

Lecture 1: Univariate Time Series

B41910: Autumn Quarter, 2008, by Mr. Ruey S. Tsay

1 Some Basic Concepts

1. *Time Series*: A sequence of random variables measuring certain quantity of interest over time.

Convention:

- In applications, a time series is a sequence of measurements of some quantity of interest taken at different time points.
- Usually, the measurements are observed at equally spaced time intervals, resulting in a discrete-time time series.
- If t is continuous, then we have a continuous-time time series. The series becomes a stochastic process.
- Notation: X_t or Y_t or Z_t for a discrete-time time series and $X(t)$, or $Y(t)$ or $Z(t)$ for the continuous-time case.
- X_t can be a continuous random variable or a discrete random variable, e.g. counts or the number of wins of a football team in a season.

2. Basic objective of time series analysis

The objective of (univariate) time series analysis is to find the dynamic dependence of X_t , i.e. the dependence of X_t on its past values $\{X_{t-1}, X_{t-2}, \dots\}$. A linear model means X_t depends linearly on its past values; a formal definition of linear model will be given later. To describe the dynamic dependence of X_t effectively, it pays to introduce the following operator.

3. Backshift (or lag) operator:

We define the backshift operator “B” (or the lag operator “L”) by

$$BX_t = X_{t-1}.$$

In other words, BX_t is the value of the time series at time index $t - 1$. Note that “B” operates on “time index” so that for the function $f_t = 2^t$ for $t > 1$, $Bf_t = f_{t-1} = 2^{t-1}$.

We can define a Lag (or backshift) polynomial as

$$\phi(B) = \phi_0 - \phi_1 B - \dots - \phi_p B^p = \phi_0 - \sum_{i=1}^p \phi_i B^i$$

where $\phi_0 = 1$, $\phi_p \neq 0$, and p is a non-negative integer referred to as the “order” of $\phi(B)$.

Applying this operator to the X_t sequence, we obtain

$$\phi(B)X_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = X_t - \sum_{i=1}^p \phi_i X_{t-i}.$$

This equation is often used in time series analysis to describe the dynamic dependence of X_t on its past values.

The equation

$$\phi(B)X_t = c, \tag{1}$$

where c is a constant, is called a “difference equation” of order p . If $c = 0$, the equation is homogeneous. The variable X_t , which satisfies the difference equation in (1), is a solution of that equation. In practice, different $\phi(B)$ can give rise to different dynamic behavior of X_t . We shall use such a difference equation to describe the dynamic pattern of a linear time series.

A variety of dynamic dependence patterns of X_t can be generated by considering the “rational” lag polynomial $\pi(B) = \phi(B)/\theta(B)$.

A simple example of the rational polynomial is

$$\pi(B) = \frac{1}{1 - \theta B}.$$

Using long division, we have

$$\frac{1}{1 - \theta B} = 1 + \theta B + \theta^2 B^2 + \cdots.$$

Therefore,

$$\frac{1}{1 - \theta B} X_t = \sum_{i=0}^{\infty} \theta^i X_{t-i}.$$

If the $\{X_t\}$ sequence is bounded, then we might want the resulting sequence to be bounded as well. This is achieved by requiring $|\theta| < 1$. This special rational polynomial shows that X_t is an infinite-order moving average of its past values, $\{X_{t-1}, X_{t-2}, \cdots\}$, with weights decaying exponentially.

If $|\theta| > 1$, then it might be reasonable to define

$$\begin{aligned} \frac{1}{1 - \theta B} &= \frac{-(\theta B)^{-1}}{1 - (\theta B)^{-1}} = \frac{-1}{\theta B} \left[1 + \frac{1}{\theta} B^{-1} + \frac{1}{\theta^2} B^{-2} + \cdots \right] \\ &= - \sum_{i=1}^{\infty} (1/\theta)^i X_{t+i}. \end{aligned}$$

This is a forward-looking moving average that relates X_t to its future values. Note that sometimes we write $B^{-1} = F$ such that $F X_t = X_{t+1}$ and refer to F as the forward operator.

