Lecture 4

We ended **Lecture 3** with the notion of differencing a time series successively till it becomes stationary and had mentioned that more would follow. Before that, for computational ease, we will show that a stationary $AR\ (p)$ model with possibly non-zero intercept α can be equivalently written as a stationary $AR\ (p)$ model with zero intercept - just like in *Probability*: to say that X is a random variable with mean μ, is *equivalent* to saying that $X - \mu$ is a random variable with mean zero.

For proving the above, we need to only recognise that for a stationary $AR\ (p)$ model with non-zero intercept α, the mean function μ_t is a constant (called, the stationary mean) given by

$$\mu = \alpha/(1 - \sum_{i=1}^{p} \beta_i).$$

Now, a stationary $AR\ (p)$ model with the stationary mean μ given above is

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t.$$

Adding and subtracting $\mu \sum_{i=1}^{p} \beta_i$ on the right side of the above equation and rearranging the terms, we get

$$Y_t = [\alpha + \mu \sum_{i=1}^{p} \beta_i] + \sum_{i=1}^{p} \beta_i [Y_{t-i} - \mu] + \epsilon_t.$$

Substituting for α in terms of μ in the first bracketed term on the right side above and simplifying it, we see that

$$Y_t = \mu + \sum_{i=1}^{p} \beta_i [Y_{t-i} - \mu] + \epsilon_t$$

or equivalently,

$$Y_t - \mu = \sum_{i=1}^{p} \beta_i [Y_{t-i} - \mu] + \epsilon_t.$$

Letting $X_t = Y_t - \mu$ and $X_{t-i} = Y_{t-i} - \mu$ we get

$$X_t = \sum_{i=1}^{p} \beta_i X_{t-i} + \epsilon_t,$$

where all the X_t's above are random variables with mean zero.

So from now on, without loss of generality, whenever we consider a stationary $ARMA$ model, we assume that its stationary mean is zero.

Back to differencing

Now, we will give the more generalised definition of an $ARMA$ model. Towards that, let $\{Y_t^{(d)}\}$ denote the d-differenced (that is, differenced d times) series of the original given time series $\{Y_t\}$, $d = 0, 1, 2, 3, ...$ with $\{Y_t^{(0)}\} \equiv \{Y_t\}$.

Autoregressive-Integrated-Moving Average Model ($ARIMA$ Model)

Let $\{Y_t\}$ be a given $ARMA$ (p, q) model. Then $\{Y_t\}$ is called an $ARIMA$ (p, d, q) model if $\{Y_t^{(d)}\}$ is stationary, where d is defined by

$$d = \min \{k \mid \{Y_t^{(k)}\} \text{ is stationary}\}, \quad d = 0, 1, 2, ...$$

Remark: Note that, if the original given $ARMA$ (p, q) model $\{Y_t\}$ is itself stationary, then $\{Y_t\}$ is $ARIMA$ $(p, 0, q)$, as no differencing is required on the original data. That is, a stationary $ARMA$ (p, q) is precisely $ARIMA$ $(p, 0, q)$.

If a given time series data exhibits a roughly linear trend, then the method of differencing once will make it stationary. But, if the data shows an exponential trend, one way is to apply the logarithmic transformation on the original data and then differencing this logarithmic data. In fact, as we will see later, this is the method to make financial time series data stationary (almost all of them like stock price, exchange rate, market index etc.).

Recall that our fundamental aim of time series analysis is to forecast future value(s) based on a given time series data set of size T (say). So, in what follows, we shall list the required steps involved in the analysis of time series data to achieve our goal of forecasting. After listing the steps, we shall briefly explain about each step involved.

<u>Note:</u> **In all our analysis of $ARIMA$ models, it will be assumed that the white noise sequence $\{\epsilon_t\}$ is Gaussian; that is, $\{\epsilon_t\}$ is a sequence of i.i.d. normal random variables with ,ean zero and variance σ^2.**

Steps involved in time series forecasting

Step 1: *Prepare the given time series data well to keep it ready for subsequent analysis.*

Step 2: *Test for stationarity of this prepared data. If it is not stationary, apply (one or more) of the methods of transformation described earlier to make it stationary.*

Step 3: *Fit an appropriate $ARMA(p, q)$ model to the resulting stationary data from the above step.*

Step 4: *Estimate the parameters of the fitted $ARMA$ (p, q) model.*

Step 5: *Forecast the future value(s) for this fitted stationary model. Then, get the forecast values for the original non stationary data by inverse transformation of these forecast values for the stationary model.*

Brief details on the above steps

Step 1 - Preparing the data set

Ensure that the observations from the data set are chronologically ordered, with the earliest (or, first) observation at the top and the latest (or, last) observation at the end, in the form of two columns. The first column is for the time stamp, and the second column for the observed values. Then, to get a feel of the data, plot its graph - clearly, with time as the x-axis and the observed values as the y-axis. If you get to see a linear or exponential trend, you can immediately make the conclusion that this data is **not** stationary (which can be verified formally in **Step 2**).

