

Stationarity, ADF, and AR Forecasting

S. Yanki Kalfa

Yanki.Kalfa@rad.ucsd.edu

University of California - San Diego

Rady School of Management

January 14, 2022

1 Introduction

This note provides a short summary of the following topics:

1. Integrated Processes
2. ADF Tests
3. AR Forecasts

Each of these topics are covered from a theoretical perspective. For an applied treatment of the topics please see my [GitHub](#) page.

2 Integrated Processes

2.1 Definitions

Before jumping directly to stationary processes and non-stationary processes, we need to define the characteristics of the series. The first moment, or the unconditional mean of a series is defined as below:

$$E[y_t] = \mu_t$$

μ_t is hardly ever observed in the data. The covariance between two element of the series is defined as:

$$E[(y_t - \mu_t) - (y_{t-j} - \mu_{t-j})] = \gamma_{j,t}$$

The covariance of the two elements displaced by j units is basically the autocovariance of order j . We want to have a series where the mean nor the autocovariance depend on time. Meaning that the expected value of the mean and the expected value of the autocovariances should not depend on time. If the series is time independent, then it is called covariance stationary, also called stationary. If the series is stationary then we have:

$$E[y_t] = \mu \quad (2.1)$$

$$E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j \quad (2.2)$$

$$\gamma_0 = E[(y_t - \mu)(y_t - \mu)] = E[(y_t - \mu)^2] = \text{var}(y_t) = \sigma^2 \quad (2.3)$$

In finite samples we have:

$$\hat{\gamma}_k = \frac{1}{T} \cdot \sum_{t=1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})$$

Another characteristic we need to take into consideration is the correlation between two elements of time series, also known as autocorrelation.

$$\rho_\tau = \frac{\gamma_\tau}{\text{var}(y)}$$

$$\hat{\rho}_\tau = \frac{1/T \cdot \sum_{t=1}^T (y_t - \bar{y})(y_{t-\tau} - \bar{y})}{1/T \cdot \sum_{t=1}^T (y_t - \bar{y})^2}$$

2.2 Stationary Processes

The first stationary process that we will look at is the white noise. It is the simplest stationary process. We define the white noise as the following:

$$y_t = \mu + \epsilon_t | \epsilon_t \sim N(0, \sigma^2) i.i.d.$$

$$E[y_t] = \mu + E[\epsilon_t] = \mu \tag{2.4}$$

$$E[(y_t - \mu)(y_t - \mu)] = E[(\epsilon_t)(\epsilon_t)] = \sigma^2 = \gamma_0 \tag{2.5}$$

$$E[(y_t - \mu)(y_{t-j} - \mu)] = E[(\epsilon_t)(\epsilon_{t-j})] = 0 = \gamma_j; j \geq 1 \tag{2.6}$$

We proved that the WN process is stationary because its mean and covariance are time independent.

One common stationary process is the Moving Average of order one (MA(1)). Defined below:

$$y_t = \mu + \epsilon_t + \theta \epsilon_{t-1} | \epsilon_t \sim N(0, \sigma^2) i.i.d.$$

Proof that MA(1) is stationary:

$$E[y_t] = E[\mu] + E[\epsilon_t] + E[\epsilon_{t-1}] = \mu \quad (2.7)$$

$$E[(y_t - \mu)^2] = \text{var}(\epsilon_t) + \text{var}(\theta\epsilon_{t-1}) = (1 + \theta)\sigma^2 \quad (2.8)$$

$$E[(y_t - \mu)(y_{t-1} - \mu)] = E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] = \theta\sigma^2 = \gamma_1 \quad (2.9)$$

$$\gamma_2 = 0 \quad (2.10)$$

It is important to note that shocks in an MA(1) process die out after 1 period.