4. First-order Difference Equations

A difference equation is a deterministic relationship between the current value X_t and its past values X_{t-i} with $i > 0$. In some cases, it may also contain the current and past values of a “forcing” or “driving” variable. A first order difference equation involves only one lagged variable:

$$X_t = \phi X_{t-1} + ba_t + c,$$

where a_t is a forcing variable, which follows a well-defined probability distribution, and ϕ , b and c are real-valued parameters. Using the backshift operator, we can write the model as

$$(1 - \phi B)X_t = c + ba_t.$$

The “solution” to this difference equation expresses the current value X_t as a function of time and the forcing variable a_t .

$$X_t = \frac{c}{1 - \phi B} + \frac{b}{1 - \phi B}a_t + \gamma\phi^t.$$

The term $\gamma\phi^t$ is included because this is the only function such that $(1 - \phi B)f_t = 0$, where γ is a unknown parameter whose value is determined by some initial condition. Note that $\gamma\phi^t$ is the solution to the homogeneous equation $(1 - \phi B)X_t = 0$. As in the theory of differential equations, the solution of a difference equation consists of the solution to the homogeneous part plus a particular solution to the inhomogeneous equation.

The solution, $X_t = c/(1 - \phi) + b \sum_{i=0}^{\infty} \phi^i a_{t-i} + \gamma\phi^t$, is fully determined by knowledge of the initial conditions of the sequence. If X_0 is known, then

$$X_t = \frac{c(1 - \phi^t)}{1 - \phi} + b \sum_{i=0}^{t-1} \phi^i a_{t-i} + \phi^t X_0.$$

If $|\phi| < 1$ and a_t is bounded with mean zero, then $E(X_t)$ approaches $c/(1 - \phi)$ from any starting point.

Obviously, the value of ϕ determines the qualitative behavior of the equation $X_t = \phi^t X_0$. For the first-order equation, there are three types of solution. If $0 < \phi < 1$, then X_t dampes smoothly to zero. If $-1 < \phi < 0$, the X_t dampes oscillatorily to zero. If $|\phi| > 1$, then X_t is explosive. [The case of $|\phi| = 1$ is obvious.]

5. Second-order Difference Equations:

The same idea can be extended to higher-order difference equations. Solutions to higher-order difference equations can exhibit more interesting sinusoidal patterns. The latter case occurs when the equation contains complex solutions.

Consider the general second-order difference equation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \delta + ba_t$$

or

$$(1 - \phi_1 B - \phi_2 B^2)X_t = \delta + ba_t.$$

Solutions of this equation can be computed by factoring the backshift polynomial as, $(1 - \phi_1 B - \phi_2 B^2) = (1 - \lambda_1 B)(1 - \lambda_2 B)$. Considering just the homogeneous equation, we find a solution of the form

$$X_t = c_1(\lambda_1)^t + c_2(\lambda_2)^t,$$

where c_1 and c_2 are unknown parameters depending on the initial conditions. Note that $1/\lambda_1, 1/\lambda_2$ are the zeros of the polynomial $1 - \phi_1 B - \phi_2 B^2$. If the homogeneous solution is to remain bounded, we would require $|\lambda_i| < 1$ for $i = 1, 2$, or equivalently that the zeros of the polynomial $1 - \phi_1 B - \phi_2 B^2$ lie outside the unit circle (modulus > 1). [Note: Zeros of $1 - \phi_1 B - \phi_2 B^2$ are roots of the equation $1 - \phi_1 B - \phi_2 B^2 = 0$.]

For a second-order equation, we have three possibilities: (a) distinct real roots, (b) equal real roots, and (c) complex roots. The quadratic formula gives the roots as

$$\lambda_i = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

We have complex roots if $\phi_1^2 + 4\phi_2 < 0$. The roots are $a \pm bi$ where i is $\sqrt{-1}$. Note that the complex roots come in a conjugate pair.

If we write the roots using polar form, we can see how oscillatory solutions are possible.

$$a \pm bi = r(\cos \theta \pm i \sin \theta)$$

where $r = \sqrt{a^2 + b^2}$ and $\cos \theta = a/r = \phi_1/(2\sqrt{-\phi_2})$, or $\theta = \cos^{-1}(\phi_1/(2\sqrt{-\phi_2}))$.

