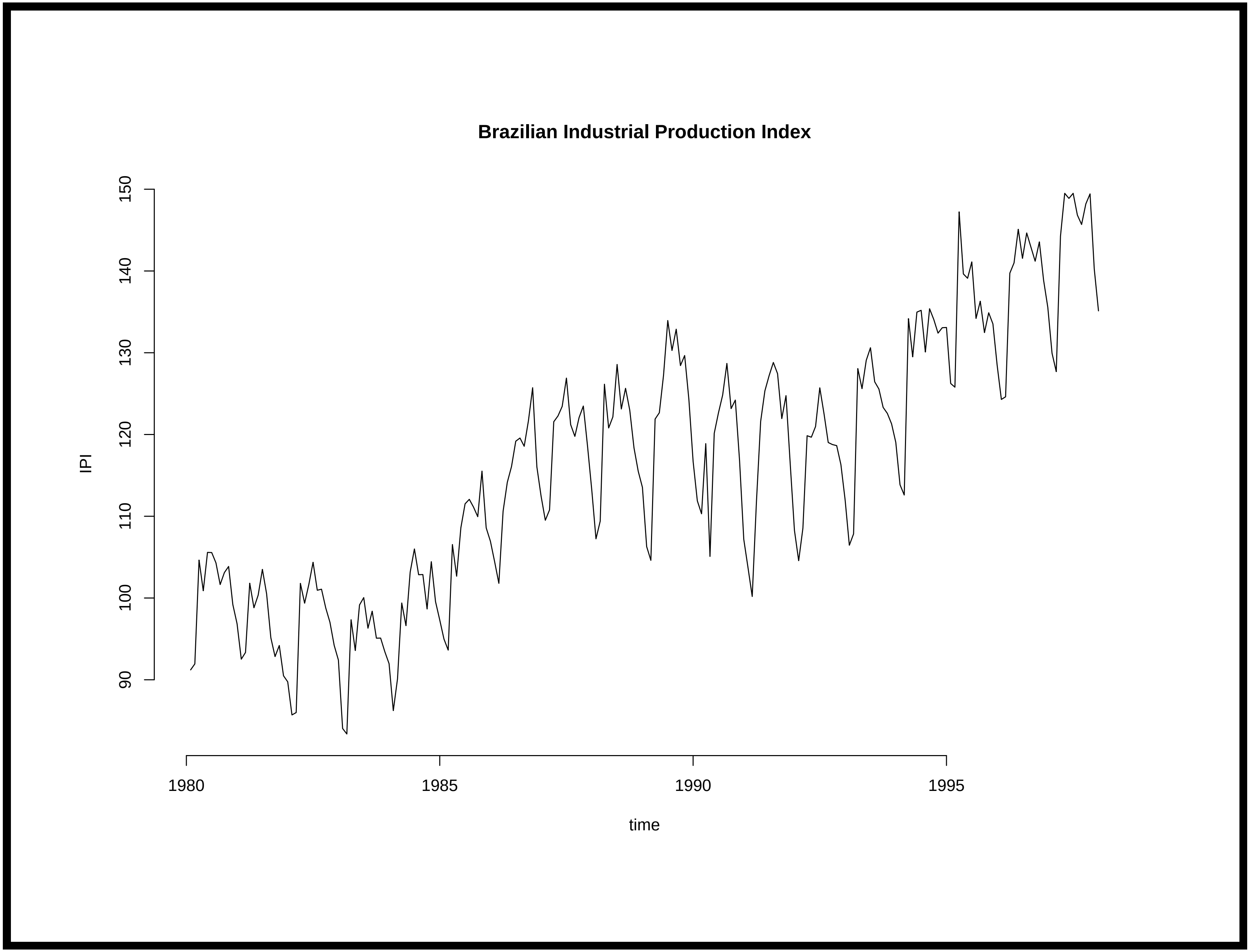
**US Industrial Production Index**

- 546 monthly observations.
- The data starts in January 1947 and ends in December 12.
- With more data it is more sensible to look into a long-term behavior.
- The data has been seasonally adjusted (periodicity has been removed).
- A "positive slope" trend is present in the data.
- Can this data be related to a deterministic regression line or to a purely stochastic mechanism?


