Lecture 3: Unit Roots and ARIMA Models

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

Autoregressive Integrated Moving-Average (ARIMA) models consist of unit-root non-stationary time series which can be rendered stationary by the difference operation. They are useful models in real applications. The well-known exponential smoothing model for forecasting is just a special case of ARIMA(0,1,1) models; details are given later in forecasting.
Many economic time series contain unit root. They are referred to as *integrated* processes. A formal definition of unit root will be given shortly.

A. Difference Operator

The first differenced series $\triangle Z_t$ of a time series $\{Z_t\}$ is defined by

$$\triangle Z_t = Z_t - Z_{t-1} = (1-B)Z_t.$$

Since $(1-B)(Z_t-\mu) = (1-B)Z_t - (1-B)\mu = (1-B)Z_t$, the mean value vanishes after the first difference. Thus, from the differenced series $\triangle Z_t$, no information on the mean level of Z_t is available. $\triangle Z_t$ is the *increment* process of Z_t.
The second differenced series $\triangle^2 Z_t$ is defined by

$$\triangle^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} = (1-B)^2 Z_t.$$

Since

$$\begin{aligned}(1-B)^2(Z_t+c+\beta t) &= (1-B)^2 Z_t + (1-B)^2 c + (1-B)^2(\beta t)\\ &= (1-B)^2 Z_t + \beta t - 2[\beta(t-1)] + \beta(t-2) = (1-B)^2 Z_t.\end{aligned}$$

Both the mean level and time-trend vanish after the second difference. $\triangle^2 Z_t$ is to study the acceleration of the Z_t process.
In general, we can define a d-th order differenced process $\triangle^d Z_t$ by $\triangle^d Z_t = (1-B)^d Z_t$. It is easy to see that any time-polynomial of order less than d vanishes after the d-th difference. In applications, first difference is often sufficient to transform a time series into a stationary one. Occasionally, the second difference is used. Other higher-order differences are rarely used.

B. The ARIMA(p,d,q) Model

A time series Z_t follows an ARIMA(p,d,q) model if $\triangle^d Z_t$ is an ARMA(p,q) process. Thus, the model of Z_t is

$$\phi(B)(1-B)^d Z_t = c + \theta(B)a_t$$

where c is a constant, $\phi(B)$ and $\theta(B)$ are defined as before. This is a general model for (unit-root) non-stationary time series. To gain insight, it pays to study some simple cases.
Remark: By writing $\Phi(B) = \phi(B)(1-B)^d$, which is a polynomial of order $p+d$, we can treat an ARIMA(p,d,q) model as an ARMA$(p+d,q)$ model with the understanding that some of the zeros of $\Phi(B)$ are on the unit circle.

Below we consider some special cases of unit-root nonstationary time series.

C. The Random Walk or ARIMA(0,1,0) Model

A time series Z_t follows a random walk if

$$Z_t = Z_{t-1} + a_t.$$

That is, the first differenced series $\triangle Z_t = a_t$ is a white noise series. This model has been widely used and studied in the literature. It is a common model used to describe the behavior of a stock price (or log price).

Alternatively, the model can be written as

$$Z_t = \phi Z_{t-1} + a_t \quad \text{with} \quad \phi = 1.$$

Since $\phi = 1$, the zero of $1-\phi B$ is **on** the unit circle, the model thus has a unit root. This also gives the definition of a unit root. Later on we shall consider the least squares estimate of the coefficient ϕ for a random-walk model which is the unit-root problem commonly discussed in the literature.

The random-walk model is non-stationary. Writing the model as

$$Z_t = \frac{1}{1-B} a_t = a_t + a_{t-1} + a_{t-2} + \cdots$$

we see that the process is the sum of the current and past innovations. The effect of any past innovation $a_{t-\ell}$ on Z_t is persistent, where $\ell > 0$. In fact, the effect never decays or disappears. This feature is referred to as strong memory in the time series literature. Therefore, the process Z_t does not have finite variance as t goes to infinite. In practice, we assume that the process starts at some time point m of the remote past. Then,

$$Z_t = \sum_{i=m}^{t} a_{t-i}$$

so that $\text{Var}(Z_t) = (t-m)\sigma^2$, which is a linear, increasing function of t.

D. Random Walk with Drift

A slightly more general, yet still simple, unit-root nonstationary time series is the random-walk process with drift. The model is

$$Z_t - Z_{t-1} = c + a_t,$$

where c is a non-zero constant. For simplicity, we assume that the process starts at time t = 0. Then, we have

$$Z_t = \frac{1}{1-B} c + \frac{1}{1-B} a_t = ct + \sum_{i=0}^{t} a_{t-i}.$$

The constant "c", therefore, is the "slope" of a time trend of Z_t. This highlights one of the differences between stationary and unit-root non-stationary time series. For stationary time series, the constant term c is related to the mean. On the other hand, for the first differenced series of a unit-root nonstationary process the constant term represents the slope of a time trend in the original series.
When $c \neq 0$, the behavior of this process is very different from that of a pure random walk. Thus, in unit-root study, the two processes have different asymptotic properties.

