# Lecture 8-a
# Time Series:
# Stationarity, AR(p) & MA(q)

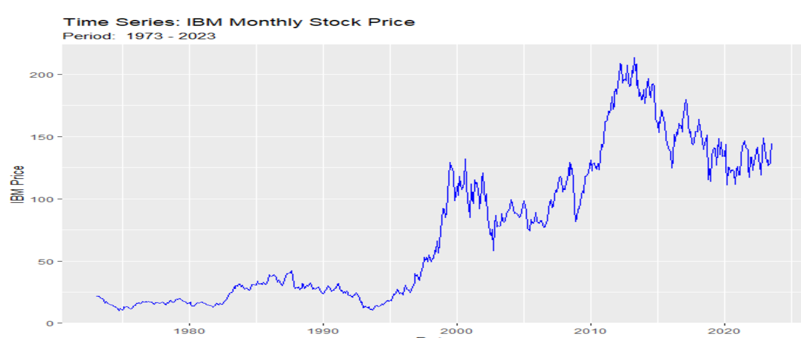
Brooks (4th edition): Chapter 6

 1

## Time Series: Introduction

• A time series $y_t$ is a process observed in sequence over time,
$t$ = 1, ...., $T$ $\Rightarrow Y_t = \{y_1, y_2, y_3, ..., y_T\}$.

**Examples**: IBM monthly stock prices from 1973:January till 2020:September (plot below); or USD/GBP daily exchange rates from February 15, 1923 to March 19, 1938.



## Time Series: Introduction

**Examples (continuation)**: Different ways to do the plot in R:

- Using plot.ts, creating a timeseries object in R:

```
# the function ts creates a timeseries object, start = 1973,1 (start of sample), frequency =
12(=monthly)
ts_ibm <- ts(x_ibm, start=c(1973,1), frequency=12)

plot.ts(ts_ibm,xlab="Time",ylab="IBM price", main="Time Series: IBM Stock Price")
```

- Using R package ggplot2

```
x_ibm <- SFX_da$IBM
x_date <- as.Date(SFX_da$Date, "%m/%d/%Y")
df <- data.frame(x_date, x_ibm)
ggplot(df, aes(x = x_date, y = x_ibm)) +
  geom_line(color="blue") +
  labs(x = "Date", y = "IBM Price", col = "blue", title = "Time Series: IBM Monthly Price",
      subtitle = "Period:  1973 - 2023")
```

## Time Series: Introduction – Categories

- Usually, time series models are separated into two categories:
  – Univariate ($y_t \in R$, it is a scalar)

    **Example:** We are interested in the behavior of IBM stock prices as function of its past.

    ⇒ Primary model: Autoregressions (ARs).

  – Multivariate ($y_t \in R^m$, it is a vector-valued)

    **Example:** We are interested in the joint behavior of IBM stock and IBM bond prices as function of their joint past.

    ⇒ Primary model: Vector autoregressions (VARs).

## Time Series: Introduction – Dependence

• Given the sequential nature of $y_t$, we expect $y_t$ & $y_{t-1}$ to be dependent. This is the main feature of time series: dependence. It creates statistical problems.

• In classical statistics, we usually assume we observe several *i.i.d.* realizations of $y_t$. We use $\bar{y}$ to estimate the mean.

• With several independent realizations we are able to sample over the entire probability space and obtain a "good" –i.e., consistent or close to the population mean– estimator of the mean.

• But, if the samples are highly dependent, then it is likely that $y_t$ is concentrated over a small part of the probability space. Then, the sample mean will not converge to the mean as the sample size grows.

## Time Series: Introduction – Dependence

Technical note: With dependent observations, the classical results (based on LLN & CLT) are not to valid. New assumptions and tools are needed: *stationarity*, *ergodicity*, CLT for martingale difference sequences (MDS CLT).

Roughly speaking, stationarity requires constant moments for $y_t$; ergodicity requires that the dependence is short-lived, eventually $y_t$ has only a small influence on $y_{t+k}$, when $k$ is relatively large.

• The amount of dependence in $y_t$ determines the 'quality' of the estimator. There are several ways to measure the dependence. The most common measure: Covariance.

$$\text{Cov}(y_t, y_{t+k}) = E[(y_{t_t} - \mu)(y_{t+k} - \mu)]$$

Note: When μ =0, then $\text{Cov}(y_t, y_{t+k}) = E[y_t\, y_{t+k}]$

## Time Series: Introduction – Forecasting

• In a time series model, we describe how $y_t$ depends on past $y_t$'s. That is, the information set is $I_t = \{y_{t-1}, y_{t-2}, y_{t-3}, ....\}$

• The purpose of building a time series model: Forecasting.

• We estimate time series models to forecast out-of-sample. For example, the *l-step ahead* forecast: $\hat{y}_{T+l} = E_t[y_{t+l} \,|\, I_t]$.

• In the 1970s it was found that very simple time series models out-forecasted very sophisticated (big) economic models.

• This finding represented a big shock to the big multivariate models that were very popular then. It forced a re-evaluation of these big models.

## Time Series: Introduction – White Noise

• In general, we assume the error term, $\varepsilon_t$, is uncorrelated with everything, with mean 0 and constant variance, $\sigma^2$. We call a process like this a *white noise* (WN) *process*.

• We denote a WN process as

$$\varepsilon_t \sim WN(0, \sigma^2)$$

• White noise is the basic building block of all time series. It can be written as:

$$z_t = \sigma\, u_t, \qquad u_t \sim i.i.d\ WN\ (0, 1) \qquad \Rightarrow z_t \sim WN(0, \sigma^2)$$

• The $z_t$'s are random shocks, with no dependence over time, representing unpredictable events. It represents a model of news.