US Industrial Production Index
IPI
20
40
60
80
100
1950
1960
1970
1980
1990
Time

**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.


Brazilian Industrial Production Index
IPI
90
100
110
120
130
140
150
1980
1985
1990
1995
time

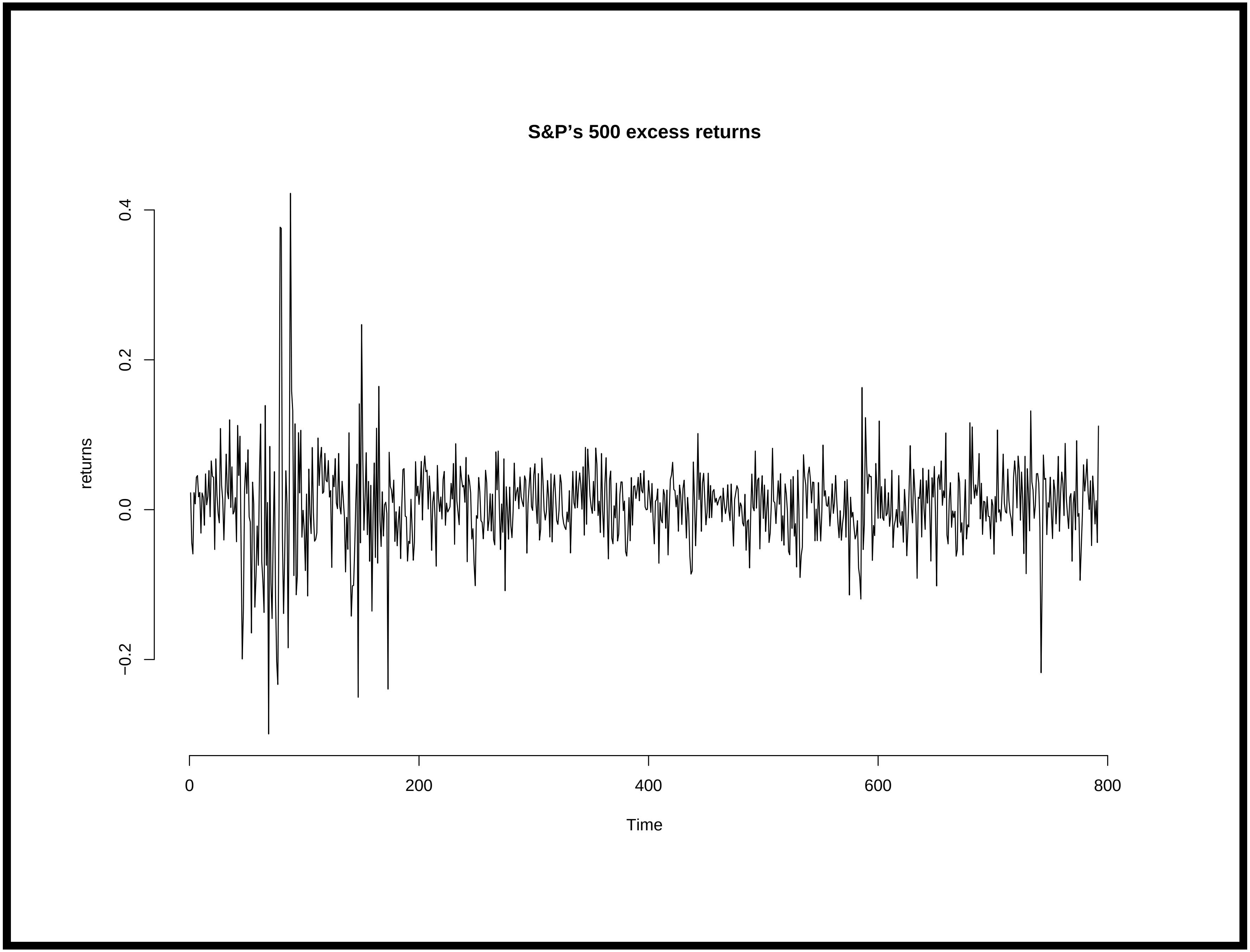
## R Code for Brazilian IPI example

Go to http://www.stat.unm.edu/∼ghuerta/tseries/braipi and save data into your directory

```
> y=read.table(``/mydata/braipi'',skip=1)
# reading data
> x=ts(y[,2],start=c(1980,2),frequency=12)
# creating a ts object
> ts.plot(x,xlab=''time'',ylab=''Brazilian IPI'')
```

## Standard and Poor's 500

- Financial index.
- The data consists of excess returns.
  $X_t = log(s_t) - log(s_{t-1})$
- The mean level of the process seems constant.
- There are sections of the data with explosive behavior (high volatility).
- The data corresponds to a non-stationary process.
- The variance (or volatility) is not constant in time.
- No *linear* time series model will be available for this data.


