SF2943 Time Series Analysis: Lecture 12

May 12, 2022

This lecture will introduce two nonstationary time series models: ARIMA and SARIMA.

1 ARIMA Models

We have already discussed the importance of the class of ARMA models for representing stationary processes. A generalization of this class, which incorporates a wide range of nonstationary processes, is provided by the ARIMA processes.

Definition 1.1 *Let $p, d, q \in \mathbb{N} \cup \{0\}$. A time series $\{X_t\}$ is an autoregressive integreted moving average process of order (p, d, q) (or an ARIMA(p, d, q) process) if*

$$Y_t \doteq (I - B)^d X_t$$

is a causal ARMA(p, q) process, where B is the backward shift operator, i.e., $BX_t = X_t - X_{t-1}$.

The definition means that $\{Y_t\}$ satisfies

$$\phi(B)Y_t = \theta(B)Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

where $\phi(z)$ and $\theta(z)$ are polynomials of degrees p and q, respectively, and $\phi(z) \neq 0$ for $|z| \leq 1$ (due to the causality). Because $Y_t \doteq (I - B)^d X_t$, we find that $\{X_t\}$ satisfies

$$\phi^*(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2), \tag{1.1}$$

where $\phi^*(z) \doteq \phi(z)(1 - z)^d$.

Remark 1.2 *In other words, a process is an ARIMA process if the process can be reduced to an ARMA process after differencing finitely many times.*

Remark 1.3 *Since* (1.1) *is an equivalent way to define an ARIMA(p, d, q) process $\{X_t\}$, we will use that equation sometimes.*

It then follows that $\phi^*(z)$ has a zero of order d at $z = 1$. Moreover, it turns out that the process $\{X_t\}$ is stationary if and only if $d = 0$, in which case it reduces to an ARMA(p, q) process.

Let's consider a particular example of an ARIMA(p, d, q) process.

Example 1.1 *Consider an ARIMA$(1,1,0)$ process $\{X_t\}$, that is, from (1.1), $\{X_t\}$ satisfies*

$$(I-\phi B)(I-B)X_t = Z_t, \quad \{Z_t\} \sim WN(0,\sigma^2)$$

with $|\phi|<1$ (for causality).

In this case, we can have an explicit expression of X_t in terms of $\{Z_t\}$ and X_0. First of all, recall that $Y_t = (I-B)X_t = X_t - X_{t-1}$ and satisfies

$$(I-\phi B)Y_t = Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0,\sigma^2).$$

Namely, $\{Y_t\}$ is an AR(1) process. From our previous lectures, we know that such Y_t can be rewritten as

$$Y_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

On the other hand,

$$X_t - X_0 = (X_t - X_{t-1}) + (X_{t-1} - X_{t-2}) + \cdots + (X_1 - X_0) = \sum_{k=1}^{t} Y_k.$$

Therefore,

$$X_t = X_0 + \sum_{k=1}^{t} Y_k = X_0 + \sum_{k=1}^{t}\sum_{j=0}^{\infty} \phi^j Z_{k-j}.$$

Using this expression, one can show that $\{X_t\}$ is not stationary. In particular, the variance of X_t increases at t increases.

Remark 1.4 *This example motivates the name of "integrated" in ARIMA since X_t is a summation (sort of integration) of an ARMA process Y_k from $k=1$ to $k=t$ (and plus X_0).*

Remark 1.5 *The name of ARIMA is consistent with the order (p,d,q) in the sense that p,d and q correspond to the orders of the AR, I and MA parts, respectively. Furthermore, they are also consistent with the representation*

$$\underbrace{\phi(B)}_{AR}\underbrace{(I-B)^d}_{I} X_t = \underbrace{\theta(B)Z_t}_{MA}.$$

An important feature of ARIMA processes is the following.