Get the descriptive statistics of this data set. Also, apart from the basic descriptive statistics, get (a) the Sample Kurtosis and (b) the percentage of the observed values lying in the interval (-3*Sample Standard Deviation, +3*Sample Standard Deviation). We will explain the need for the last two computed values later in our course.

Another important aspect is about the quality of the data to be used in the analysis. If there are missing observations, as it is a time series data, we can replace the missing observation by the method of interpolation (available in both R and Python). Similarly, if there is an inconsistency in the data - an extreme value or a value whose unit of measurement does not match with the other observations etc. - care should be taken to deal with that observation in terms of replacing it.

My personal rule of thumb for the data size is that *there should be at least 100* *observations* for building a model. In fact, I follow this thumb rule as it was advocated by two well-known Statisticians (Time Series experts) George E.P. Box and George C. Tiao (*cf* Box, G.E.P. and Tiao, G.C. (1975): "Intervention analysis with applications to economic and environmental problems." *Journal of the American Statistical Association* **Volume 70**: pp 70-92).

Step 2 - Test for stationarity

To test the stationarity of the data set prepared in **Step 1** above, we apply the following three test procedures.

PP Test: The Phillips Perron (PP) test tests the null hypothesis of whether a unit root is present in a time series sample, against a stationary alternative. That is, the null hypothesis is **H_0: The time series has a unit root and hence not stationary**, against the alternative hypothesis **H_1: The time series is stationary**.

ADF Test: The Augmented Dickey Fuller (ADF) test, tests the null hypothesis of whether a unit root is present in a time series sample. That is, the null hypothesis is **H_0: The time series has a unit root and hence not stationary**, against the alternative hypothesis **H_1: The time series is stationary**.

KPSS Test: The Kwiatkowski Phillips Schmidt Shin (KPSS) test, tests a null hypothesis that an observable time series is stationary, against the alternative of a unit root. That is, the null hypothesis is **H_0: The time series is stationary**, against the alternative hypothesis **H_1: The time series has a unit root and hence not stationary**.

Remarks:

(i) If all the above three tests reject the stationarity, differencing needs to applied on the data, and the three tests should be applied again, on this *differenced/transformed* data. This procedure should be repeated till we achieve stationarity, as the forecasts will be made only for the stationary model. So, this procedure of differencing (if required) to get stationarity will result in a stationary $ARIMA\ (p, d, q)$ model (with $d = 0$ if the original data is stationary) for which the forecasts will be made.

(ii) The PP and the ADF tests are similar in the sense that the null hypothesis of non stationarity is the same for both the tests, whereas the null hypothesis of the KPSS test is that of stationarity. In general, all the three tests return the same verdict - stationary or not. Very rarely they differ.

Step 3 - Fit an appropriate $ARMA$ (p, q)

Fitting an appropriate $ARMA$ (p, q) model to the stationary data set from **Step 2** essentially involves in finding the appropriate values for p and q.

The most commonly used criterion used in finding the *appropriate* p and q is one the following two quantities: (a) the AIC (*Akaike Information Criterion*) and (b) the BIC (*Bayesian Information Criterion*). We just mention that these quantities are based on the likelihood function of the given model. Since we get a model for each choice of p and q, we get the associated values of the AIC (or the BIC) as $AIC(p, q)$ (or, $BIC(p, q)$). The *most appropriate* model is the one for which the computed AIC (or, BIC) value is the least. That is, the final model is $ARMA$ (p_0, q_0) where (p_0, q_0) is such that

$$AIC(p_0, q_0) = min_{(p,q)} \{AIC(p, q)\}$$

or

$$BIC(p_0, q_0) = min_{(p,q)} \{BIC(p, q)\}.$$

There is no evidence to suggest that one approach outperforms the other (that is, one of the two, AIC or BIC, is better) in a real application (*cf.* Ruey S. Tsay (2010), p. 49).

In practice, it is common to fit different models with each p and q ranging from 1 and 10 (this upper limit 10 will certainly depend on (a) size of the data, (b) the processing speed of the computer). So, we will get a square matrix of size 10, corresponding to the values of AIC (or, BIC).

Remark:

While fitting different $ARMA$ (p, q) models, you need to use the syntax for fitting an $ARIMA$ $(p, 0, q)$ model for each choice of the pair (p, q) while keeping $d = 0$. And, for each of the fitted model, the output will also contain the *maximum likelihood estimators, or the MLEs* of the associated parameters which are the parameters given by the intercept α and the different coefficients β_i's and θ_j's. So, along with the final best model, you will also get the MLEs of the parameters of this final chosen model.