S&P's 500 excess returns
returns
0.4
0.2
0.0
−0.2
0
200
400
600
800
Time

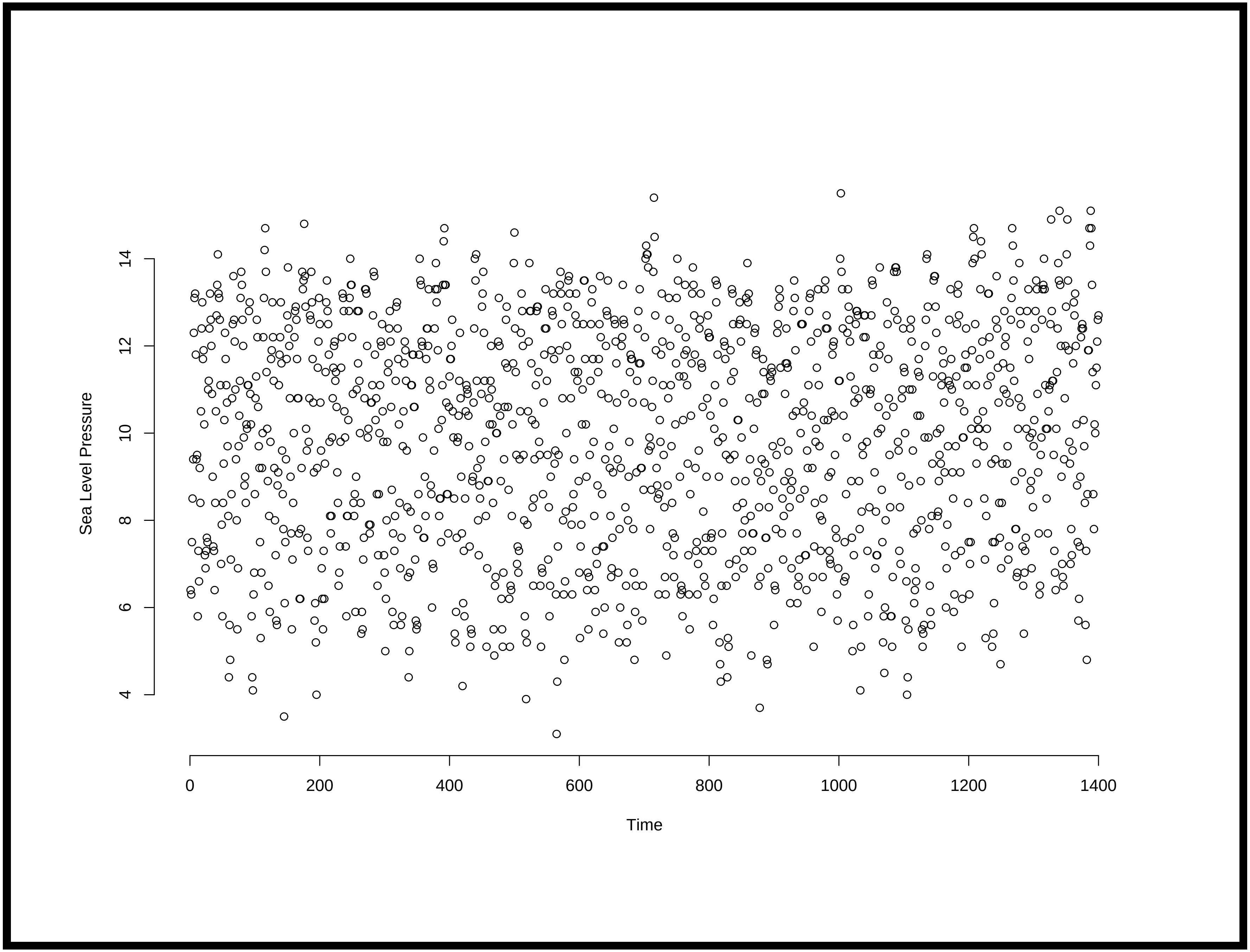