Proposition 1.6 *If $\{X_t\}$ is ARIMA(p,d,q) with $d \geq 1$, and consider a process $\{\tilde{X}_t\}$ satisfying*

$$\tilde{X}_t = X_t + g(t),$$

where $g(t)$ is a polynomial in t of degree $d-1$, then $\{\tilde{X}_t\}$ is also an ARIMA process with the same order (p,d,q). Moreover, if we define $Y_t = (I-B)^d X_t$ and $\tilde{Y}_t = (I-B)^d \tilde{X}_t$, then $Y_t = \tilde{Y}_t$.

Remark 1.7 *Simply speaking, this result says that adding a polynomial trend with degree $d-1$ to an ARIMA(p,d,q) process remains ARIMA with the same order. These two ARIMA processes share the same "underlying" ARMA process. From here we can also see that ARIMA models are useful for representing data with trend as well as without trend.*

We will only "partially" prove this proposition by considering a specific case.

Example 1.2 *Let $\{X_t\}$ be an ARIMA(p,d,q) with $d=2$. Consider $\tilde{X}_t = X_t + at + b$ for some $a, b \in \mathbb{R}$. Then observe that*

$$\begin{aligned}(I-B)^2(at+b) &= (I-B)(I-B)(at+b) = (I-B)[(at+b)-(a(t-1)+b)] \\ &= (I-B)a = a-a = 0.\end{aligned}$$

Thus,

$$(I-B)^2\tilde{X}_t = (I-B)^2(X_t+at+b) = (I-B)^2X_t + (I-B)^2(at+b) = (I-B)^2X_t.$$

This implies that $\tilde{Y}_t = (I-B)^2\tilde{X}_t = (I-B)^2X_t = Y_t$ and $\{\tilde{X}_t\}$ is also ARIMA(p,d,q) with $d=2$.

2 Choice of d

We know from the previous section that we can use ARIMA models to fit data with a trend. As long as we know how many times we should difference our data, i.e., how large d should be, then we can fit the remaining model parameters, including p, q, σ^2, using the methods that we mentioned before for fitting ARMA models. So how should we choose d in practice?

2.1 "Brute force"

One most straightforward way is by first plotting out the observations $\{x_1, \ldots, x_n\}$ and the sample acvf. If the data exhibit no apparent deviations from "stationarity" and have a rapidly decreasing sample acvf, we attempt to fit an ARMA model to the mean-corrected data using the techniques we mentioned in lecture 10. Otherwise[1], we transform the data by differencing, leading us to consider the class of ARIMA models. Then after the transformation, we go back to the previous step. Continue this procedure and hope that after finitely many times of differencing, we can fit an ARMA model to the transformed data.

2.2 Dicklet-Fuller test.

Another more systematic way is to use the Dicklet-Fuller test. We will motivate the test by considering ARIMA$(1,1,0)$ $\{X_t\}$ which satisfies

$$(I-\phi B)(I-B)X_t = Z_t.$$

[1] This happens when the sample acvf decreases "slowly" or the data looks very "nonstationary".

If our data truly comes from ARIMA$(1,1,0)$, then apparently we should transform our data by differencing once and fitting an ARMA model to the transformed data. However, multiplying out the left side of the above equality gives

$$X_t - (1+\phi)X_{t-1} + \phi X_{t-2} = Z_t.$$

Without considering the well-definedness, the above form suggests that $\{X_t\}$ is an AR(2) process with the AR polynomial

$$\phi(z) = 1 - (1+\phi)z + \phi z^2 = (1-\phi z)(1-z).$$

Of course, $\{X_t\}$ is not a well-defined AR(2) process since $\phi(z)$ has a root at $z=1$ (i.e., $\phi(1)=0$) which implies that $\{X_t\}$ can not be stationary. Nevertheless, this is not obvious when we only have data. We need to estimate the coefficients of $\phi(z)$, in this case ϕ and $-(1+\phi)$, from the data. Small errors of the estimates would make the corresponding estimate $\hat{\phi}(z)$ such that $\hat{\phi}(1) \neq 0$ (but might be close to 0). Therefore, intuitively speaking, if $\hat{\phi}(1) \approx 0$, we fit the data with ARIMA. Otherwise, if $\hat{\phi}(1)$ is "significantly" different from zero, then we fit the data with ARMA. A question remains to answer now is **how close is close**?