An AR(1) is defined as shown below:

$$y_t = c + \phi y_{t-1} + \epsilon_t | \epsilon_t \sim N(0, \sigma^2) i.i.d.$$

We can express the AR(1) process with recursive substitution in the following form:

$$y_t = \sum_{i=0}^k \phi^i \cdot c + \phi^k \cdot y_{t-k} + \sum_{i=0}^k \phi^i \epsilon_{t-i}$$

$$E[y_t] = E[c] + \phi E[y_{t-1}] + E[\epsilon_t] = \frac{c}{1 - \phi} \quad (2.11)$$

$$E[(y_t - \mu)^2] = \text{var}(c) + \phi \text{var}(y_{t-1}) + \text{var}(\epsilon_t) = \frac{\sigma^2}{1 - \phi^2} \quad (2.12)$$

These prove that the mean and the variance of the AR(1) are time independent. Lastly, we need to prove that the autocovariance of the series is also time invariant. First let's consider the following:

$$(y_t - \mu) = c + \phi y_{t-1} + \epsilon_t - \frac{c}{1 - \phi}$$

$$(y_t - \mu) = \frac{c}{1 - \phi} - \frac{c}{1 - \phi} - (\phi \cdot \frac{c}{1 - \phi}) + \phi y_{t-1} + \epsilon_t$$

$$(y_t - \mu) = (y_{t-1} - \mu + \epsilon_t)$$

Keeping this in mind let's look at the autocovariance:

$$E[(y_t - \mu)(y_{t-1} - \mu)] = E[(y_{t-1} - \mu + \epsilon_t)(y_{t-1} - \mu)] \quad (2.13)$$

$$E[(y_{t-1} - \mu)(y_{t-1} - \mu)] + E[\epsilon_t(y_{t-1} - \mu)] = E[(y_{t-1} - \mu)(y_{t-1} - \mu)] \quad (2.14)$$

$$\gamma_1 = \phi \text{var}(y_t) = \phi \cdot \frac{\sigma^2}{1 - \phi^2} \quad (2.15)$$

$$\gamma_k = \phi^k \cdot \frac{\sigma^2}{1 - \phi^2} \quad (2.16)$$

Looking at the above proof, we can now say that the AR(1) process is stationary as long as the assumption $|\phi| < 1$ holds. Another important thing to know about the AR(1) process is its autocorrelation function. The first order autocorrelation of the AR(1) is simply:

$$\rho_1 = \frac{\gamma_1}{\gamma_0}$$

And the k^{th} autocorrelation is represented as:

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Of course, these only hold if the process is stationary.

2.3 Non-Stationary - Integrated Processes

When we think of non-stationary processes, we usually think of a random walk or a unit-root process. The difference between a stationary process and a non-stationary process is that the series/process is inherently unstable, meaning that the mean and variance are unbounded. Let $\phi = 1$ then:

$$\begin{aligned} E[y_t] &= E[c] + \phi E[y_{t-1}] + E[\epsilon_t] = \frac{c}{1 - \phi} = \infty \\ E[(y_t - \mu)^2] &= var(c) + \phi var(y_{t-1}) + var(\epsilon_t) = \frac{\sigma^2}{1 - \phi^2} = \infty \\ \gamma_1 &= \phi var(y_t) = \phi \cdot \frac{\sigma^2}{1 - \phi^2} = \infty \end{aligned}$$

This means that these values are time dependent and unbounded. If the variance and covariance of a process, its statistics will be unreliable. But this does not mean that we cannot work with a unit root process. If we can get the process to be stationary by first differencing then we say that the process is integrated of order 1 or I(1). We can show this in the following form:

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \epsilon_t \\ y_t - y_{t-1} &= c + \phi y_{t-1} - y_{t-1} + \epsilon_t \\ \Delta y_t &= c + (1 - \phi)y_{t-1} + \epsilon_t \end{aligned}$$

Now we can run tests on the $(1 - \phi)$ coefficient, and this is called an I(1).

3 Augmented Dickey-Fuller Tests

The purpose of conducting a Dickey-Fuller (DF) test or an Augmented Dickey Fuller (ADF) test is to see if the series we are interested in has a unit-root. The ADF test transforms a regression in levels to a regression in changes. I will demonstrate it in the sections below.

