

# Time Series Fundamentals

Jakob Sverre Alexandersen  
GRA4159 Trends, Cycles & Signals Extraction  
Lecture 2

January 14, 2026

## Contents

<b>1</b>	<b>What is a time series?</b>	<b>2</b>
1.1	What is the cycle? . . . . .	2
1.2	Our focus . . . . .	3
1.3	Defining time series . . . . .	3
<b>2</b>	<b>Tools, definitions, and concepts</b>	<b>4</b>
2.1	Lag operators, diff.eq and the autocorrelation function . . . . .	4
2.2	White noise and moving averages . . . . .	4
2.3	Difference equations . . . . .	4
2.4	Lag operators . . . . .	5
2.5	Conditional and unconditional expectations . . . . .	5
2.6	Stationarity . . . . .	7
2.7	Autocorrelation function (ACF) . . . . .	7
2.8	Ergodicity . . . . .	7
<b>3</b>	<b>Some basic time series processes</b>	<b>8</b>
3.1	Moving average process (MA) . . . . .	8
3.1.1	ACF for an MA process . . . . .	9
3.1.2	Summary thus far . . . . .	9
3.2	Autoregressive process (AR) . . . . .	9
3.2.1	AR(1) to MA( $\infty$ ) by recursive substitution . . . . .	10
3.2.2	AR(1) to MA( $\infty$ ) with lag operators . . . . .	10
3.2.3	Mean, variance and autocovariance of an AR(1) . . . . .	11
3.2.4	Adding a constant . . . . .	11
3.2.5	Higher order AR processes . . . . .	12
3.2.6	Autoregressive moving average (ARMA) process . . . . .	13
3.2.7	Estimation . . . . .	13
3.2.8	Lag selection . . . . .	13
3.2.9	Multipliers and impulse responses . . . . .	14
3.3	Random walk (RW) . . . . .	14

# 1 What is a time series?

- Most data in macroeconomics and finance can be described as time series: a set of repeated observations over time of the same variable, such as consumer prices, GDP, stock prices, exchange rates, etc.
- Accordingly, to be able to understand macroeconomic fluctuations and financial markets, we need to know time series
- In terms of the notation, we will from now specify  $y_t$ , where the subscript  $t$  runs over time, i.e.  $t = 1, 2, 3, \dots, T$ .
- Time series data can have a wide range of frequencies. This has implications for what type of analysis we can do on the data
- The sample available for a given time series varies greatly. Many results in statistics and econometrics depend on having many observations
- Time series data behave very differently. The logarithm of a given time series  $y_t$  can be thought of as the sum of four (additive) components:

$$y_t = g_t + c_t + s_t + \epsilon_t$$

where  $g_t$  is a trend,  $c_t$  is a cycle,  $s_t$  is a seasonal component, and  $\epsilon_t$  is noise

Trend:

- Either deterministic or stochastic trends

Seasonal component:

- Deterministic wave movements
- In economics, think e.g. Christmas shopping, holidays, weekends etc.

Noise:

- Unpredictable/unanticipated innovation/shock

## 1.1 What is the cycle?

Cycle:

- Burns and Mitchell (1943):
  - expansions and recessions occurring at about the same time in many economic activities
  - sequence of changes is **recurrent but not periodic**
  - duration of business cycles vary from more than one year to ten or twelve years
- Lucas (1977):
  - movements about the trend in GNP
  - do **not** resemble deterministic wave motions
  - cycles characterized by co-movement among different aggregative time series
  - they are characterized by qualitative similarity and not by e.g. country or time period

## 1.2 Our focus

- in this course we will focus mostly on the cyclical part of time series (+ noise), less on trends, and hardly nothing on the seasonal component
  - most of the time we are not that interested in the seasonal component. However, estimating and removing the seasonal component from time series can be demanding and complicated and worth a course itself
  - luckily, the statistical agencies typically do the job for us so that we can download seasonally adjusted NAS data directly from their databases
- indeed, the theoretical model you will learn about in the following weeks will focus on business cycles, i.e., the cyclical part of time series. But, most of the time we need to estimate the trend to get the cycle. I.e.:
  - $\hat{c}_t = y_t + \hat{g}_t + \hat{\epsilon}_t$  (when working with seasonally adjusted data)
  - note also that in macro, depending on what variable  $y_t$  is,  $c_t$  is often referred to as the output gap (for GDP), core inflation (for inflation), credit gap (for credit), etc.
  - more on this later!