Using DeMoivre's formula, namely $\cos \theta + i \sin \theta = e^{i\theta}$, we can write

$$\begin{aligned} X_t &= c_1(re^{i\theta})^t + c_2(re^{-i\theta})^t \\ &= r^t(c_1e^{it\theta} + c_2e^{-it\theta}) \\ &= r^t[(c_1 + c_2) \cos(t\theta) + i(c_1 - c_2) \sin(t\theta)]. \end{aligned}$$

Since X_t is real, $c_1 + c_2$ must be real while $c_1 - c_2$ must be imaginary. Thus, c_1 and c_2 are complex conjugates. Using DeMoivre's formula again and some identities from trigonometrics, we obtain

$$X_t = kr^t \cos(t\theta + \omega).$$

Equal roots case:

$$(1 - \lambda B)^2 X_t = 0$$

The solution of which is

$$X_t = c_1 \lambda^t + c_2 t \lambda^t.$$

6. General Case:

The above results can readily be extended to the general higher-order difference equations.

7. Stochastic Difference Equations: When the “forcing” factor is stochastic, we have a general difference equation. In particular, the case in which the forcing variable is a sequence of independent and identically distributed normal random variables $\{a_t\}$ plays an important role in time series analysis. Here the solution X_t is usually correlated and follows certain statistical distribution.

8. Let a_t and X_t be input and output at time t , respectively. Consider the linear system

$$X_t = \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \psi(B) a_t,$$

where $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$ and $\psi_0 = 1$. Consider the relationship between ψ_i and coefficients $\pi(B)$ discussed earlier.

2 Linear Time Series Models

Recall that a time series is a collection of random variables $\{X_t\}$ indexed by “time”. In most economic and business applications, we only observe one *realization* of the time series. Such a realization is called a sample path. The goal of time series analysis is to make inference of the series based on the observed realization. A typical approach of the analysis is to identify a model within a given class of flexible models which can adequately approximate the process under study.

Model choice in time series analysis involves evaluation of a joint distribution function for the sample data:

$$F(x_1, x_2, \dots, x_n) = Pr(X_1 \leq x_1, \dots, X_n \leq x_n)$$

where n is the sample size and x_i 's are real numbers. Note that $\{X_t\}$ may not be a random sample. It is usually an observed series. To succeed in modeling the process, we must restrict the class of joint distributions under consideration. Furthermore, to predict future values of the process, we must be able to identify some key features of the distribution that are **time invariant**.

Stationarity: A particular time-invariant feature that has proven to be useful is the stationarity. A time series $\{X_t\}$ is strictly stationary if

$$F_{X_t, \dots, X_{t+s}}(*) = F_{X_{t+r}, \dots, X_{t+r+s}}(*) \quad \text{for all } r \text{ and } s.$$

In other words, X_t is strictly stationary if (a) the distribution of X_t and X_s are the same for all t and s , (b) the joint distribution of (X_t, X_{t+s}) is the same as that of (X_{t+r}, X_{t+r+s}) for all r and s , (c) the joint distribution of $(X_t, X_{t+s}, X_{t+s+u})$ is identical to that of $(X_{t+r}, X_{t+r+s}, X_{t+r+s+u})$ for all r, s and u , and so on.

In practice, we often relax the requirement of stationarity by considering only the weak stationarity. A time series X_t is weakly stationary if

$$\begin{aligned} E(X_t) &= \mu, \quad \text{a constant} \\ \text{Cov}(X_t, X_{t+k}) &= \gamma_k, \quad \text{a function depending only on } k. \end{aligned}$$

In other words, X_t is weakly stationary if its first two moments are time invariant. Of course, here we assume that the first two moments of X_t exist. Some people refer to weakly stationary processes as covariance stationary processes.

Clearly, strict stationarity implies weak stationarity provided that the first two moments of the series exist. On the other hand, a weakly stationary series may not be strictly stationary.