## Time Series: Introduction – Conditionality

• We make a key distinction: *Conditional* vs *Unconditional* moments. In time series we model the conditional mean as a function of its past, for example in an AR(1) process, we have:

$$y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t.$$

Then, the **conditional mean** forecast at time $t$, conditioning on information at time $I_{t-1}$, is:

$$E_t[y_t \,|\, I_{t-1}] = E_t[y_t] = \alpha + \beta\, y_{t-1}$$

Notice that the **unconditional mean**, μ, is given by:

$$E[y_t] = \alpha + \beta\, E[y_{t-1}] = \frac{\alpha}{1-\beta} = \mu = \text{constant} \qquad (\beta \neq 1)$$

The conditional mean is time varying; the unconditional mean is not!

Key distinction: Conditional vs. Unconditional moments.

## Time Series: Introduction – AR and MA models

• Two popular models for $E_t[y_t \,|\, I_{t-1}]$ :

– An **autoregressive** (**AR**) **process** models $E_t[y_t \,|\, I_{t-1}]$ with lagged dependent variables:

$$E_t[y_t \,|\, I_{t-1}] = f(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, y_{t-p})$$

**Example**: AR(1) process, $\quad y_t = \alpha + \beta\, y_{t-1} + \varepsilon_t.$

– A **moving average** (**MA**) **process** models $E_t[y_t \,|\, I_{t-1}]$ with lagged errors, $\varepsilon_t$:

$$E_t[y_t \,|\, I_{t-1}] = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots, \varepsilon_{t-q})$$

**Example**: MA(1) process, $\quad y_t = \mu + \theta_1\, \varepsilon_{t-1} + \varepsilon_t$

• There is a third model, ARMA, that combines lagged dependent variables and lagged errors.

## Time Series: Introduction – Forecasting (again)

• We want to select an appropriate time series model to forecast $y_t$. In this class, we will use linear models, with choices: AR($p$), MA($q$) or ARMA($p$, $q$).

• Steps for forecasting:

(1) Identify the appropriate model. That is, determine $p$, $q$.

(2) Estimate the model.

(3) Test the model.

(4) Forecast.

• In this lecture, we go over the statistical theory (stationarity, ergodicity), the main models (AR, MA & ARMA) and tools that will help us describe and identify a proper model.

## CLM Revisited: Time Series Implications

• With autocorrelated data, we get dependent observations. For example, with autocorrelated errors:

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + u_t,$$

the independence assumption is violated. The LLN and the CLT cannot be easily applied in this context. We need new tools.

• We introduce the concepts of **stationarity** and **ergodicity**. The ergodic theorem will give us a counterpart to the LLN.

To get asymptotic distributions, we also need a CLT for dependent variables, using new technical concepts: mixing and stationarity. Or we can rely on a new CLT: The *martingale difference sequence CLT*.

• We will not cover these technical points in detail.

## Time Series – Stationarity

- Consider the joint probability distribution of the collection of RVs:

$$F(y_{t_1}, y_{t_2}, \ldots, y_{t_T}) = F(Y_{t_1} \le y_{t_1}, Y_{t_2} \le y_{t_2}, \ldots, Y_{t_T} \le y_{t_T})$$

To do statistical analysis with dependent observations, we need extra assumptions. We need some form of invariance on the structure of the time series.

If the distribution $F$ is changing with every observation, estimation and inference become very difficult.

- Stationarity is an invariant property: the statistical characteristics of the time series do not change over time.

- There different definitions of stationarity, they differ in how strong is the invariance of the distribution over time.

## Time Series – Stationarity

- We say that a process is stationary of

*1st order* if $F(y_{t_1}) = F(y_{t_{1+k}})$ for any $t_1$, $k$

*2nd order* if $F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}})$ for any $t_1$, $t_2$, $k$

*Nth-order* if $F(y_{t_1}, \ldots, y_{t_T}) = F(y_{t_{1+k}}, \ldots, y_{t_{T+k}})$ for any $t_1$, ..., $t_T$, $k$

- *Nth-order* stationarity is a strong assumption (& difficult to verify in practice). *2nd order* stationarity is weaker: only consider mean and covariance (easier to verify in practice).

- Moments describe a distribution. We calculate moments as usual:

$$\text{E}[Y_t] = \mu$$

$$\text{Var}(Y_t) = \sigma^2 = E[(Y_t - \mu)^2]$$

$$\text{Cov}(Y_{t_1}, Y_{t_2}) = E[(Y_{t_1} - \mu)(Y_{t_2} - \mu)] = \gamma(t_1 - t_2)$$

## Time Series – Stationarity & Autocovariances

• $\text{Cov}(Y_{t_1}, Y_{t_2}) = \gamma(t_1 - t_2)$ is called the **auto-covariance function**.

Notes: $\gamma(t_1 - t_2)$ is a function of $k = t_1 - t_2$
$\gamma(0)$ is the variance.

• The autocovariance function is symmetric. That is,

$$\gamma(t_1 - t_2) = \text{Cov}(Y_{t_1}, Y_{t_2}) = \text{Cov}(Y_{t_2}, Y_{t_1}) = \gamma(t_2 - t_1)$$

$$\Rightarrow \gamma(k) = \gamma(-k)$$

• Autocovariances are unit dependent. We will have different values if we calculate the autocovariance for IBM returns in % terms or in decimal terms.

Remark: The autocovariance measures the (linear) dependence between two $Y_t$'s separated by $k$ periods.