2.2.1 Unit roots in autoregressions (underdifferencing)

Consider a process $\{X_t\}$ which is AR(p) with mean μ given by

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

This model can be rewritten as (after some algebra, we will skip here)

$$\nabla X_t = \phi_0^* + \phi_1^* X_{t-1} + \phi_2^* \nabla X_{t-1} + \cdots + \phi_p^* \nabla X_{t-p+1} + Z_t, \tag{2.1}$$

where $\nabla X_t \doteq (I-B)X_t = X_t - X_{t-1}$, $\phi_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$,

$$\phi_1^* = \sum_{i=1}^{p} \phi_i - 1 \quad \text{and} \quad \phi_j^* = -\sum_{i=j}^{p} \phi_i, \quad j = 1, 2, \ldots, p.$$

If the AR polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ has a unit root at 1, then $0 = \phi(1) = -\phi_1^*$. Moreover, (2.1) becomes

$$\nabla X_t = \phi_0^* + \phi_2^* \nabla X_{t-1} + \cdots + \phi_p^* \nabla X_{t-p+1} + Z_t, \tag{2.2}$$

which suggests that the differenced series $\{\nabla X_t\}$ is an AR$(p-1)$ process. Consequently, testing the hypothesis of a unit root at 1 of the AR polynomial is equivalent to testing $\phi_1^* = 0$. We can estimate ϕ_1^* by using ordinary least squares (OLS) estimator, $\hat{\phi}_1^*$, found by regressing ∇X_t onto $1, X_{t-1}, \nabla X_{t-1}, \ldots, \nabla X_{t-p+1}$.

It turns out that, under the unit root assumption $\phi_1^* = 0$, there exists some kind of CLT (central limit theorem) result (proved by Dickey and Fuller) which guarantees that for large sample size n, the t-ratio

$$\hat{\tau}_\mu \doteq \frac{\hat{\phi}_1^*}{\widehat{\mathrm{SE}}(\hat{\phi}_1^*)}$$

has a certain limit distribution, where $\widehat{\mathrm{SE}}(\hat{\phi}_1^*)$ is the estimated standard error of $\hat{\phi}_1^*$. Moreover, the $0.01, 0.05$, and 0.10 quantiles of the limit distribution of $\hat{\tau}_\mu$ are $-3.43, -2.86$, and -2.57, respectively. The augmented Dickey-Fuller test then rejects the null hypothesis $H_0 : \phi_1^* = 0$, at say, level 0.05 if $\hat{\tau}_\mu < -2.86$.

Further extensions of the augmented Dickey-Fuller test to ARMA(p, q) model can be found in Said and Dickey (1984).

Remark 2.1 *To sum up, we can use the augmented Dickey-Fuller test (or a generalized version) to see if our data requires further differencing.*

2.2.2 Unit roots in moving averages (overdifferencing)

In the previous subsubsection, we introduce a method for testing if our data needs to be differenced. A related problem is to understand if our data has been overdifferenced. We will present a method for testing overdifferencing.

Let $\{X_t\}$ be a causal and invertible ARMA(p, q) process satisfying

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

Then consider the differenced series $\{Y_t\}$ with $Y_t = (I - B)X_t$ and multily the above equation by $I - B$, since the operators $\phi(B)$ and $I - B$ are exchangeable, we find

$$\phi(B)Y_t = \phi(B)(I - B)X_t = (I - B)\phi(B)X_t = (I - B)\theta(B)Z_t.$$

If we define $\theta^*(z) = \theta(z)(1 - z)$, then the above equation becomes

$$\phi(B)Y_t = \theta^*(B)Z_t.$$

This shows that $\{Y_t\}$ is an ARMA$(p, q + 1)$ process with the AR polynomial $\phi(z)$ and the MA polynomial $\theta^*(z)$ which has a unit root at $z = 1$. Consequently, testing for a unit root in the MA polynomial is equivalent to testing that the time series has been overdifferenced.