**R code for SP-500 data example**

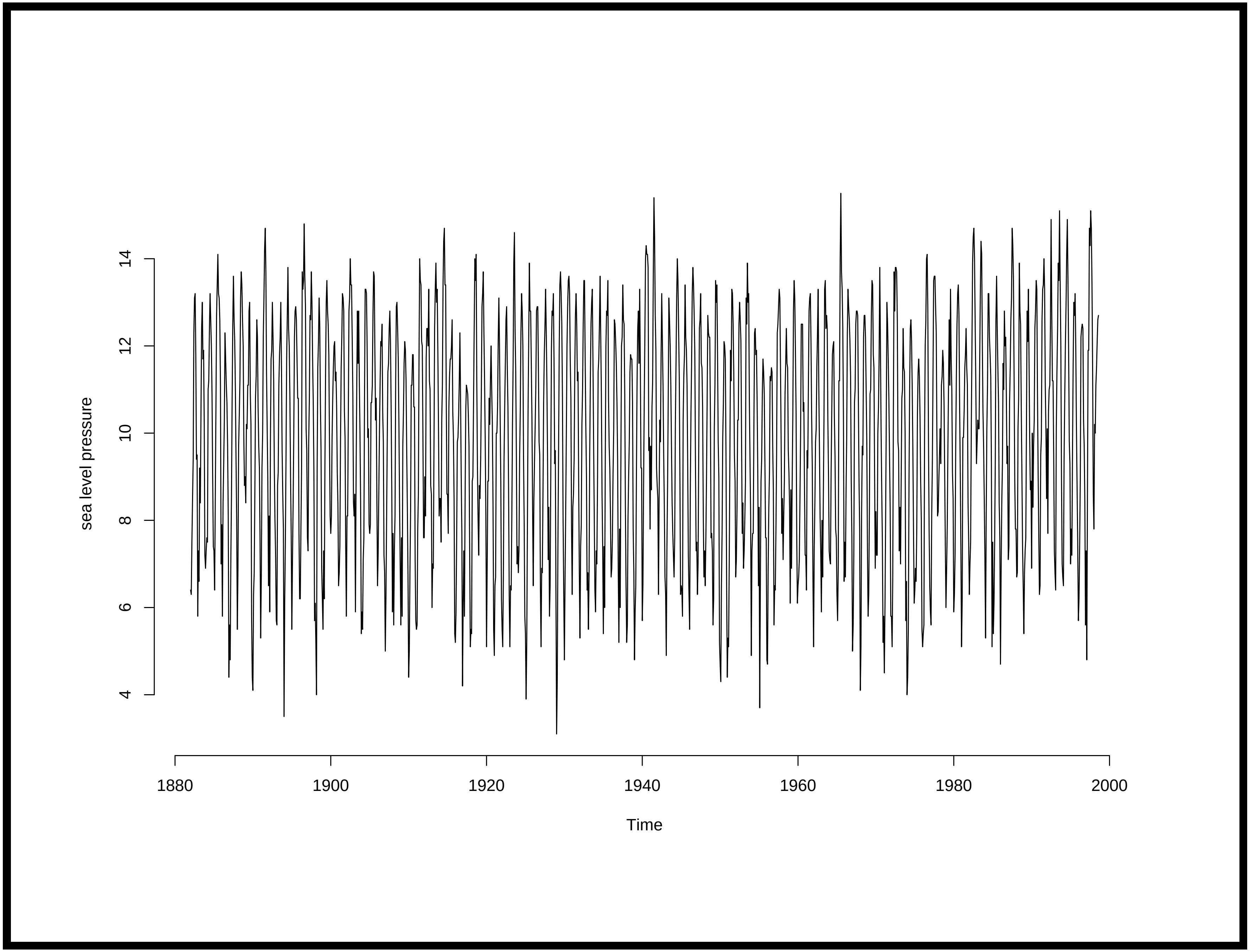
If we have the values of $s_t$ as a vector-file stored in *sp-500.dat*.