## Time Series – Stationarity & Autocorrelations

• From the autocovariances, we derive the **autocorrelations**:

$$\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2}) = \frac{\gamma(t_1 - t_2)}{\sigma_{t_1}\sigma_{t_2}} = \frac{\gamma(t_1 - t_2)}{\gamma(0)}$$

the last step takes assumes: $\sigma_{t_1} = \sigma_{t_2} = \sqrt{\gamma(0)}$

• $\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2})$ is called the **auto-correlation function** (**ACF**), –think of it as a function of $k = t_2 - t_1$. The ACF is also symmetric.

• Unlike autocovoriances, autocorrelations are not unit dependent. It is easier to compare dependencies across different time series.

• Stationarity requires all these moments to be independent of time. If the moments are time dependent, we say the series is **non-stationary**.

## Time Series – Stationarity & Constant Moments

- For a strictly stationary process (constant moments), we need:

$$\mu_t = \mu$$
$$\sigma_t = \sigma$$

because $F(y_{t_1}) = F(y_{t_{1+k}}) \Rightarrow \quad \mu_{t_1} = \mu_{t_{1+k}} = \mu$

$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$

Then,

$$F(y_{t_1}, y_{t_2}) = F(y_{t_{1+k}}, y_{t_{2+k}}) \Rightarrow \text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$

$$\Rightarrow \rho(t_1, t_2) = \rho(t_1 + k,\ t_2 + k)$$

Let $t_1 = t - k$ & $t_2 = t$

$$\Rightarrow \rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t - k) = \rho(k) = \rho_k$$

The correlation between any two RVs depends on the time difference. Given the symmetry, we have $\rho(k) = \rho(-k)$.

## Time Series – Weak Stationary

- A **Covariance stationary** process (or *2nd -order weakly stationary*) has:
  - constant mean, $\mu$
  - constant variance, $\sigma^2$
  - covariance depends on time difference, $k$, between two RVs, $\gamma(k)$

That is, $Z_t$ is covariance stationary if:

$$E(Z_t) = \text{constant} = \mu$$
$$\text{Var}(Z_t) = \text{constant} = \sigma^2$$
$$\text{Cov}(Z_{t_1}, Z_{t_2}) = E[(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})] = \gamma(k = t_1 - t_2)$$

Remark: Covariance stationarity is only concerned with the covariance of a process, only the mean, variance and covariance are time-invariant. $N^{th}$*-order* stationarity is stronger and assumes that the whole distribution is invariant over time.

## Time Series – Stationarity: Example

**Example**: Assume $y_t$ follows an AR(1) process:

$$y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

- **Mean**

Taking expectations on both side:

$$\text{E}[y_t] = \phi\, \text{E}[y_{t-1}] + \text{E}[\varepsilon_t]$$
$$\mu = \phi\, \mu + 0$$
$$\text{E}[y_t] = \mu = 0 \qquad \text{(assuming } \phi \neq 1)$$

- **Variance**

Applying the variance on both side:

$$\text{Var}[y_t] = \gamma(0) = \phi^2\, \text{Var}[y_{t-1}] + \text{Var}[\varepsilon_t]$$
$$\gamma(0) = \phi^2\, \gamma(0) + \sigma^2$$
$$\gamma(0) = \frac{\sigma^2}{1-\phi^2} \qquad \text{(assuming } |\phi| < 1)$$

## Time Series – Stationarity: Example

**Example (continuation)**: $y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{WN}(0, \sigma^2)$

- **Covariance**

$$\gamma(1) = \text{Cov}[y_t, y_{t-1}] = \text{E}[y_t\, y_{t-1}] = \text{E}[(\phi\, y_{t-1} + \varepsilon_t)\, y_{t-1}]$$
$$= \phi\, \text{E}[y_{t-1}\, y_{t-1}] + \text{E}[\varepsilon_t\, y_{t-1}]$$
$$= \phi\, \text{E}[{y_{t-1}}^2]$$
$$= \phi\, \text{Var}[{y_{t-1}}^2]$$
$$= \phi\, \gamma(0)$$

$$\gamma(2) = \text{Cov}[y_t, y_{t-2}] = \text{E}[y_t\, y_{t-2}] = \text{E}[(\phi\, y_{t-1} + \varepsilon_t)\, y_{t-2}]$$
$$= \phi\, \text{E}[y_{t-2}\, y_{t-2}]$$
$$= \phi\, \text{Cov}[y_t, y_{t-1}]$$
$$= \phi\, \gamma(1)$$
$$= \phi^2\, \gamma(0)$$

$\vdots$

$$\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k\, \gamma(0)$$

## Time Series – Stationarity: Example

**Example (continuation)**: $y_t = \phi\, y_{t-1} + \varepsilon_t$, $\quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$

- **Covariance**

$$\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k\, \gamma(0)$$

$\Rightarrow$ If $|\phi| < 1$, $y_t$ process is covariance stationary: mean, variance and covariance are constant.

Note: From the autocovariance function, we can derive the auto-correlation function (ACF):

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi^k\, \gamma(0)}{\gamma(0)} = \phi^k$$

If $|\phi| < 1$, both autocovariance function and ACF show exponential decay.