3.1 Dickey Fuller Test

Let's take a simple AR(1):

$$x_t = \phi \cdot x_{t-1} + \epsilon_t$$

We can transform this to obtain the DF regression:

$$x_t = \phi \cdot x_{t-1} + \epsilon_t$$

$$x_t - x_{t-1} = \phi \cdot x_{t-1} - x_{t-1} + \epsilon_t$$

$$\Delta x_t = (\phi - 1) \cdot x_{t-1} + \epsilon_t$$

we can call $(\phi - 1) = \pi$.

$$\Delta x_t = \pi \cdot x_{t-1} + \epsilon_t$$

The DF test's null hypothesis is:

$$H_0 : \pi = 0$$

This is counter-intuitive, however, it is important to keep in mind that $\pi = (\phi - 1)$ hence testing $\pi = 0$ tests if $\phi = 1$.

There is one issue with the DF test, it ignores serial correlation. Therefore, we usually run an Augmented Dickey Fuller test which includes differences of different orders.

3.2 Augmented Dickey Fuller (ADF) Test

Consider an AR(4):

$$x_t = \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} + \phi_3 \cdot x_{t-3} + \phi_4 \cdot x_{t-4}$$

We can transform this regression into an ADF(3) and obtain:

$$\Delta x_t = \pi \cdot x_{t-1} + \gamma_1 \cdot \Delta x_{t-1} + \gamma_2 \cdot x_{t-2} + \gamma_3 \cdot x_{t-3}$$

where: $\pi = \phi_1 + \phi_2 + \phi_3 + \phi_4 - 1$

By including the lags of the differences of the series x_t , the ADF solves the problem of serial correlation. The test has two different components, the first part is the lag selection and the second part is the unit root test on π .

Analytically, we can run the regression above by dropping the lag length one by one and look at the joint significance of the dropped lags. You can implement this by running these regressions and running a progress report, which will give you the likelihood ratio test for each model, and will also present AIC, BIC, and HQ information criterion. We select the smallest when looking at these information criteria. In practice, we can just run a unit root test in Python/R and select 3 lags. The unit root test includes AIC for lag selection. After looking at lag selection, we can look at unit-root tests. The null hypothesis is: $H_0 : \pi = 0$, if we fail to reject the null, then we say that the process is unit root, and is at least integrated of order 1 ($I(1)$).

4 AR Forecasting

We will start with a simple AR(1) process to showcase that as the forecast horizon grows the forecast itself will converge to the mean of the series. To make the demonstration simpler we will first subtract the mean from y and then forecast.

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t$$

suppose now we want to forecast 3 steps ahead. We will iterate our equation forward

$$(\widehat{y_{t+1}} - \mu) = \phi[\phi(y_{t-1} - \mu) + \varepsilon_t] + \varepsilon_{t+1}$$

$$= \phi^2(y_t - \mu) + \phi\varepsilon_t + \varepsilon_{t+1}$$

$$(\widehat{y_{t+2}} - \mu) = \phi^3(y_t - \mu) + \phi^2\varepsilon_t + \phi\varepsilon_{t+1} + \varepsilon_{t+2}$$

$$\widehat{y_{t+3}|t} = \mu + \phi^3(y_t - \mu) + \phi^2\varepsilon_t + \phi\varepsilon_{t+1} + \varepsilon_{t+2}$$

$$E[y_{t+3}|y_t, y_{t-1}, \dots] = \mu + \phi^3 E[y_t - \mu]$$

Now suppose that we roll this forward s times to get an s step-ahead forecast.

$$E[y_{t+s}|y_t, y_{t-1}, \dots] = \mu + \phi^s E[y_t - \mu]$$

$$\lim_{s \rightarrow \infty} E[y_{t+s}|y_t, y_{t-1}, \dots] = \mu$$

Hence we can see that the forecast is converging to the mean of the process.