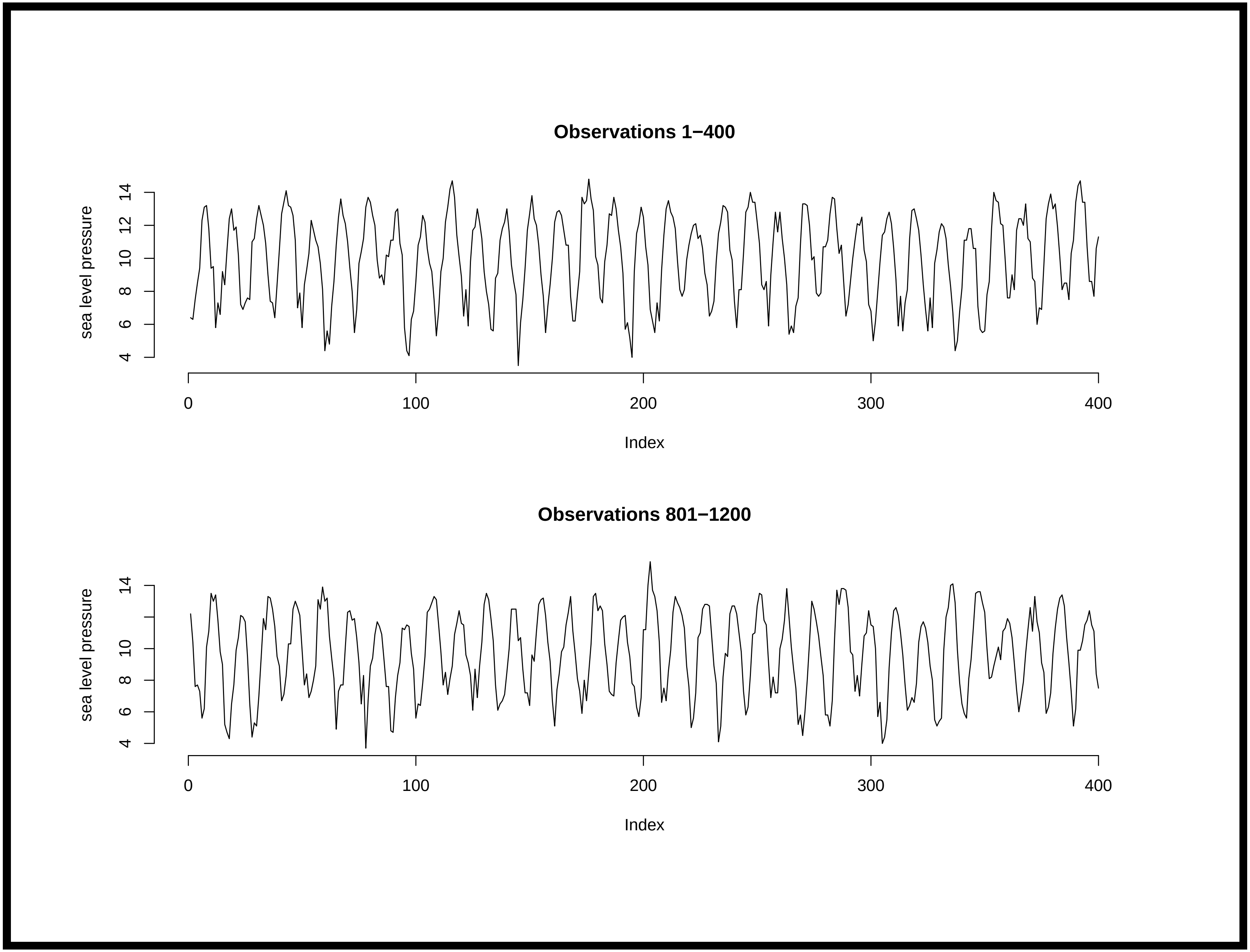
```
> x=scan(‘‘/mydata/sp-500.dat’’)
# Read in data
> y=diff(log(x),lag=1)
# First difference of log-data
> ts.plot(y,xlim=’’time’’,ylab=’’returns’’)
```