In many applications, we assume that the time series X_t is Gaussian, that is, jointly normal. This is mainly for statistical convenience. Since the distribution of a normal distribution is determined by its first two moments, weak stationarity is equivalent to strict stationarity for a Gaussian time series.

We shall discuss non-Gaussian time series later.

Autocovariance Function: For a weakly stationary process X_t , $\gamma_k = \text{Cov}(X_t, X_{t+k})$ is called the lag- k autocovariance. Treating γ_k as a function of k , we call γ_k the autocovariance function of X_t . For a Gaussian process, the sample $\mathbf{Z} = (X_1, \dots, X_n)'$ has a multivariate normal distribution:

$$\mathbf{Z} \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu} = \mu(1, \dots, 1)'$ is a n -dimensional vector of μ and $\boldsymbol{\Sigma}$ is a $n \times n$ symmetric matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-3} & \gamma_{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}$$

where we have used the property $\gamma_\ell = \gamma_{-\ell}$ for $\ell < 0$.

In (linear) time series analysis, we focus mainly on various models that parameterize the above covariance matrix in terms of a smaller number of parameters. This is essential, because for an un-restricted covariance matrix there are too many parameters.

Autocorrelation Function. The lag- ℓ autocorrelation function (ACF) of a stationary time series X_t is defined as

$$\rho_\ell = \frac{\gamma_\ell}{\gamma_0},$$

where γ_k is the lag- k autocovariance function of X_t . Note that it is easy to see that (i) $\rho_0 = 1$, (ii) $|\rho_\ell| < 1$, and (iii) $\rho_\ell = \rho_{-\ell}$.

Ergodicity: Since we often have a single realization from the time series under study, we must estimate the parameters of a particular time series model using observations of this realization. The basic reason that we can do so is the theory of *ergodicity*. This is another time invariant property we shall use. A simple way to discuss ergodicity is as follows:

Consider the random variable X_t . The traditional way to estimate the mean of this random variable is to have a *random sample* of m observations drawn from the distribution of X_t . Denote the sample by $X_{t,1}, \dots, X_{t,m}$. Then, the mean μ of X_t is estimated by

$$\tilde{\mu} = \frac{1}{m} \sum_{i=1}^m X_{t,i}.$$

By the law of large number, we have $\tilde{\mu} \rightarrow_p \mu$ as $m \rightarrow \infty$ provided that μ exists. This estimate $\tilde{\mu}$ is an “ensemble” average.

In time series analysis, we have only **ONE** realization. That is, we have only one observation X_t at time t . How can we estimate the mean of X_t ? Recall that for a stationary time series the distribution of X_t is the same as that of X_s . Consequently, we can treat the single realization X_1, X_2, \dots, X_n as a sample of n observations from the underlying distribution. Then, a natural way to estimate the mean μ of X_t is

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n X_t.$$

Here $\hat{\mu}$ is a “time” average. This type of estimate can only be justified if $\hat{\mu} \rightarrow_p \mu$ as $n \rightarrow \infty$.

More generally, for a time series X_t , the question is

$$\hat{r}_k = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})(X_{t+k} - \bar{X}) \stackrel{?}{=} \gamma_k = \text{Cov}(X_t, X_{t+k}),$$

where $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$. If a time series satisfies the requirement that the “time” averages converge to the “ensemble” averages, then the series is said to be ergodic.

Not all stationary time series are ergodic. However, all stationary, linear Gaussian time series are ergodic.

Note that the method of moments in time series analysis depends on ergodicity. What is the method of moments?

Linear time series and Wold decomposition: The simplest time series is a sequence of *iid* $N(0, \sigma^2)$:

$$\dots, a_{-2}, a_{-1}, a_0, a_1, a_2, \dots \text{ or simply } \{a_t\}_{t=-\infty}^{\infty}.$$

This series is called a Gaussian white noise series.

If $\{a_t\}$ are *iid*, but not Gaussian, then we have a strictly stationary series.

Linear Time Series. A univariate time series X_t is linear if it can be written as

$$X_t = \mu + a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i},$$

where $\{a_t\}$ is an *iid* sequence. The requirement that a_t are *iid* is rather strong. In practice, X_t may contain some deterministic trend component or is subjected to the effect of exogenous variables. We shall discuss this definition further when we introduce nonlinear time series later.