## 1.3 Defining time series

Defined formally, a time series is a collection of observations indexed by date of each observation, say starting in time  $t = 1$  and ending in time  $t = T$ :

$$\{y_1, y_2, y_3, \dots, y_T\}$$

where the time index can be of any frequency, e.g., daily or quarterly. Usually this (finite) sample is a subset of an infinite sample, indexed by  $\{y_t\}_{t=-\infty}^{\infty}$ . A time series is usually identified by describing the  $t$ -th element

We will generally treat  $y_t$  as a random variable, implying that a time series is a sequence of random variables ordered in time. We call such a sequence a stochastic process.

But, non-stochastic time series are also easy to envision (e.g. deterministic processes), i.e.

$$y_t = t$$

$$y_t = c$$

## 2 Tools, definitions, and concepts

### 2.1 Lag operators, diff.eq and the autocorrelation function

To introduce **randomness** into a time series process, we work with white noise.

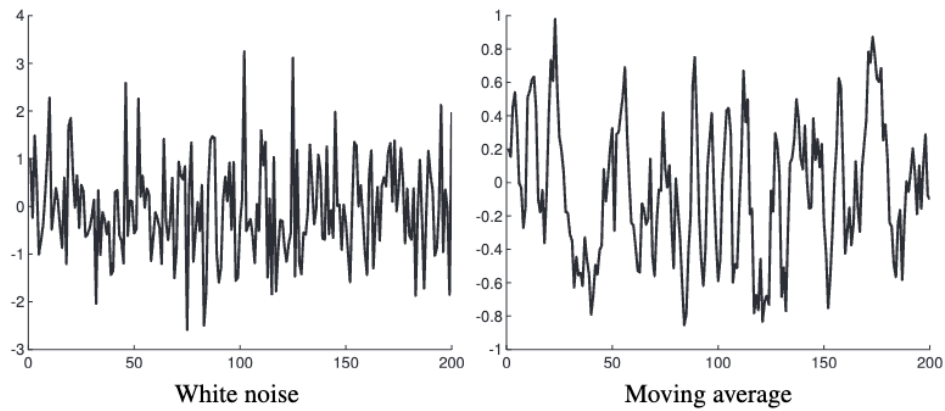
$$y_t = \epsilon_t$$

where  $\{\epsilon_t\}_{t=-\infty}^{\infty}$  is a sequence of independent and random variables each of which has a distribution:

$$\epsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$$

That is,  $\epsilon_t$  has zero mean and constant variance  $\sigma^2$

### 2.2 White noise and moving averages



Moving average is here simply computed as  $y_t = \frac{1}{5}(\epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t + \epsilon_{t+1} + \epsilon_{t+2})$  which we call a centered moving average

### 2.3 Difference equations

To introduce **dynamics** we work with difference equations

$$y_t = \phi y_{t-1} + \epsilon_t$$

It is customary to assume that  $\epsilon_t$  is white noise. If  $\phi < 1$ , we can show that the series will always return to its mean, i.e., it is (covariance) stationary

If  $\phi = 1$  we get a special, and important, case, called a random walk:

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

more on this later

## 2.4 Lag operators

To find solutions to various difference equations, it is helpful to employ lag operators,  $L$ , which transform an observation at time  $t$  to period  $t - 1$ , that is, backward one period in time:

$$Ly_t = y_{t-1}$$

Raising the lag operator to the power of  $-1$  transforms the series one period forward, i.e.,  $L^{-1}y_t = y_{t+1}$ . Generally, the lag operator can be raised to arbitrary integer powers  $k$  s.t.:

$$L^k y_t = y_{t-k} \quad ; \quad L^{-k} y_t = y_{t+k}$$

A convenient use of the lag operator is to express the first difference of a series as:

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$$

A polynomial of lag operators is called a lag polynomial, defined as  $\phi(L)$ , where:

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$$

where  $p$  is the lag order

## 2.5 Conditional and unconditional expectations

### Expectations:

We refer to the first moments of a stochastic process  $y_t$  as the mean:

$$\mu = \mathbb{E}[y_t] \quad t = 1, \dots, T$$

The first moments are interpreted as the average value of  $y_t$  taken over all possible realizations. Note that in this chapter we will not allow the mean of a time series to be time dependent, i.e., vary over time