## Sea Level Pressures at Darwin

- Monthly values starting from 1882 and ending in 1998.
- This series is a key indicator for climatological studies.
- Expectedly, there is a strong seasonality in this data.
- The first plot is the actual data points. No lines are connected between points.
- Second plot is the standard time series plot with points connected by lines. The seasonality is now clear.
- There is no obvious change in mean so the process seems stationary.
- The third graph includes plots for observations 1-400, 401-801.


Sea Level Pressure
4
6
8
10
12
14
0
200
400
600
800
1000
1200
1400
Time


sea level pressure
4
6
8
10
12
14
1880
1900
1920
1940
1960
1980
2000
Time


Observations 1−400
sea level pressure
4
6
8
10
12
14
0
100
200
300
400
Index
Observations 801−1200
sea level pressure
4
6
8
10
14
0
100
200
300
400
Index

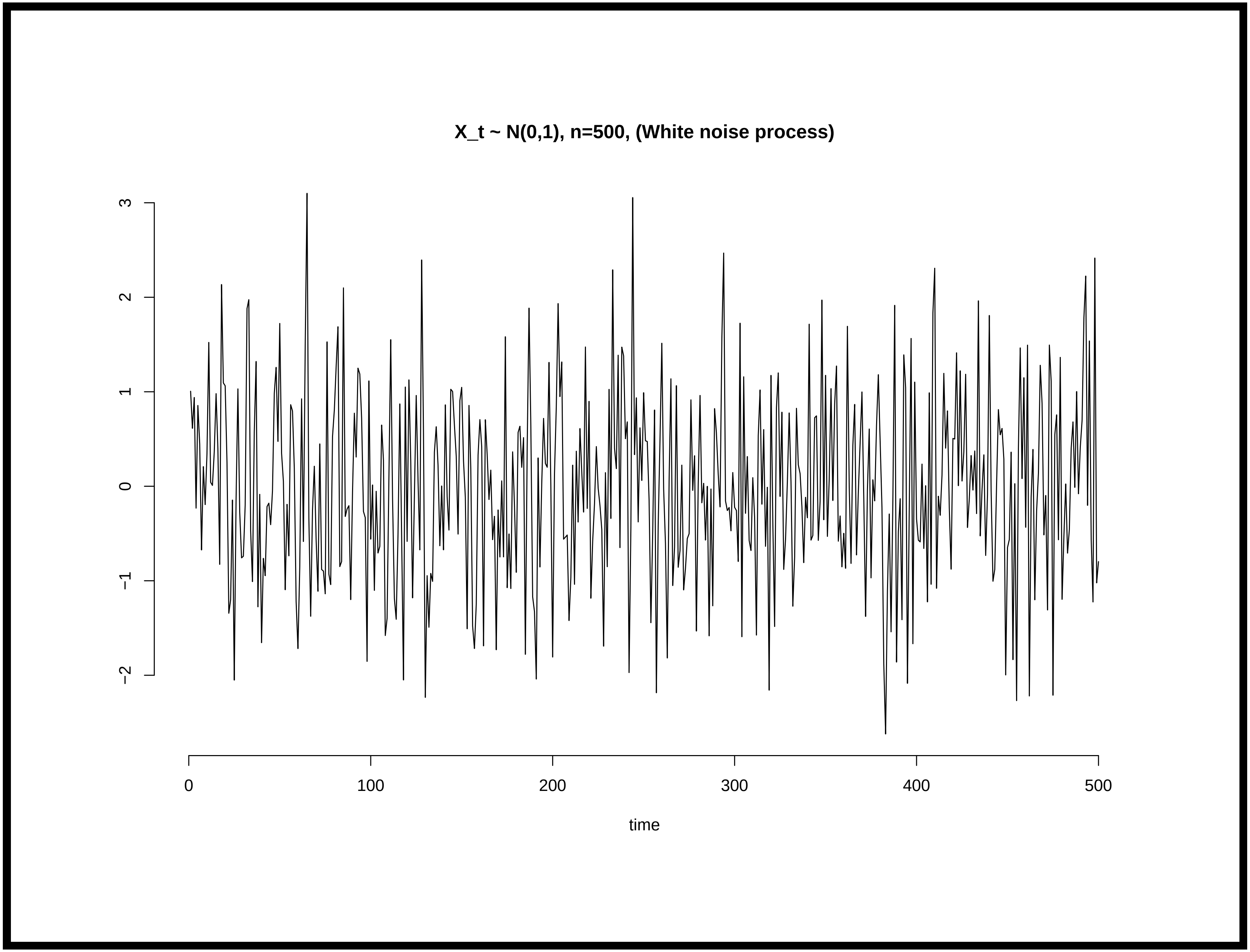
## Some R code

```
 x <- scan("/mydata/darwin1")
# Reading Darwin data
 darw <- ts(x,start=c(1882,1),frequency=12)
# Transforming into a time series object..
plot(darw,ylab=''sea level pressure'')
title(``Time Series Plot of Darwin Data'')
# Produces the time series plot.
par(mfrow=c(2,1))
 plot(darw[1:400],ylab="sea level pressure",type="l")
title("Observations 1-400")
 plot(darw[801:1200],ylab="sea level pressure",type="l")
title("Observations 801-1200")
```

The darwin data and others are available from the class web page.

**White Noise Process**

- The $\omega_t's$ are iid and each follows a $N(0, \sigma_\omega^2)$ distribution.
- No time correlation to model.
- It is a stationary process.
- The time series plot will not show any patterns or any changes in time.
- The figure in the next page, shows a realization of size $n = 500$ of a white noise process with $\sigma_\omega^2 = 1$. (N(0,1)).


X_t ~ N(0,1), n=500, (White noise process)
3
2
1
0
−1
−2
0
100
200
300
400
500
time

**R code for White noise and Moving Average process (Ex 1.9 in text)**

```
> w=rnorm(500,0,1) # 500 N(0,1) variates
> v=filter(w,sides=2, rep(1,3)/3) # moving average
> par(mfrow=c(2,1))
> plot.ts(w)
> plot.ts(v)
```