White Noise. A white noise series $\{a_t\}$ is defined as follows: (1) $E(a_t) = 0$ for all t , (2) $E(a_t^2) = \sigma^2$ for all t , and (3) $\text{Cov}(a_t a_s) = 0$ for $t \neq s$. That is, a white noise series is a sequence of uncorrelated random variables with mean zero and variance σ^2 . Note that a white noise series is serially uncorrelated, but not necessarily serially independent.

If we want to generate a series which is non-independent, we can take linear combinations of white noise terms:

$$X_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = (1 + \psi_1 B + \psi_2 B^2 + \cdots) a_t.$$

This is a one-sided linear filter of the a_t series; we average the current and past values of the a_t 's to generate the observations X_t 's. A process generated in this way is called a linear process or more specifically a moving average process.

Note that the X_t process considered above has zero mean; we can simply add a constant term μ to the right hand side to X_t so that it has a non-zero mean.

If X_t is weakly stationary, we require that its variance exist which in turn requires that

$$\sum_{i=1}^{\infty} \psi_i^2 < \infty.$$

The autocovariance function of $\{X_t\}$ is then given by

$$\gamma_k = \text{Cov}(X_t, X_{t+k}) = E(X_t X_{t+k}) = E\left[\left(\sum_{i=0}^{\infty} \psi_i a_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j a_{t+k-j}\right)\right] = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{k+i},$$

where $\psi_0 = 1$.

A more convenient way of obtaining the autocovariances of a linear process is by using the moment generating function. The function is defined

$$\Gamma(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k.$$

This generating function just serves to store the sequence of autocovariances in a convenient form with the device that the k -th coefficient is γ_k . Generating functions of this sort are

useful in many ways, for instance, they are convenient in book-keeping. By substituting the generating function in the formula for the autocovariance, we can obtain

$$\Gamma(z) = \sigma^2 \psi(z) \psi(z^{-1}).$$

All linear time series models have an infinite moving average representation. That is, any linear time series model can be written in the form of a moving average model of order infinity. The only difference between different models is the different restrictions on the ψ weights. The general linear model considered here has even greater applicability than one might think. We shall gradually see the flexibility of the model in this course. Here we simply rely on a very important theorem due to Wold which states that **any stationary** process can be decomposed into two parts:

$$Y_t = D_t + X_t$$

where X_t is a general linear process, D_t is a deterministic process (a process which can be perfectly predicted under the MSE criterion from past values of the process) and $E(X_t D_s) = 0$ for all t and s . Even processes which are generated by non-linear functions of the observations but which are strictly stationary can be represented by a linear process of infinite order.

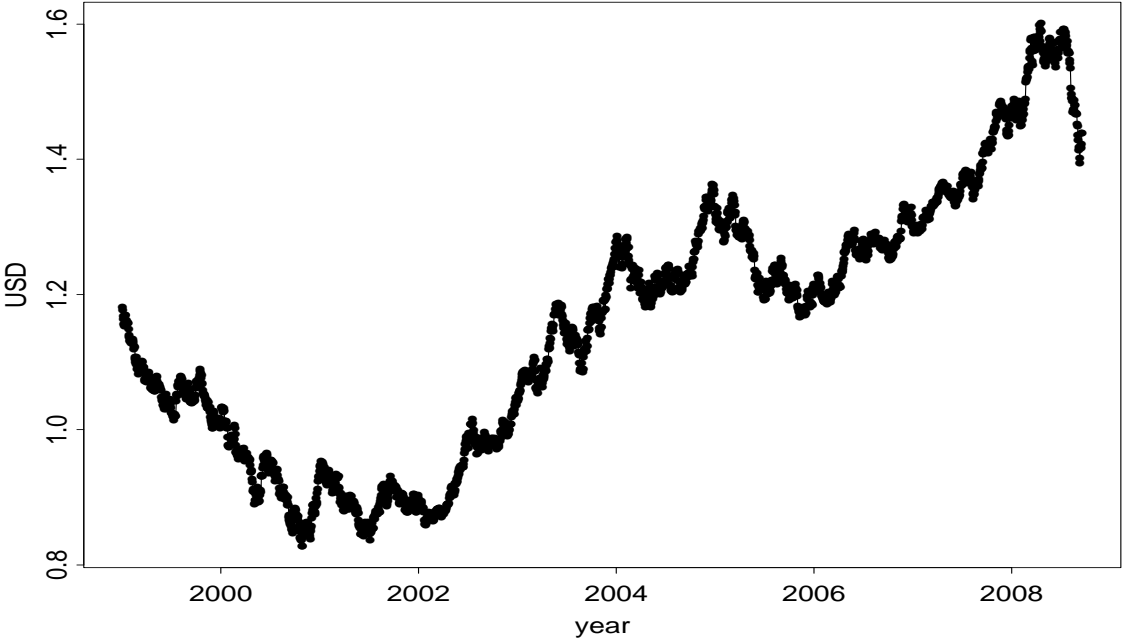
Note that in the Wold decomposition, X_t is a linear function of “uncorrelated” (not necessarily independent) process. Thus, a linear function in the Wold decomposition may still be a non-linear time series.