The second moments are defined as the variance:

$$\text{var}(y_t) = \mathbb{E}[y_t, y_t] = \mathbb{E}[(y_t - \mathbb{E}[y_t])^2] \quad t = 1, \dots, T$$

and the covariance, for lag  $j$ :

$$\text{cov}(y_t, y_{t-j}) = \mathbb{E}[y_t, y_{t-j}] = \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-j} - \mathbb{E}[y_{t-j}])] \quad t = j + 1, \dots, T$$

### Conditional moments:

The conditional distribution is based on the observation of some realization of random variables. Consider the following first-order difference equation:

$$y_t = \phi y_{t-1} + \epsilon_t$$

where  $\epsilon \sim \text{i.i.d. } N(0, \sigma^2)$  and  $|\phi| < 1$ . Since the error terms are from the normal distribution, we know that  $y_t$  will also be a normally distributed variable. Conditional on information up to time  $t$ , i.e. knowing  $y_{t-1}$ , the mean, variance and covariance of the process in the diff.eq are:

$$\begin{aligned} \mathbb{E}[y_t | y_{t-1}] &= \phi y_{t-1} \\ \text{var}(y_t | y_{t-1}) &= \sigma^2 \\ \text{cov}((y_t | y_{t-1}), (y_{t-j} | y_{t-j-1})) &= 0 \quad \text{for } j > 1 \end{aligned}$$

This follows as  $\epsilon_t$  is equal to 0 by construction, further that  $\text{var}(y_t | y_{t-1}) = \mathbb{E}[\phi y_{t-1} + \epsilon_t - \phi y_{t-1}]^2 = \mathbb{E}[\epsilon_t]^2 = \sigma^2$  and by employing the standard covariance formula. Hence, conditional on  $y_{t-1}$ , the distribution of  $y_t$  does not depend on previous values of  $y_{t-j}$ , for  $j > 1$ .

What happens if we instead condition on  $y_{t-2}$ ?

$$\begin{aligned}\mathbb{E}[y_t|y_{t-2}] &= \phi^2 y_{t-2} \\ \text{var}(y_t|y_{t-2}) &= (1 + \phi^2) \sigma^2 \\ \text{cov}((y_t|y_{t-2}), (y_{t-j}|y_{t-j-2})) &= \phi \sigma^2 \quad \text{for } j = 1 \\ \text{cov}((y_t|y_{t-2}), (y_{t-j}|y_{t-j-2})) &= 0 \quad \text{for } j > 1\end{aligned}$$

### Unconditional moments

Unconditional moments are the same as the populations moments, but explicitly specified in a time series context. The unconditional distribution of a time series process is arrived at under the hypothesis that we are at the observation preceding the realized observation. Thus, we only know the process generating the observation

The unconditional moments of the process in  $y_t = \phi y_{t-1} + \epsilon_t$  can be shown to be:

$$\begin{aligned}\mathbb{E}[y_t] &= 0 \\ \text{var}(y_t) &= \frac{\sigma^2}{1 - \phi^2} \\ \text{cov}(y_t, y_{t-j}) &= \phi^j \text{var}(y_t)\end{aligned}$$

Hence, in the conditional moments, the location of the mean depends on the conditioning information set, i.e.,  $y_{t-1}$ . The unconditional mean in these aforementioned moments, on the other hand, is zero as we assume it only depends on the parameter  $\phi$ . Also, knowing  $y_{t-1}$  changes the size of the variance of  $y_t$  and the degree of covariance with lagged values of  $y_t$  relative to the unconditional case.

### What's the challenge?

In reality we only ever observe one realization of a stochastic process, i.e. a sequence of  $T$  random variables. However, this one realization is only one of infinitely many possible “paths of histories / alternative worlds”

We require time series properties that ensure that the amount of information contained in observing one realization over time is equivalent to observing the process over and over again at the same point in time.