## Time Series – Non-Stationarity: Example

**Example**: Assume $y_t$ follows a Random Walk with drift process:

$$y_t = \mu + y_{t-1} + \varepsilon_t, \qquad \text{with } \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

Doing backward substitution:

$$\begin{aligned} y_t &= \mu + (\mu + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= 2 * \mu + y_{t-2} + \varepsilon_t + \varepsilon_{t-1} \\ &= 2 * \mu + (\mu + y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \varepsilon_{t-1} \\ &= 3 * \mu + y_{t-3} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} \end{aligned}$$

$$\Rightarrow y_t = \mu\, t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$$

- **Mean** & **Variance**

$$E[y_t] = \mu\, t + y_0$$
$$\text{Var}[y_t] = \gamma(0) = \sum_{j=0}^{t-1} \sigma^2 = \sigma^2\, t$$

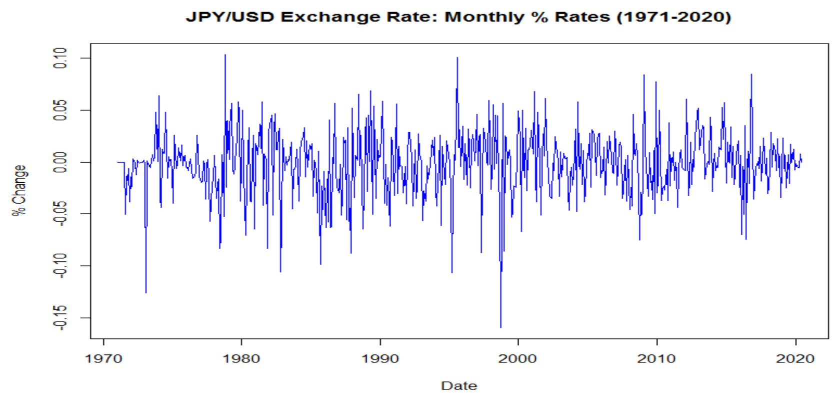
$\Rightarrow$ the process $y_t$ is non-stationary: moments are time dependent.

## Stationary Series: Examples

**Examples**: Assume $\varepsilon_t \sim$ WN(0, $\sigma^2$).

$y_t = 0.08 + \varepsilon_t + 0.4\,\varepsilon_{t-1}$ - MA(1) process

$y_t = 0.13\,y_{t-1} + \varepsilon_t$ - AR(1) process

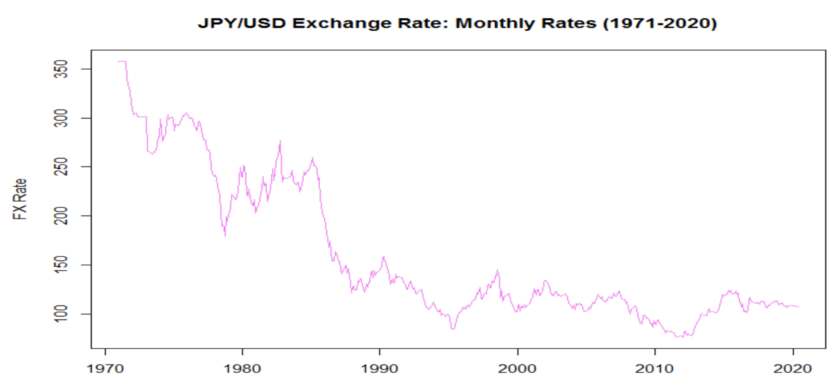



## Non-Stationary Series: Examples

**Examples**: Assume $\varepsilon_t \sim$ WN(0, $\sigma^2$).

$y_t = \mu\,t + \phi_1\,y_{t-1} + \phi_2\,y_{t-2} + \varepsilon_t$ - AR(2) with deterministic trend

$y_t = \mu + y_{t-1} + \varepsilon_t$ - Random Walk with drift



## Time Series – Stationarity: Remarks

- The main characteristic of time series is that observations are dependent.

- To analyze time series, however, we need to assume that some features of the series are not changing. If we have non-stationary series (say, mean or variance are changing with each observation), it is not possible to make inferences.

- Stationarity is an invariant property: the statistical characteristics of the time series do not vary over time.

- If IBM is weak stationary, then, the returns of IBM may change month to month or year to year, but the average return and the variance in two equal lengths time intervals will be more or less the same.

## Time Series – Stationarity (Again)

- In the long run, say 100-200 years, the stationarity assumption may not be realistic. After all, technological change has affected the return of IBM over the long run. But, in the short-run, stationarity seems likely to hold.

- In general, time series analysis is done under the stationarity assumption.

## Ergodicity

• We want to estimate the mean of the process $\{Z_t\}$, μ($Z_t$). But, we need to distinguishing between *ensemble average* (with $m$ observations) and *time average* (with $T$ observations):

- Ensemble Average: $\bar{\bar{z}} = \frac{\sum_{i=1}^{m} z_i}{m}$

- Time Series Average: $\bar{z} = \frac{\sum_{t=1}^{T} z_t}{T}$

Q: Which estimator is the most appropriate?
A: Ensemble Average. But, it is impossible to calculate. We only observe one $Z_t$, with dependent observations.

• Q: Under which circumstances we can use the time average (with only one realization of $\{Z_t\}$)? Is the time average an unbiased and consistent estimator of the mean? The *Ergodic Theorem* gives us the answer.

## Time Series – Ergodicity

• Intuition behind Ergodicity:
We go to a casino to play a game with 20% return, but on average, one gambler out of 100 goes bankrupt. If 100 gamblers play the game, there is a 99% chance of winning and getting a 20% return. This is the *ensemble scenario*. Suppose that gambler 35 is the one that goes bankrupt. Gambler 36 is not affected by the bankruptcy of gamble 35.

Suppose now that instead of 100 gamblers you play the game 100 times. This is the *time series* scenario. You win 20% every day until day 35 when you go bankrupt. There is no day 36 for you (dependence at work!).

<u>Result</u>: The probability of success from the group (ensemble scenario) does not apply to one person (time series scenario).

Ergodicity describes a situation where the ensemble scenario outcome applies to the time series scenario.

## Ergodicity

• With dependent observation, we cannot use the LLN used before. The *ergodicity theorem* plays the role of the LLN with dependent observations.

The formal definition of ergodicity is complex and is seldom used in time series analysis. One consequence of ergodicity is the ergodic theorem, which is extremely useful in time series.