Step 4 - Estimate the parameters of the fitted $ARMA\ (p, q)$ model

See the remark under **Step 3 - Fit an appropriate $ARMA\ (p, q)$ model...**.

Now, if no confusion would arise, we will use Y_t also to denote the observation y_t. We have the basic data set of size T, given by $\{Y_1, Y_2, ..., Y_T\}$ and from **Step 3** we have the best $ARMA$ model fit for this data given by

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t.$$

Step 3 also provides us with the MLEs of the model parameters, denoted by $\hat{\alpha}$, the $\hat{\beta}_i$'s and the $\hat{\theta}_j$'s.

Note that we only have the observations only on Y_t's and the errors ϵ_t's are **not** observable. But, we can construct or guess or estimate what the error was for each observation Y_t. The *residual* (or, the past forecast error) at time t, $t = 1, 2, ..., T$, denoted by e_t, is nothing but the "estimated error value" at t corresponding to the observation Y_t. It is defined by

$$e_1 = Y_1 - \hat{\alpha},$$

and

$$e_t = Y_t - \{\hat{\alpha} + \sum_{i=1}^{p} \hat{\beta}_i Y_{t-i} + \sum_{j=1}^{q} \hat{\theta}_j e_{t-j}\}.$$

Given the data set $\{Y_1, Y_2, ..., Y_T\}$, these residuals will play their role in the forecasts of $Y_{T+k}, k = 1, 2, ...,$ which we will discuss in what follows.

Step 5 - Forecasting

Now, apart from the original data of observations $\{Y_1, Y_2, ..., Y_T\}$, we also have the estimated data on the unobservable errors, given by the set of residuals $\{e_1, e_2, ..., e_T\}$. So, the total information we have with us is the set of $2T$ values of the T observations on Y_t and the T residuals e_t. We will denote this information set by $\mathcal{F}_T$, where

$$\mathcal{F}_T = \{Y_1, Y_2, ..., Y_T, e_1, e_2, ..., e_T\}.$$

Firstly, **we assume that the given time series $\{Y_t\}$ is stationary**. Now, our aim is to forecast Y_{T+1} based on the information $\mathcal{F}_T$ up to T, and then provide step-by-step forecasts of Y_{T+2}, Y_{T+3}, and so on. These are called *one-step-ahead, two-step-ahead, three-step-ahead* etc. forecasts based on the given information.

Let the forecast of Y_{T+1} be denoted by $\hat{Y_T}(1)$. Then, using the result that the best forecast (or, predictor) of Y_{T+1} given the information $\mathcal{F}_T$ up to time T, is the conditional expectation of Y_{T+1} given $\mathcal{F}_T$. That is, the best forecast of $Y_{T+!}$ based on $\mathcal{F}_T$ is

$$\hat{Y_T}(1) = E[Y_{T+1} \mid \mathcal{F}_T].$$

If our data $\{Y_1, Y_2, ..., Y_T\}$ is from a stationary $ARMA\ (p, q)$ given by

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

then the one-step-ahead forecast based on $\mathcal{F}_T$ is

$$\begin{aligned}
\hat{Y_T}(1) &= E[Y_{T+1} \mid \mathcal{F}_T] \\
&= E[(\hat{\alpha} + \sum_{i=1}^{p} \hat{\beta}_i Y_{(T+1)-i} + \sum_{j=1}^{q} \hat{\theta}_j e_{(T+1)-j} + \epsilon_{T+1}) \mid \mathcal{F}_T] \\
&= \hat{\alpha} + \sum_{i=1}^{p} \hat{\beta}_i Y_{(T+1)-i} + \sum_{j=1}^{q} \hat{\theta}_j e_{(T+1)-j} + E[\epsilon_{T+1}) \mid \mathcal{F}_T]
\end{aligned}$$

In the above, note that for each i and j, the respective $Y_{(T+1)-i}$'s, the MLEs of the parameters and, the residuals $e_{(T+1)-j}$'s are elements of $\mathcal{F}_T$ and hence all the terms on the right side are *constants*t as far the conditional expectation is concerned.

Additionally, noting that tha last term on the right side of the last equality above is zero as ϵ_{T+1} is independent of $\mathcal{F}_T$ as $\mathcal{F}_T$ involves only the information up to time T and replacing the parameters by their respective MLEs and the errors ϵ_t's by their respective residuals e_t's, we finally get our one-step-ahead forecast as

$$\hat{Y_T}(1) = \hat{\alpha} + \sum_{i=1}^{p} \hat{\beta}_i Y_{(T+1)-i} + \sum_{j=1}^{q} \hat{\theta}_j e_{(T+1)-j}.$$