As we will see, the necessary properties are:

- stationarity
- ergodicity

## 2.6 Stationarity

- stationarity is a fundamental concept in time series analysis
- A time series is **strictly stationary** if for any avalues of  $j_1, j_2, \dots, j_n$ , the joint distribution of  $(y_t, y_{t+j_1}, y_{t+j_2}, \dots, y_{t+j_n})$  depends only on the intervals separating the dates  $(j_1, \dots, j_n)$  and not on the date itself  $(t)$ .
- If neither the mean  $(\mu)$  nor the covariances  $\text{cov}(y_t, y_{t-j})$  depend on the date  $t$ , then the process for  $y_t$  is said to be **covariance (weakly) stationary**:

$$\begin{aligned}\mathbb{E}[y_t] &= \mu \quad \forall j \\ \mathbb{E}[(y_t - \mu)(y_{t-j} - \mu)] &= \text{cov}(y_t, y_{t-j}) \quad \forall t, j\end{aligned}$$

- strong stationarity implies weak stationarity but not vice versa. Special case: gaussian distribution

## 2.7 Autocorrelation function (ACF)

When a process is stationary, its time domain properties can be summarized by computing the covariance of the process against a given number of lags. This is called the autocovariance function, and denoted  $\gamma(j) \equiv \text{cov}(y_t, y_{t-j})$  for  $t = 1, \dots, T$  against  $j$ .

The autocovariance may be standardized by dividing by the variance of the process. This yields the autocorrelation function (ACF):  $\rho(j) \equiv \frac{\gamma(j)}{\gamma(0)}$ . Often it is useful to make a plot of  $\rho(j)$  against (non-negative)  $j$  to learn about the properties of a given time series process. Note that  $\rho(0) = 1$  by definition, as  $\rho(0) \equiv \frac{\gamma(0)}{\gamma(0)} = 1$

## 2.8 Ergodicity

A covariance-stationary process is said to be ergodic of the mean if the (sample) average  $\bar{y} \equiv (1/T) \sum_{t=1}^T y_t$  converges in probability to  $\mathbb{E}[y_t]$  as  $T \rightarrow \infty$ .

If ergodicity holds, it implies that the sample average and variance provide consistent estimates of their populations' counterparts (Ergodic theorem).

One implication of ergodicity is that the autocorrelation function goes to zero quickly as  $j$  becomes large (observations sufficiently apart should be almost uncorrelated):

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty$$

That is, the dependencies between  $y_t$  and  $y_{t-j}$  weaken with increasing  $j$ . A stationary process is ergodic, if it is asymptotically independent, i.e. if any two rv's positioned far apart in the sequence are almost independently distributed.

An example of this may be daily temperatures with site-specific oddset. Here one sensor's average  $B$  will be different from all the sites' averages.

### 3 Some basic time series processes

#### 3.1 Moving average process (MA)

We will now use the fundamentals discussed so far to construct and analyze the properties of time series models that make linear combinations of  $\epsilon_t$  in different ways.

The first process we will consider in detailed is the moving average (MA), which is a linear combination of white noise. The simplest case is the MA(1):

$$y_t = \delta + \epsilon_t + \theta\epsilon_{t-1}$$

where  $\delta$  is a constant,  $y_t$  is the weighted sum of the two most recent values of  $\epsilon$ , and  $\epsilon_t$  is i.i.d. white noise,  $\epsilon_t \sim N(0, \sigma^2)$ .

##### Is the MA stationary?

To examine this we need to calculate the different moments. We start with the mean, i.e., the first moment of the process:

$$\begin{aligned}\mathbb{E}[y_t] &= \mathbb{E}[\delta + \epsilon_t + \theta\epsilon_{t-1}] \\ &= \delta + \mathbb{E}[\epsilon_t] + \theta\mathbb{E}[\epsilon_{t-1}] \\ &= \delta\end{aligned}\tag{1}$$

Since the  $\epsilon$ 's are i.i.d., their expectations are zero. Hence, the mean is found to be  $\delta$ , which is constant

Turning to the second moments, we first compute the variance as:

$$\begin{aligned}\text{var}(y_t) &= \mathbb{E}[y_t - \mathbb{E}[y_t]]^2 \\ &= \mathbb{E}[(\delta + \epsilon_t + \theta\epsilon_{t-1}) - \delta]^2 \\ &= \mathbb{E}[\epsilon_t]^2 + 2\theta\mathbb{E}[\epsilon_t\epsilon_{t-1}] + \mathbb{E}[\theta\epsilon_{t-1}]^2 \\ &= \sigma^2 + 0 + \theta^2\sigma^2 \\ &= (1 + \theta^2)\sigma^2\end{aligned}\tag{2}$$

the first covariance as:

$$\begin{aligned}\text{cov}(y_t, y_{t-1}) &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-1} - \mathbb{E}[y_{t-1}])] \\ &= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] \\ &= \mathbb{E}[\epsilon_t\epsilon_{t-1}] + \mathbb{E}[\theta\epsilon_{t-1}^2] + \mathbb{E}[\theta^2\epsilon_{t-1}\epsilon_{t-2}] \\ &= 0 + \theta\sigma^2 + 0 + 0 \\ &= \theta\sigma^2\end{aligned}\tag{3}$$

and for lag order  $j > 1$ :

$$\begin{aligned}\text{cov}(y_t, y_{t-j}) &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-j} - \mathbb{E}[y_{t-j}])] \\ &= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-j} + \theta\epsilon_{t-j-1})] \\ &= 0 \quad \text{for } j > 1\end{aligned}\tag{4}$$

Neither the mean (1), the variance (2) nor the covariances (3 and 4) depend on time, the MA(1) process is covariance stationary. Moreover, the MA(1) process is stationary regardless of the value of  $\theta$



### 3.1.1 ACF for an MA process

What does the ACF for an MA(1) process look like? Recall that the ACF is given by  $\frac{\gamma(j)}{\gamma(0)}$ . Accordingly, the autocorrelation for the MA(1) is:

$$\begin{aligned}\rho(1) &= \frac{\theta}{(1 + \theta^2)} \\ \rho(j) &= 0 \quad \text{for } j > 1\end{aligned}$$

We see that for lag orders  $j > 1$ , the autocorrelations are zero. Hence, the autocorrelation function goes to zero as  $j$  becomes large, i.e., the MA(1) process is ergodic.

### 3.1.2 Summary thus far

1. The noise component is most often assumed to be Gaussian white noise
2. Linear combinations of Gaussian white noise generate time series that behave similarly to many real world time series
3. A time series is said to be covariance-stationary if neither its first nor second moments depend on the time index
4. When a process is stationary, its time domain properties can be summarized by the autocorrelation function (ACF)
5. For a stationary process, the ACF goes to zero quickly as the number of lags increases
6. The MA model is stationary regardless of its parameters

## 3.2 Autoregressive process (AR)

Another important time series process is the AR process:

$$y_t = \phi y_{t-1} + \epsilon_t$$

The AR(1) relates the value of a variable  $y$  at time  $t$ , to its previous value at time  $(t - 1)$  and a random disturbance  $\epsilon$ , also at time  $t$ . As always,  $\epsilon_t$  is i.i.d. white noise,  $\epsilon_t \sim N(0, \sigma^2)$ .

If  $|\phi| < 1$ , the AR(1) is covariance-stationary, with finite variance, i.e.:

$$\begin{aligned}\mathbb{E}[y_t] &= 0 \\ \text{var}(y_t) &= \frac{\sigma^2}{1 - \phi} \\ \text{cov}(y_t, y_{t-1}) &= \phi \text{var}(y_t)\end{aligned}$$

We will now prove this. There are two ways to do so, either through recursive substitution, or by using the lag operator

### 3.2.1 AR(1) to MA( $\infty$ ) by recursive substitution

We can solve  $y_t = \phi y_{t-1} + \epsilon_t$  with recursive substitution (starting at some infinite time  $j$ ).

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t \\ &= \phi(\phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &\quad \Downarrow \\ &= \phi^{j+1} y_{t-(j+1)} + \phi^j \epsilon_{t-j} + \dots + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \end{aligned}$$

The final line explains  $y$  as a linear function of an initial value  $y_{t-j-1}$  and the historical values of  $\epsilon_t$ . If  $|\phi| < 1$  and  $j$  become large,  $\phi^{j+1} y_{t-(j+1)} \rightarrow 0$ . Thus, the AR(1) can be expressed as an MA( $\infty$ ).

$$y_t = \phi^j \epsilon_{t-j} + \dots + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

### 3.2.2 AR(1) to MA( $\infty$ ) with lag operators

Using the lag operator,  $y_t = \phi y_{t-1} + \epsilon_t$  can be written as:

$$(1 - \phi L)y_t = \epsilon_t$$

A sequence  $\{y_t\}_{t=-\infty}^{\infty}$  is said to be bounded if there exists a finite number  $k$ , s.t.  $|y_t| < k \quad \forall t$ . Provided  $|\phi| < 1$  and we restrict ourselves to bounded sequences, the expression can be multiplied by  $(1 - \theta L)^{-1}$  on both sides of the equality sign to obtain:

$$y_t = (1 - \theta L)^{-1} \epsilon_t$$

and solved s.t.:

$$y_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

Importantly, we can only perform this inversion if  $|\phi| < 1$ . However, when it holds, we say that the process is invertible.