In the book, they confine the discussion of unit root tests for the MA polynomial to MA(1) models. Please see the book for more details. We will omit it here. The general case is considerably more complicated and not fully resolved.

Remark 2.2 *To sum up, we can use a test, analogous to (but not the same as) the augmented Dickey-Fuller test, to see if our data has been overdifferenced.*

3 SARIMA Models

As we mentioned earlier, ARIMA models can be useful for data with a trend, but how about seasonality? It turns out that if our data is periodic with period s, then instead of applying differencing $I - B$, we apply $I - B^s$. This type of differencing removes the seasonal component of period s.

Definition 3.1 *Consider $d, D \in \mathbb{N} \cup \{0\}$. The time series $\{X_t\}$ is a seasonal ARIMA process of order $(p, d, q) \times (P, D, Q)_s$, also known as SARIMA$(p, d, q) \times (P, D, Q)_s$, if the differenced series*

$$Y_t \doteq (I - B)^d (I - B^s)^D X_t$$

is a causal ARMA process defined by

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2), \tag{3.1}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, and $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$.

Remark 3.2 *If we define new notations*

$$\phi^*(z) \doteq \phi(z)\Phi(z^s) \quad \text{ and } \theta^*(z) \doteq \theta(z)\Theta(z),$$

then (3.1) *becomes*

$$\phi^*(B)Y_t = \theta^*(B)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2), \tag{3.2}$$

where $\phi^*(z), \theta^*(z)$ are polynomials of degree $p + sP$ and $q + sQ$, respectively, whose coefficients can all be expressed in terms of $\phi_1, \ldots, \phi_p$, $\Phi_1, \ldots, \Phi_P$, $\theta_1, \ldots, \theta_q$, and $\Theta_1, \ldots, \Theta_Q$.

Remark 3.3 *SARIMA is very similar to ARIMA. The additional seasonal parts are denoted by capital letters.*

An advantage of SARIMA models over the classical decomposition

$$X_t = m_t + s_t + Y_t$$

with a deterministic seasonal component m_t is that SARIMA models allow randomness in the seasonal cycle. This property makes SARIMA more realistic models.

An important case of SARIMA models is the pure seasonal ARMA$(P, Q)_s$ processes. To be precise, consider a time series $\{X_t\}$ satisfying

$$\Phi(B^s)X_t = \Theta(B^s)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2).$$

Then one can show that the acf and pacf are zero for all lags that are not multiples of s. For lags that are multiples of s, we can compute the values as before. Let's consider a concrete example.

Example 3.1 *Consider $\{X_t\}$ to be the pure seasonal ARMA$(1, 0)_{12}$ process. Namely, $\{X_t\}$ satisfies*

$$X_t = \Phi X_{t-12} + Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2).$$

Recall that the acvf for an AR(1) process is

$$\gamma(h) = \phi^{|h|} \frac{\sigma^2}{1 - \phi^2}.$$

Therefore, the acvf for the pure seasonal ARMA$(1, 0)_{12}$ process is

$$\gamma(h) = \begin{cases} \Phi^{|h|/12} \frac{\sigma^2}{1-\Phi^2}, & \text{if } h = 0, 12, 24, \ldots \\ 0, & \text{otherwise} \end{cases}.$$

Example 3.2 *Consider $\{X_t\}$ to be the pure seasonal ARMA$(0,1)_{12}$ process. Namely, $\{X_t\}$ satisfies*

$$X_t = Z_t + \Theta Z_{t-12}, \quad \{Z_t\} \sim WN(0, \sigma^2).$$

Recall that the acvf for a MA(1) process is

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2, & \text{if } h = 0 \\ \theta\sigma^2, & \text{if } h = 1, -1 \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, the acvf for the pure seasonal ARMA$(0,1)_{12}$ process is

$$\gamma(h) = \begin{cases} (1+\Theta^2)\sigma^2, & \text{if } h = 0 \\ \Theta\sigma^2, & \text{if } h = 12, -12 \\ 0, & \text{otherwise} \end{cases}.$$