For those who are interested in a formal proof of Wold decomposition, see Brockwell and Davis (1991, page 187).

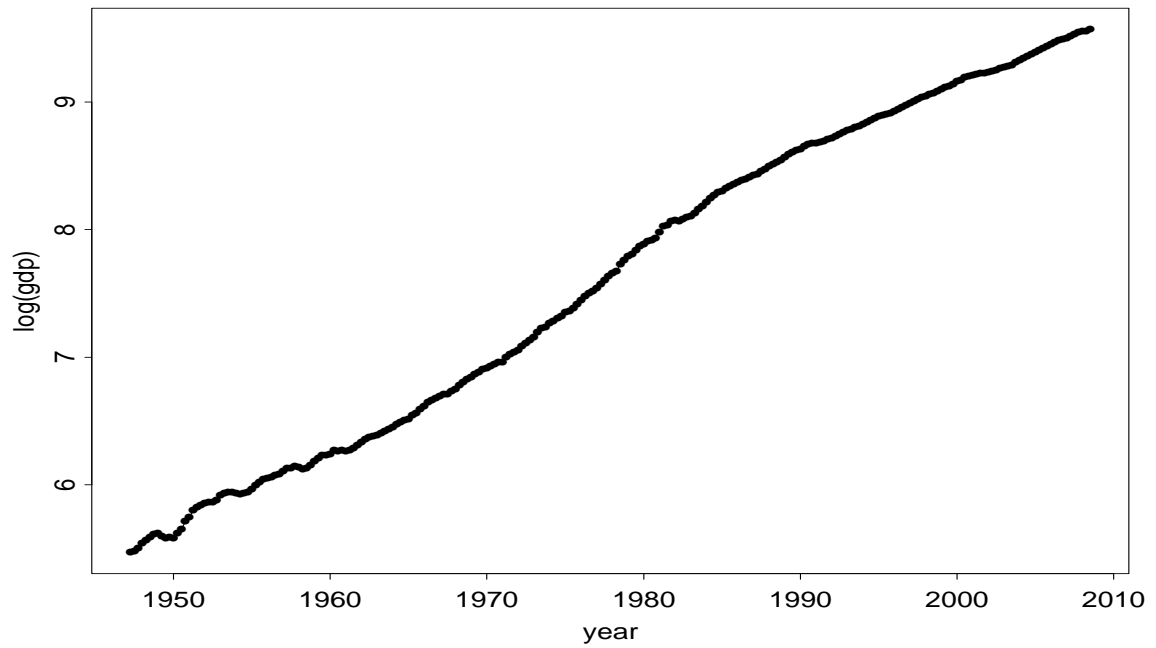
Example. Below are some examples of univariate time series.

