Lecture 2: Linear Time Series Models

Bus 41910, Univariate Time Series Analysis, Mr. R. Tsay

A time series is a collection of random variables $\{X_t\}$ which are ordered by "time". In most economic and business applications, we only observe one *realization* from the time series which is called the sample path. The goal of time series analysis is to make inference of the process 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 reasonably approximate the process under study.

Choice of a model for a time series involves evaluation of a joint distribution function for the sample data:

$$F(x_1, x_2, \cdots, x_n) = Pr(X_1 \le x_1, \cdots, X_n \le x_n)$$

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. Obviously, to succeed in modeling the process, we must restrict the class of joint distributions under consideration. Furthermore, to predict the future of the process, we must be able to identify some key features of the distribution of the process 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

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

In other words, X_t is 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

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

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. Thus, for a Gaussian time series, weak stationarity is equivalent to strict stationarity.

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 Guassian process, the sample $\boldsymbol{Z} = (X_1, \cdots, X_n)'$ has a multivariate normal distribution:

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

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

$$\boldsymbol{\Sigma} = \left[\begin{array}{cccccc} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-3} & \gamma_{n-2} \\ \vdots & & & & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_1 & \gamma_0 \end{array}\right]$$

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

In (linear) time series analysis, we focus mainly on various models which 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 Funaction. 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 $X_{t,1}, \cdots, X_{t,m}$. 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 $\tilde{\mu} \to_p \mu$ as $m \to \infty$ 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 a stationary time series is time invariant so that the distribution of X_t is the same as the distribution of X_s. Consequently, we can treat the single realization $X_1, X_2, \cdots, X_n$ as a sample of n observations from a 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 sort 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

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

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* $N(0, \sigma^2)$:

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

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 be contains some deterministic trend component or is subjected to the effect of exogenerous variables. We shall discuss the 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) $\text{Vov}(a_t a_s) = 0$ 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

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

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 to keep track of the 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. Logarithum of electric power consumption (KWHC) of food industry of the U.S. The data were for the 15th day of each month from 1972 to 1993.

2. Daily U.S. and EU exchange rates

3. Quarterly U.S. real GDP in billions of chained 2000 dollars, seasonally adjusted. 1947-2005.

4. Monthly Comsumer Price Index for all urban consumers (all items), not seasonally adjusted, Index 1982-84 = 100, 1921-2005.7.