It states that if $Z_t$ is an ergodic stochastic process then

$$\frac{1}{T}\sum_{t=1} g(Z_t) \xrightarrow{a.s} \text{E}[g(Z_t)]$$

for any function g(.). And, for any time shift $k$

$$\frac{1}{T}\sum_{t=1} g(Z_{t_1+k}, Z_{t_2+k}, \ldots, Z_{t_\tau+k}) \xrightarrow{a.s} \text{E}[g(Z_{t_1}, Z_{t_2}, \ldots, Z_{t_\tau}))]$$

where a.s. means *almost sure convergence*, a strong form of convergence.

## Ergodicity of the Mean

• *Definition*: A covariance-stationary process is *ergodic* for the mean if

$$\bar{z} \xrightarrow{p} \text{E}[Z_t] = \mu$$

• This result needs the variance of $\bar{z}$ to collapse to 0. It can be shown that the var[$\bar{z}$] can be written as a function of the autocorrelations, $\rho_k$:

$$\text{var}[\,\bar{z}\,] = \text{var}[\,(z_1 + z_2 + \cdots + z_T)/T] = \frac{\gamma_0}{T}\sum_k (1 - \frac{|k|}{T})\rho_k$$

**Theorem**: A sufficient condition for ergodicity for the mean is that the autocorrelations $\rho_k$ between two observations, say $(y_{t_i}, y_{t_j})$, $\rho(t_i, t_j)$ = $\rho_{t_i - t_j}$, go to zero as $t_i$ & $t_j$ grow further apart.

<u>Condition for ergodicity</u>: $\rho_k \to 0$, as $k \to \infty$

## Time Series – Lag Operator

- Define the operator $L$ as

$$L^k z_t = z_{t-k}.$$

- It is usually called *Lag operator*. But it can produce lagged or forward variables (for negative values of $k$). For example:

$$L^{-3} z_t = z_{t+3}.$$

- Also note that if $c$ is a constant $\Rightarrow L\, c = c$.

- Sometimes the notation for $L$ when working as a lag operator is $B$ *(backshift operator),* and when working as a forward operator is $F$.

- Important application: Differencing

$$\Delta\, z_t = (1 - L)\, z_t = z_t - z_{t-1}.$$
$$\Delta^2 z_t = (1 - L)^2 z_t = z_t - 2z_{t-1} + z_{t-2}.$$

## Time Series – Useful Result: Geometric Series

- The function $f(x) = (1-x)^{-1}$ can be written as an infinite geometric series (use a Maclaurin series around $c$ =0):

$$f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{n=0}^{\infty} x^n$$

- If we multiply $f(x)$ by a constant, $a$:

$$\sum_{n=0}^{\infty} a x^n = \frac{a}{1-x} \rightarrow \quad \sum_{n=1}^{\infty} a x^n = a\left(\frac{1}{1-x} - 1\right)$$

**Example:** In Finance we have many applications of the above results.
- A stock price, $P$, equals the discounted some of all futures dividends. Assume dividends are constant, $d$, and the discount rate is $r$. Then:

$$P_t = \sum_{t=1}^{\infty} \frac{d}{(1+r)^t} = d\left(\frac{1}{1-\frac{1}{1+r}} - 1\right) = d\left(\frac{1}{\frac{1+r-1}{1+r}} - 1\right) = \frac{d}{r}$$

where $x = \frac{1}{1+r}$

## Time Series – Useful Result: Geometric Series

• We will use this result when, under certain conditions, we invert a lag polynomial (say, $\theta(L)$) to convert an AR (MA) process into an infinite MA (AR) process.

**Example:** Suppose we have an MA(1) process:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\,\varepsilon_t \qquad - \theta(L) = (1 + \theta_1 L)$$

Recall,

$$f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{n=0}^{\infty} x^n$$

Let $x = -\theta_1 L$. Then, assuming that $\theta(L)^{-1}$ is well defined,

$$\theta(L)^{-1} = \frac{1}{1-(-\theta_1 L)} = 1 + (-\theta_1 L) + (-\theta_1 L)^2 + (-\theta_1 L)^3 + (-\theta_1 L)^4 + \ldots$$

$$= \sum_{n=0}^{\infty}(-\theta_1 L)^n = 1 - \theta_1 L + \theta_1^{\,2} L^2 - \theta_1^{\,3} L^3 + \theta_1^{\,4} L^4 + \cdots$$

## Time Series – Useful Result: Geometric Series

**Example (continuation):**

$$\theta(L)^{-1} = \sum_{n=0}^{\infty}(-\theta_1 L)^n = 1 - \theta_1 L + \theta_1^{\,2} L^2 - \theta_1^{\,3} L^3 + \theta_1^{\,4} L^4 + \cdots$$

Now, we multiply $\theta(L)^{-1}$on both sides of the MA process

$$y_t = \mu + \theta(L)\,\varepsilon_t.$$

Then,

$$\theta(L)^{-1}\, y_t = \theta(L)^{-1}\mu + \varepsilon_t = \mu^* + \varepsilon_t$$

$$\theta(L)^{-1}\, y_t = y_t - \theta_1 y_{t-1} + \theta_1^2\, y_{t-2} - \theta_1^3\, y_{t-3} + \theta_1^4\, y_{t-4} + \cdots$$
$$= \mu^* + \varepsilon_t$$

Then, solving for $y_t$:

$$y_t = \mu* + \theta_1 y_{t-1} - \theta_1^2\, y_{t-2} + \theta_1^3\, y_{t-3} - \theta_1^4\, y_{t-4} + \cdots + \varepsilon_t$$

That is, we get an AR(∞)!

## Moving Average Process

- An MA process models $E_t[y_t \mid I_{t-1}]$ with lagged error terms. An MA($q$) model involves $q$ lags.