1. Daily U.S. and EU exchange rates ($\$ = 1$ Euro): 1999-2008.
2. Quarterly U.S. GDP in billions dollars, seasonally adjusted annual rate. 1947-2008.
3. Monthly Producer Price Index: all commodities, index 1982=100, not seasonally adjusted, 1921-2008.
4. Monthly U.S. M2 Money Stock in billions of dollars (1959.1-2008.8)
5. Monthly U.S. Unemployment rate: civilian 16 years and older; seasonally adjusted, 1948-2008.
6. Monthly U.S. Total Nonfarm Payrolls: All employees, seasonally adjusted, 1939-2008.

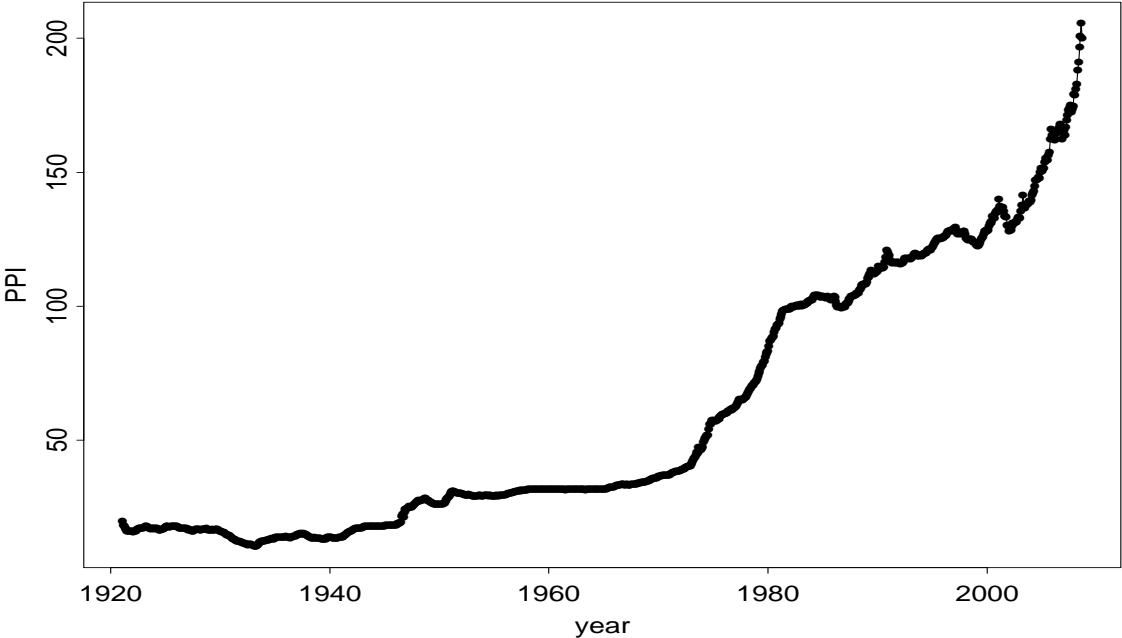
Daily US–EU exchange rate: 1999.1–2008.9