5. Monthly U.S. M2 Money Stock in billions of dollars (1959.1-2005.7)

6. Monthly U.S. Producer Price Index: all commodities, index 1982 = 100, not seasonally adjusted, 1921 to 2005.

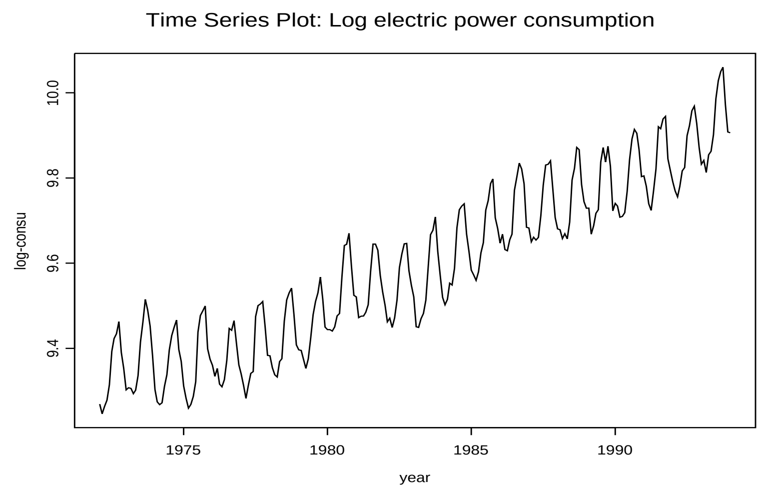

Figure 1: Logarithm of electric power consumption

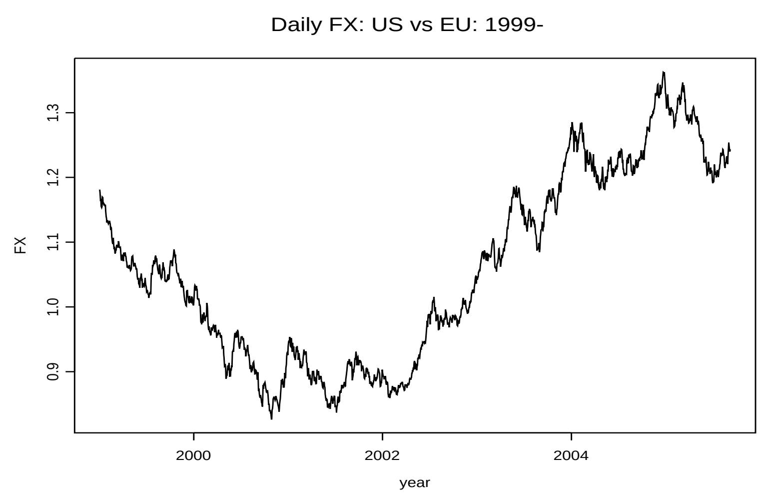

Figure 2: Daily exchange rates: Dollar vs Eurp

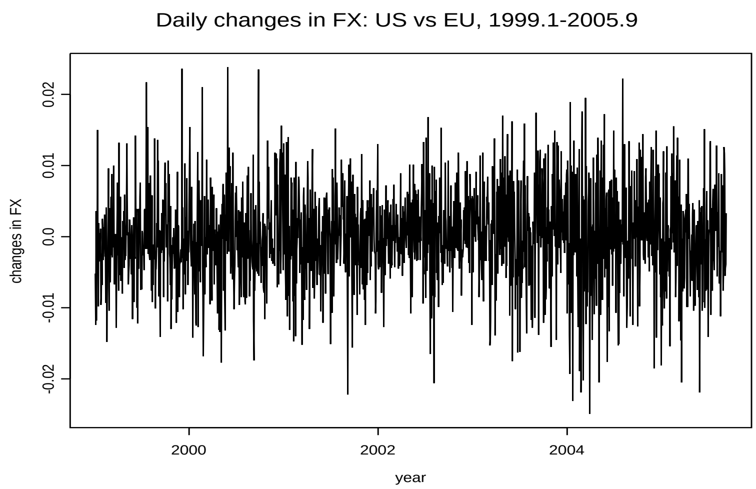

Figure 3: Changes of daily exchange rates: Dollar vs Eurp

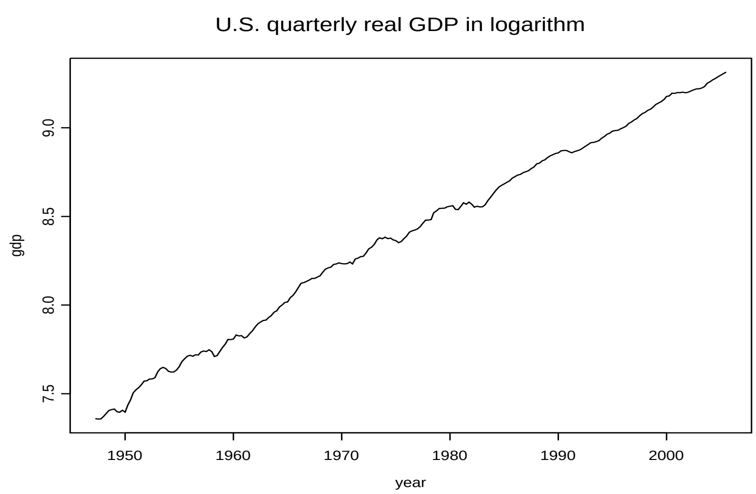

Figure 4: Quarterly U.S. real GDP, seasonally adjusted

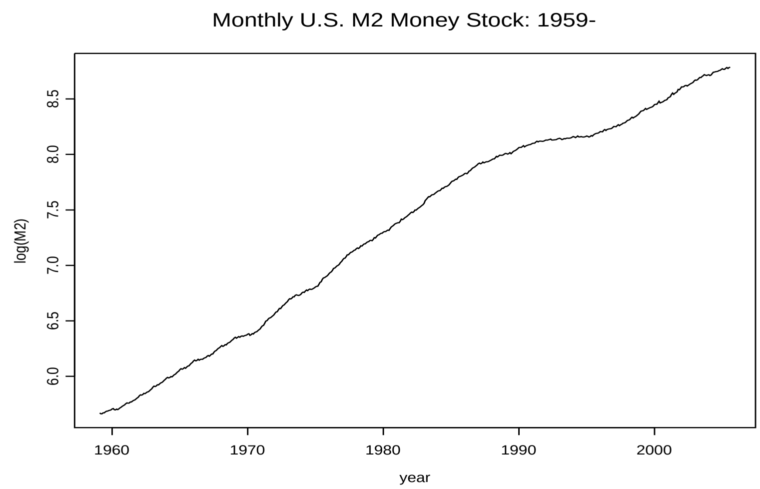

Figure 5: Monthly logarithms of M2 Money Stock, not seasonally adjusted

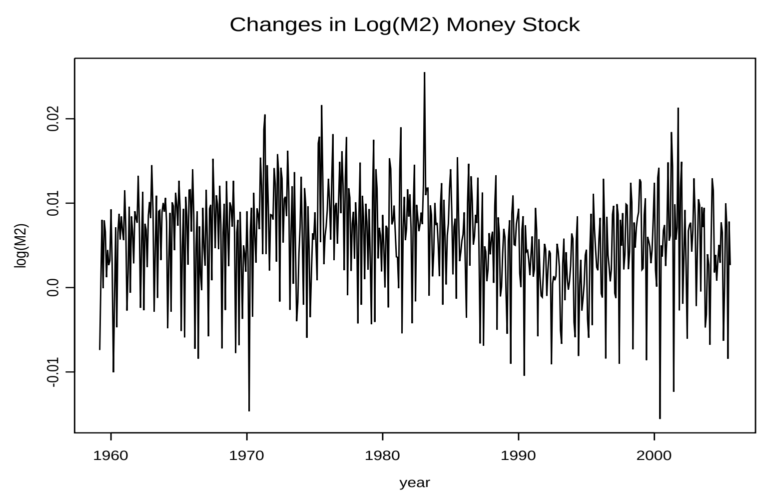

Figure 6: Changes in Monthly M2 Money Stock