- We keep the white noise assumption for $\varepsilon_t$: $\varepsilon_t \sim WN(0, \sigma^2)$

**Example**: A linear MA($q$) model:

$$y_t = \mu + \theta_1\, \varepsilon_{t-1} + \theta_2\, \varepsilon_{t-2} + \dots + \theta_q\, \varepsilon_{t-q} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t,$$

where

$$\theta(L) = 1 + \theta_1\, L + \theta_2\, L^2 + \theta_2\, L^3 + \dots + \theta_q\, L^q$$

- In time series, the constant does not affect the properties of AR and MA process. It is usually removed (think of the data analyze as demeaned). Thus, in this situation we say "without loss of generalization", we assume $\mu = 0$.

## Moving Average Process – Stationarity

- Q: Is MA($q$) stationary? Check the moments (assume $\mu = 0$).

$$y_t = \varepsilon_t + \theta_1\, \varepsilon_{t-1} + \theta_2\, \varepsilon_{t-2} + \dots + \theta_q\, \varepsilon_{t-q}$$

- **Mean**

$$E[y_t] = E[\varepsilon_t] + \theta_1\, E[\varepsilon_{t-1}] + \theta_2\, E[\varepsilon_{t-2}] + \dots + \theta_q\, E[\varepsilon_{t-q}] = 0$$

- **Variance**

$$Var[y_t] = Var[\varepsilon_t] + \theta_1^2\, Var[\varepsilon_{t-1}] + \theta_2^2\, Var[\varepsilon_{t-2}] + \dots + \theta_q^2 Var[\varepsilon_{t-q}]$$
$$= (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\, \sigma^2.$$

To get a positive variance, we require

$$(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) > 0,$$

which is always positive.

## Moving Average Process – Stationarity

- **Covariance**

$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$

$y_{t-1} = \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + \ldots + \theta_{q-1} \varepsilon_{t-q} + \theta_q \varepsilon_{t-(q+1)}$

- First, we compute $\gamma(1)$:

$$\begin{aligned}
\gamma(1) &= \text{Cov}[y_t, y_{t-1}] = \text{E}[y_t\, y_{t-1}] \\
&= \text{E}[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots + \theta_q \varepsilon_{t-q}) \\
&\qquad * (\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + \ldots + \theta_q \varepsilon_{t-(q+1)})] \\
&= \text{E}[\varepsilon_t\, \varepsilon_{t-1}] + \theta_1 \text{E}[\varepsilon_t\, \varepsilon_{t-2}] + \theta_2 \text{E}[\varepsilon_t\, \varepsilon_{t-3}] + \ldots. \\
&\quad + \theta_1 \text{E}[\varepsilon_{t-1}\, \varepsilon_{t-1}] + \theta_1^2 \text{E}[\varepsilon_{t-1}\, \varepsilon_{t-2}] + \theta_1\theta_2 \text{E}[\varepsilon_{t-1}\, \varepsilon_{t-3}] + \ldots \\
&\quad + \theta_2 \text{E}[\varepsilon_{t-2}\, \varepsilon_{t-1}] + \theta_2\theta_1 \text{E}[\varepsilon_{t-2}\, \varepsilon_{t-2}] + \theta_2\theta_1 \text{E}[\varepsilon_{t-2}\varepsilon_{t-3}] + \ldots \\
&\quad \ldots \\
&\quad + \theta_q \text{E}[\varepsilon_{t-q}\varepsilon_{t-1}] + \theta_q\theta_1 \text{E}[\varepsilon_{t-q}\varepsilon_{t-2}] + \ldots + \theta_q\, \theta_{q-1} \text{E}[\varepsilon_{t-q}\varepsilon_{t-q}] \\
&\quad + \theta_q^2 \text{E}[\varepsilon_{t-q}\, \varepsilon_{t-(q+1)}] \\
&= \theta_1 \sigma^2 + \theta_2\, \theta_1 \sigma^2 + \theta_3\, \theta_2 \sigma^2 + \ldots + \theta_q\, \theta_{q-1} \sigma^2 + 0
\end{aligned}$$

## Moving Average Process – Stationarity

- **Covariance**

$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots + \theta_q \varepsilon_{t-q}$

$y_{t-1} = \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + \ldots + \theta_{q-1} \varepsilon_{t-q} + \theta_q \varepsilon_{t-(q+1)}$

- We can also derive $\gamma(1)$ without computing the expectation of the cross products of errors. It is easier to look at the sum of $\text{E}[y_t\, \varepsilon_{t-j}]$'s:

$$\begin{aligned}
\gamma(1) &= \text{E}[y_t\, y_{t-1}] \\
&= \text{E}[y_t * (\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + \ldots + \theta_{q-1} \varepsilon_{t-q} + \theta_q \varepsilon_{t-(q+1)})] \\
&= \text{E}[y_t\, \varepsilon_{t-1}] + \theta_1 \text{E}[y_t\, \varepsilon_{t-2}] + \theta_2 \text{E}[y_t\, \varepsilon_{t-3}] + \ldots + \theta_q \text{E}[y_t\, \varepsilon_{t-(q+1)}] \\
&= \theta_1 \sigma^2 + \theta_2\, \theta_1 \sigma^2 + \theta_3\, \theta_2 \sigma^2 + \ldots + \theta_q\, \theta_{q-1} \sigma^2 + 0
\end{aligned}$$

- We continue with the derivation of $\gamma(k)$, for $k$ = 2, 3, …., using the sums of $\text{E}[y_t\, \varepsilon_{t-j}]$'s.