How does this work? When  $|\phi| < 1$ , and  $t \rightarrow \infty$ , the following approximation holds:

$$(1 - \phi L)^{-1} = \lim_{j \rightarrow \infty} (1 + \phi L + (\phi L)^2 + \dots + (\phi L)^j)$$

which is based on the expansion  $(1 - z)^{-1} = 1 + z + z^2 + \dots$ , i.e. the geometric rule. Thus:

$$y_t = (1 - \theta L)^{-1} \epsilon_t$$

can be written as:

$$y_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

### 3.2.3 Mean, variance and autocovariance of an AR(1)

The mean is computed as:

$$\mathbb{E}[y_t] = \mathbb{E}[\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots] = 0$$

The variance is:

$$\begin{aligned}\gamma(0) &= \text{var}(y_t) = \mathbb{E}[y_t - \mathbb{E}[y_t]]^2 \\ &= \frac{1}{1 - \phi^2} \sigma^2\end{aligned}$$

The first order covariance is:

$$\begin{aligned}\gamma(1) &= \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-1} - \mathbb{E}[y_{t-1}])] \\ &= \phi \frac{1}{1 - \phi^2} \sigma^2 = \phi \text{var}(y_t)\end{aligned}$$

while for  $j > 1$  we have:

$$\gamma(j) = \mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-j} - \mathbb{E}[y_{t-j}])] = \phi^j \text{var}(y_t)$$

### 3.2.4 Adding a constant

Let us now add a constant to the AR(1) model, so that we can write:

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

We can easily prove that the results will still be the same, except for the mean, that is no longer zero. To see this define  $v_t = \mu + \epsilon_t$ , then:

$$\begin{aligned}y_t &= \phi y_{t-1} + v_t \\ y_t &= (1 - \phi L)^{-1} v_t \\ y_t &= \left( \frac{1}{1 - \phi} \right) \mu + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots\end{aligned}$$

Accordingly, the unconditional first moment is:

$$\mathbb{E}[y_t] = \left( \frac{1}{1 - \phi} \right) \mu$$

which is no longer zero

### 3.2.5 Higher order AR processes

The results for an MA(1) process generalize to an infinite MA. For higher order AR processes, things become a bit more complicated.

Assume an AR(2) process:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$$

Whether this process is stationary no longer hinges on  $\phi_1$  alone, as we now also have to consider  $\phi_2$ . As it turns out, there are different ways of doing this.

First, rewrite and define the  $(2 \times 1)$  vector  $Z_t$ :

$$Z_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$$

the  $(2 \times 1)$  vector  $v_t$  by:

$$v_t = \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}$$

and the  $(2 \times 2)$  matrix  $\Gamma$  by:

$$\Gamma = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$$

Then the following first-order vector difference equation can be written as

$$Z_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix} \quad (5)$$

or

$$Z_t = \Gamma Z_{t-1} + v_t$$

$\Gamma$  is called the companion form matrix. Any higher AR process can be represented as a first-order AR process using this formulation and generalizations of it. We can check stationarity by computing the eigenvalues of matrix  $\Gamma$ . That is, the eigenvalues of  $\Gamma$  are those numbers  $\lambda$  for which  $|\Gamma - \lambda I| = 0$ . Substituting in for (5), the eigenvalues are the solution to:

$$\begin{aligned} \left| \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| &= 0 \\ \left| \begin{bmatrix} (\phi_1 - \lambda) & \phi_2 \\ 1 & -\lambda \end{bmatrix} \right| &= \lambda^2 - \phi_1 \lambda - \phi_2 = 0 \end{aligned}$$

These eigenvalues can be found from the formula:

$$\lambda_1, \lambda_2 = \frac{(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2})}{2}$$

Stationarity requires that the eigenvalues are less than one in absolute value. In the AR(2) case, one can show that this will be the case if:

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ -\phi_1 + \phi_2 &> -1 \\ \phi_2 &> -1 \end{aligned}$$