E. ARIMA(0,1,1) Model
This is perhaps the most commonly used model in forecasting. It is the exponential smoothing model. The general form of the model is

$$Z_t - Z_{t-1} = c + a_t - \theta a_{t-1}, \quad \text{with} \quad |\theta| < 1.$$

Following the common practice, we shall assume $c = 0$. Since the model is invertible, the π-weights are $\pi_i = \theta^{i-1}(1-\theta)$ for $i \geq 1$. Thus,

$$Z_t = (1-\theta)Z_{t-1} + \theta(1-\theta)Z_{t-2} + \theta^2(1-\theta)Z_{t-3} + \cdots + a_t = \sum_{i=1}^{\infty} \theta^{i-1}(1-\theta)Z_{t-i} + a_t.$$

Notice that $\sum_{i=1}^{\infty} \pi_i = \sum_{i=1}^{\infty} \theta^{i-1}(1-\theta) = (1-\theta)\sum_{i=1}^{\infty} \theta^{i-1} = 1$. The current value Z_t is, therefore, a weighted average of the past values plus an innovation. The weights decay exponentially at the rate θ. As will be seen later in forecasting, such a weighted average makes common sense as the most recent past values are more relevant than the remove past in determining the current value.
If $0 < \theta < 1$, the weights remain positive. This is the exponential smoothing model in the forecasting literature. The only difference is that in the forecasting literature the rate θ is determined usually by minimizing sum of squares of one-step ahead forecast errors or by pre-specification. In time series analysis, we often estimate θ by the (exact) maximum likelihood method.
The ψ-weights of the model are $\psi_i = (1-\theta)$ for all $i > 0$. Thus,

$$Z_t = a_t + (1-\theta)a_{t-1} + (1-\theta)a_{t-2} + \cdots.$$

It is seen that the effect of past innovations on Z_t is persistent, but at a discounted rate $(1-\theta)$ when $\theta > 0$.
This integrated moving average model can also be viewed as follows: The "true" time series is a random walk

$$Y_t - Y_{t-1} = b_t$$

where $\{b_t\}$ is a Gaussian white noise process with mean zero and variance σ_b^2. However, we cannot observe Y_t directly. Instead, what we observe is contaminated by an independent Gaussian white noise $\{e_t\}$ with mean zero and variance σ_e^2, i.e. $Z_t = Y_t + e_t$. You may treat e_t as a measurement error such as rounding error. By applying $(1-B)$ to Z_t, we have

$$(1-B)Z_t = (1-B)(Y_t + e_t) = b_t + (1-B)e_t = b_t + e_t - e_{t-1}.$$

Let $w_t = b_t + e_t - e_{t-1}$. It is easily seen that $\mathrm{E}(w_t) = 0$, $\mathrm{Var}(w_t) = \sigma_b^2 + 2\sigma_e^2$, and

$$\mathrm{Cov}(w_t, w_{t-\ell}) = \begin{cases} -\sigma_e^2 & \text{for } \ell = 1 \\ 0 & \text{for } \ell > 1. \end{cases}$$

Thus, w_t is an MA(1) process and can be written as $w_t = a_t - \theta a_{t-1}$ where $\{a_t\}$ is a Gaussian white noise series with parameters θ and σ_a^2 satisfying

$$(1+\theta^2)\sigma_a^2 = \sigma_b^2 + 2\sigma_e^2 \quad \text{and} \quad \theta\sigma_a^2 = \sigma_e^2.$$

Remark: The above illustration points out an important message. Consider an AR(p) time series, say $\phi(B)Y_t = b_t$. Very often we cannot observe Y_t precisely. Instead, only a contaminated version of it is observed, say $Z_t = Y_t + e_t$. Here, among many possible sources, e_t can simply represent a measurement error which, for simplicity, is serially uncorrelated. Then, we have

$$\phi(B)Z_t = \phi(B)(Y_t + e_t) = b_t + \phi(B)e_t$$

It is easy to check that the right hand side of the prior equation is an MA(p) process. Therefore, Z_t follows an ARMA(p, p) model. This is one of the reasons for using MA models. This point, however, appears not to be fully appreciated in the economic literature. Of course, MA parameters are usually hard to interpret and the AR models are "relatively" easier to estimate.

F. <u>ARIMA(0,2,2) Model</u>

Another model often used in forecasting is the ARIMA(0,2,2) model

$$(1-B)^2 Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$$

where the zeros of $\theta(B)$ are outside the unit circle. Write the model as

$$(1-B)(1-B)Z_t = (1-\lambda_1 B)(1-\lambda_2 B)a_t$$

and define a new time series W_t by $(1-B)W_t = (1-\lambda_2 B)a_t$. Clearly, this newly defined series is an exponential smoothing process. That is, W_t follows an ARIMA(0,1,1) model. It is also easy to see that

$$(1-B)Z_t = (1-\lambda_1 B)W_t,$$

saying that Z_t is an exponential smoothing model with an exponential innovational series. Thus, Z_t is referred to as a double exponential smoothing model.

It is informative to calculate the π-weights and ψ-weights of this ARIMA(0,2,2) model. (Exercise!) The double exponential smoothing model uses two parameters θ_1 and θ_2 (or equivalently λ_1 and λ_2) to control the decay rate of the weights for past values in computing the current value. It is, therefore, more flexible than the exponential smoothing model. However, the ARIMA(0,2,2) model is "relatively" harder to apply in practice.

Remark: In real applications, it is rare that one needs more than the second-order differencing $(1-B)^2$ to render a non-periodic time series stationary. Only in the velocity of a missile, I experienced the need of using $(1-B)^3$. Another case that higher-order difference might be necessary is to model the annual maximum execution speed of modern computer. I have not encountered any business and economic time series that needs more than $(1-B)^2$.

G. Some examples: Time series with unot root is common. Some examples include

1. The VIX index of the Chicago Board of Options Exchange.
2. The U.S. quarterly GDP series
3. The U.S. consumer price index