## Moving Average Process – Stationarity

- For $\gamma(2)$, we have:

$$y_t = \varepsilon_t + \theta_1\, \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q\, \varepsilon_{t-q}$$
$$y_{t-2} = \varepsilon_{t-2} + \theta_1\, \varepsilon_{t-3} + \theta_2\, \varepsilon_{t-4} + \theta_3\, \varepsilon_{t-5} + \dots + \theta_q\, \varepsilon_{t-(q+2)}$$

$$\gamma(2) = \text{Cov}[y_t, y_{t-2}] = \text{E}[y_t\, y_{t-2}]$$
$$= \text{E}[y_t * (\varepsilon_{t-2} + \theta_1\, \varepsilon_{t-3} + \theta_2\, \varepsilon_{t-4} + \dots + \theta_q\, \varepsilon_{t-q-2})]$$
$$= \text{E}[y_t\, \varepsilon_{t-2}] + \theta_1\, \text{E}[y_t\, \varepsilon_{t-3}] + \theta_2\, \text{E}[y_t\, \varepsilon_{t-4}] + \dots + \theta_q\, \text{E}[y_t\, \varepsilon_{t-q-2}]$$
$$= \theta_2\, \sigma^2 + \theta_3\, \theta_1\, \sigma^2 + \theta_4\, \theta_2\, \sigma^2 + \dots + \theta_q\, \theta_{q-2}\, \sigma^2 + 0$$

- Similarly, for $\gamma(q)$

$$y_{t-q} = \varepsilon_{t-q} + \theta_1\, \varepsilon_{t-(q+1)} + \theta_2\, \varepsilon_{t-(q+2)} + \dots + \theta_q\, \varepsilon_{t-(q+q)}$$

$$\gamma(q) = \text{E}[y_t\, y_{t-q}] =$$
$$= \text{E}[\varepsilon_t y_{t-q}] + \theta_1\, \text{E}[\, \varepsilon_{t-1} y_{t-q}] + \theta_2\, \text{E}[\varepsilon_{t-2} y_{t-q}] + \dots + \theta_q\, \text{E}[\varepsilon_{t-q} y_{t-q}]$$
$$= \theta_q\, \sigma^2$$

## Moving Average Process – Stationarity

Now, we have $\gamma(k)$ for all $k$:

$$\gamma(1) = \theta_1\, \sigma^2 + \theta_2\, \theta_1\, \sigma^2 + \theta_3 \theta_2\, \sigma^2 + \dots + \theta_q\, \theta_{q-1} \sigma^2$$
$$= \sigma^2 \textstyle\sum_{j=1}^{q} \theta_j\, \theta_{j-1} \qquad \text{(where } \theta_0 = 1)$$

$$\gamma(2) = \theta_2\, \sigma^2 + \theta_3 \theta_1\, \sigma^2 + \theta_4 \theta_2\, \sigma^2 + \dots + \theta_q\, \theta_{q-2}\, \sigma^2$$
$$= \sigma^2 \textstyle\sum_{j=2}^{q} \theta_j\, \theta_{j-2} \qquad \text{(where } \theta_0 = 1)$$

⋮

$$\gamma(q) = \theta_q\, \sigma^2$$
$$= \sigma^2 \textstyle\sum_{j=q}^{q} \theta_j\, \theta_{j-q} \qquad \text{(where } \theta_0 = 1)$$

- In general, for the $k$ *autocovariance:*

$$\gamma(k) = \sigma^2 \textstyle\sum_{j=k}^{q} \theta_j\, \theta_{j-k} \qquad \text{for } |\, k\, | \leq q$$
$$\gamma(k) = 0 \qquad \text{for } |\, k\, | > q$$

Remark: After lag $q$, the autocovariances (& ACFs) are 0.

## Moving Average Process – Stationarity

• In general, for the $k$ *autocovariance:*

$$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \, \theta_{j-k} \qquad \text{for } | \, k \, | \leq q$$

$$\gamma(k) = 0 \qquad \text{for } | \, k \, | > q$$

Remark: After lag $q$, the autocovariances are 0.

• It is easy to verify that the sums $\sum_{j=k}^{q} \theta_j \, \theta_{j-k}$ are finite. Then, mean, variance and covariance are constant.

⇒ MA($q$) is always stationary.

• It can be shown that for $\varepsilon_t$ with same distribution (say, normal) the autocovariances are non-unique.

## MA Process – Invertibility

**Example**: Two MA(1) processes that produce the same $\gamma(k)$:

$y_t = \varepsilon_t + .2\, \varepsilon_{t-1}$, $\quad \varepsilon_t \sim$ *i.i.d.* N(0, 25)

$z_t = \upsilon_t + 5\, \upsilon_{t-1}$, $\quad \upsilon_t \sim$ *i.i.d.* N(0; 1)

We only observe the time series, $y_t$ or $z_t$, and not the noise, $\varepsilon_t$ or $\upsilon_t$, thus, we cannot distinguish between the models using the autocovariances. Suppose you want to select one process to forecast. Which one? We select the model with an AR(∞) representation.

• Assuming $\theta(L) \neq 1$, we can invert $\theta(L)$. Then, by inverting $\theta(L)$, an MA($q$) process generates an AR process:

$$y_t = \mu + \theta(L)\, \varepsilon_t \qquad \Rightarrow \theta(L)^{-1}\, y_t = \Pi(L)\, y_t = \mu^* + \varepsilon_t.$$

Then, we have an infinite sum polynomial on $\theta L$. (Recall the geometric series result.) That is, we convert an MA($q$) into an AR(∞).