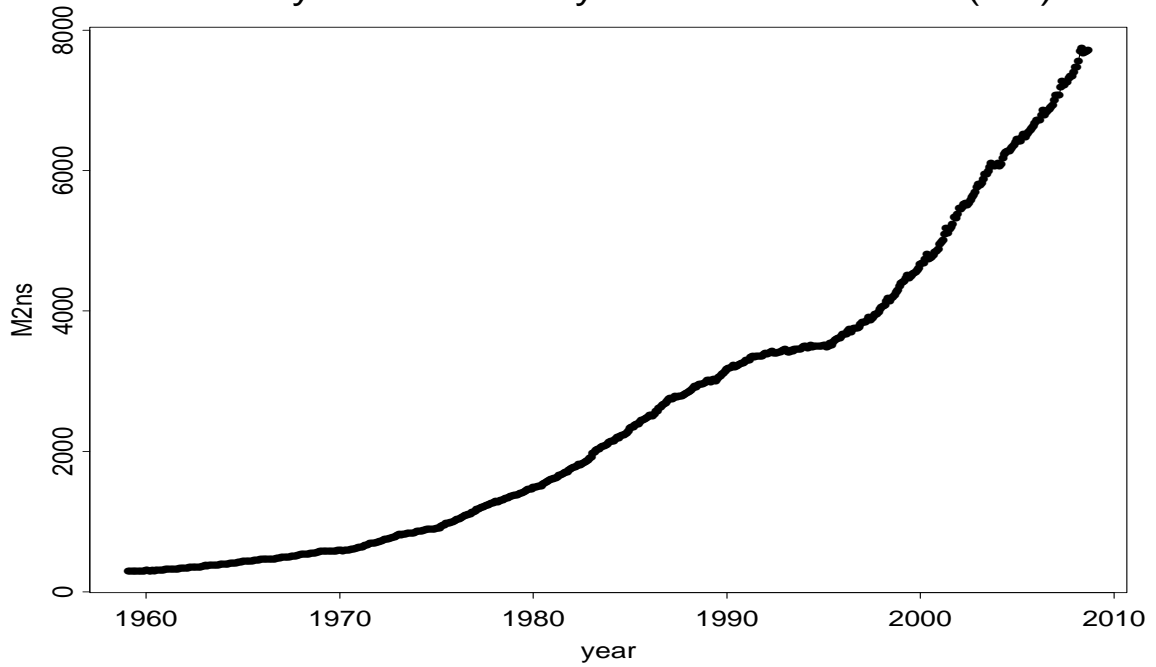
Quarterly US GDP: 1947.I–2008.II, seasonally adjusted



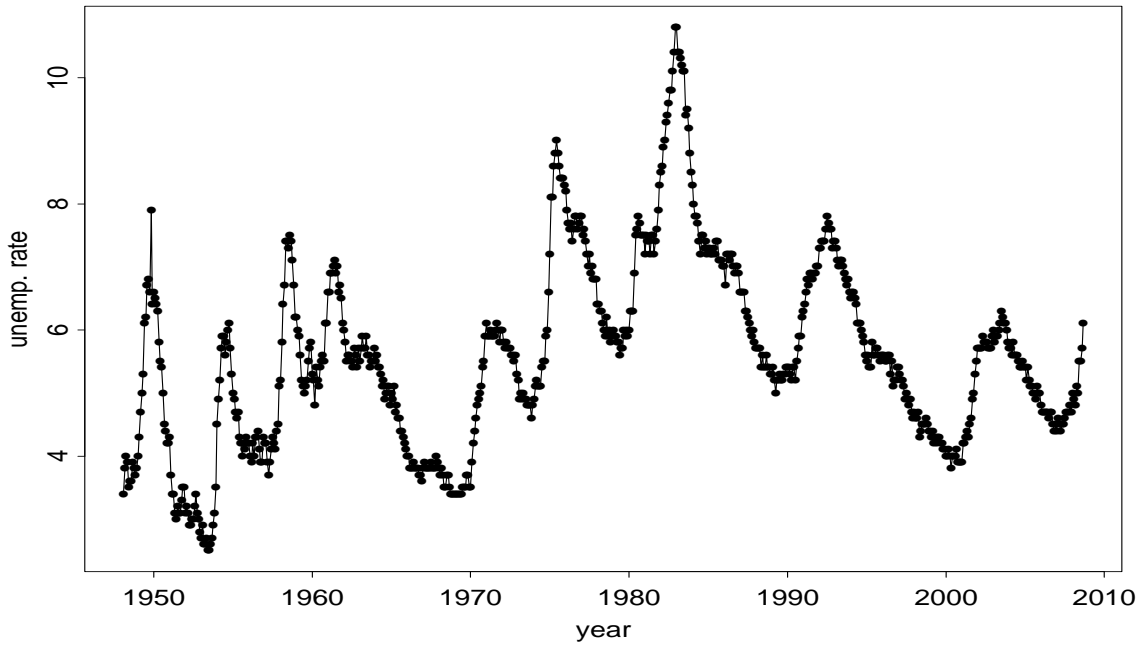
Monthly US Producer Price Index: 1921–2008



Monthly US M2 Money Stock: 1959–2008 (NS)



Monthly US unemployment rate: 1948–2008



Monthly US Nonfarm Payrolls: all employees