### 3.2.6 Autoregressive moving average (ARMA) process

By combining the MA and AR processes, we get an ARMA process. In the simplest case, we can specify an ARMA(1, 1) process, which would equal:

$$y_t = \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

Using the lag polynomial, a more general formulation of an ARMA model is:

$$\phi(L)y_t = \theta(L)\epsilon_t$$

where

$$\theta(L) = 1 - \sum_{i=1}^p \theta_i L^i \quad \text{and} \quad \theta(L) = 1 + \sum_{i=1}^q \theta_i L^i$$

The number of lags,  $p$  and  $q$ , can differ. For instance, ARMA(2, 1) combines an AR(2) with an MA(1):

$$\begin{aligned} (1 - \phi_1 L - \phi_2 L^2) y_t &= (1 + \theta_1 L) \epsilon_t \\ y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} \end{aligned}$$

Whether an ARMA process is stationary depends solely on its autoregressive part.

### 3.2.7 Estimation

An AR( $p$ ) model can be estimated using the OLS estimator. An estimate of the population parameters  $\phi_p$ , for  $p = 1, \dots, P$  and  $\sigma^2$  can be found by using the OLS quantities  $\hat{\theta}_p$  and  $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t$ , where  $\hat{\epsilon}_t = y_t - \sum_{p=1}^P \hat{\theta}_p y_{t-p}$ . But note, this is a valid procedure only when the process is stationary. Recall the OLS assumption to produce unbiased estimates:

$$\mathbb{E}[\epsilon_t | y_1, \dots, y_T] = 0 \quad \forall t = 1, \dots, T$$

In an autoregressive model like

$$y_t = \alpha + \theta_1 y_{t-1} + \epsilon_t$$

this does not hold. Clearly, if  $\theta_1 \neq 0$ ,  $\mathbb{E}[\epsilon_t | y_{t+1}] \neq 0$ , and the relationship breaks down. Then the OLS estimator of  $\theta_1$  will be biased.

Using large sample theory it can be shown that valid OLS estimation and inference can be conducted when modeling stationary time series.

### 3.2.8 Lag selection

Alternative methods for determining the lag length are often based on minimizing an information criterion. Two popular information criterion functions are Bayes Information Criterion (BIC) and the Akaike (AIC) information criterion:

$$\text{BIC}(p) = \ln \left( \frac{\text{SSR}(p)}{T} \right) + (p+1) \frac{\ln(T)}{T}$$

and

$$\text{AIC}(p) = \ln \left( \frac{\text{SSR}(p)}{T} \right) + (p+1) \frac{2}{T}$$

where  $\text{SSR}(p) = \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t$ . The first term will typically decrease as  $p$  increases, while the second term increases as the model grows. Difference is in the last term: BIC is more conservative, ‘penalizes’ the size of the model more.

In practice, neither the BIC nor AIC (or a regular  $t$ -statistic), will likely provide us with one answer. Therefore, depending on the application and the question being asked, the criterion we use will differ.

### 3.2.9 Multipliers and impulse responses

In time series, we often want to study causes (i.e., an unexpected movement in a given variable or a shock) and the effect (the response to the shock). Assume now an autoregressive AR(1) process, that can be written as infinite moving average equation:

$$y_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

Thus,  $y_t$  can be described entirely by its past and present errors or shocks. If we assume the dynamic simulation started at time  $j$ , (taking  $y_{t-(j+1)}$  as given), the effect of a change in the initial shock (assuming the rest remains the same) on  $y_t$  is:

$$\frac{\partial y_t}{\partial \epsilon_{t-j}} = \phi^j$$

which we call the dynamic multiplier. It depends only on  $j$ , the length of time separating the disturbance to the input ( $\epsilon_{t-j}$ ) and the observed value of output ( $y_t$ ). It does not depend on time

The cumulative effect of this temporary shock is:

$$\sum_{j=0}^{\infty} \frac{\partial y_t}{\partial \epsilon_{t-j}} = 1 + \phi + \phi^2 + \dots + \phi^j = \frac{1}{1 - \phi}$$

When  $|\phi| < 1$  the process decays geometrically towards zero, (positive coefficients – smooth decay; negative coefficients – decays with alternating signs). We say that a system described in this way is stable.

The dynamic multipliers can easily be moved forward in time s.t.:

$$\frac{\partial y_{t+1}}{\partial \epsilon_t} = \phi^j$$

Stated like this, the dynamix multipliers for  $j = 1, \dots, J$  are often referred to as the impulse response function. Impulse responses are important tools that enables us to study the effects of structural shocks over time

## 3.3 Random walk (RW)