## MA Process – Invertibility

- We convert an MA($q$) into an AR($\infty$):

$$\sum_{j=0}^{\infty} \pi_j(L)\, y_t = \mu^* + \varepsilon_t$$

We need to make sure that $\Pi(L) = \theta(L)^{-1}$ is defined: We require $\theta(L) \neq 0$. When this condition is met, we can write $\varepsilon_t$ as a causal function of $y_t$. We say the MA is *invertible*. For this to hold, we require:

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

Technical note: An invertible MA($q$) is typically required to have roots of the lag polynomial equation $\theta(z) = 0$ greater than one in absolute value (**outside the unit circle**). In the MA(1) case, we require $|\theta_1| < 1$.

In the previous example, we select the model with $\theta_1 = .2$.

## MA Process – MA(1)

**Example**: MA(1) process:

$$y_t = \theta_1\, \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t, \qquad \text{with } \theta(L) = (1 + \theta_1 L)$$

- **Moments**

$E[y_t] = 0$

$Var[y_t] = \gamma(0) = \sigma^2 + \theta_1^2\, \sigma^2 = \sigma^2\,(1 + \theta_1^2)$

$Cov[y_t, y_{t-1}] = \gamma(1) = E[y_t\, y_{t-1}]$

$\qquad = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t)*(\theta_1\varepsilon_{t-2} + \varepsilon_{t-1})] = \theta_1\sigma^2$

$Cov[y_t, y_{t-2}] = \gamma(2) = E[y_t\, y_{t-2}]$

$\qquad = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t) * (\theta_1\, \varepsilon_{t-3} + \varepsilon_{t-2})] = 0$

⋮

$\gamma(k) = E[y_t\, y_{t-k}] = E[(\theta_1\varepsilon_{t-1} + \varepsilon_t) * (\theta_1\varepsilon_{t-(k+1)} + \varepsilon_{t-k})] = 0 \quad (\text{for } k>1)$

That is, for $|k| > 1$, $\gamma(k) = 0$.

## MA Process – MA(1)

**Example (continuation)**: MA(1) process:

$$y_t = \theta_1\, \varepsilon_{t-1} + \varepsilon_t$$

Check derivations with formula for autocovariances

$$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j\, \theta_{j-k} \quad \text{for } |\, k \,| \le q \qquad \text{(where } \theta_0 = 1\text{)}$$
$$\gamma(k) = 0 \quad \text{for } |\, k \,| > q$$

• $\gamma(k)$

$k = 0$ $\quad \gamma(0) = \sigma^2 \sum_{j=0}^{1} \theta_j\, \theta_{j-0} = \sigma^2 (1 + \theta_1{}^2)$

$k = 1$ $\quad \gamma(1) = \sigma^2 \sum_{j=1}^{1} \theta_j\, \theta_{j-1} = \sigma^2 (\theta_1)$

$k > 1$ $\quad \gamma(k) = 0$

## MA(1) Process – ACF

**Example (continuation)**: To get the ACF, we divide the autocovariances by $\gamma(0)$. Then, the autocorrelation function (ACF):

$$\rho(0) = \gamma(0)/\gamma(0) = 1$$

$$\rho(1) = \gamma(1)/\gamma(0) = \frac{\theta_1 \sigma^2}{\sigma^2 (1 + \theta_1{}^2)} = \frac{\theta_1}{(1 + \theta_1{}^2)}$$

$$\vdots$$

$$\rho(k) = \gamma(k)/\gamma(0) = 0 \qquad \text{(for } k > 1\text{)}$$

Remark: The autocovariance function is **zero** after lag 1. Similarly, the ACF is also **zero** after lag 1, that is, $y_t$ is correlated with itself ($y_t$) and $y_{t-1}$, but not $y_{t-2}$, $y_{t-3}$, ... Contrast this with the AR(1) model, where the correlation between $y_t$ and $y_{t-k}$ is never zero.

The ACF is usually shown in a plot. When we plot $\rho(k)$ against $k$, we plot also $\rho(0)$ which is 1.

## MA(1) Process – ACF

**Example (continuation)**:

$$\rho(1) = \frac{\theta_1}{(1 + \theta_1^2)}$$

Note that $|\rho(1)| \leq 0.5$.

When $\theta_1 = 0.5 \Rightarrow \rho(1) = 0.4$.
$\theta_1 = -0.9 \Rightarrow \rho(1) = -0.497238$.
$\theta_1 = -2 \Rightarrow \rho(1) = -0.4$.
$\theta_1 = 2 \Rightarrow \rho(1) = 0.4$. (same $\rho(1)$ for $\theta_1$ & $1/\theta_1$)

Thus, the two MA processes, with $\theta_1 = 0.5$ and $\theta_1 = 2$, have the same ACF.

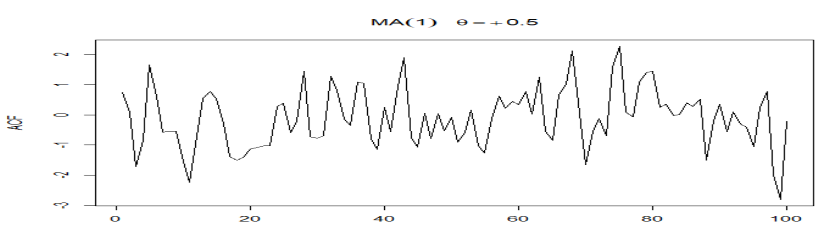
## MA(1) Process – ACF: Simulations

**Example**: We simulate and plot three MA(1) processes, with standard normal $\varepsilon_t$ -i.e., $\mu = 0$ & $\sigma = 1$:

$$y_t = \varepsilon_t + 0.5\, \varepsilon_{t-1}$$
$$y_t = \varepsilon_t - 0.9\, \varepsilon_{t-1}$$
$$y_t = \varepsilon_t - 2\, \varepsilon_{t-1}$$

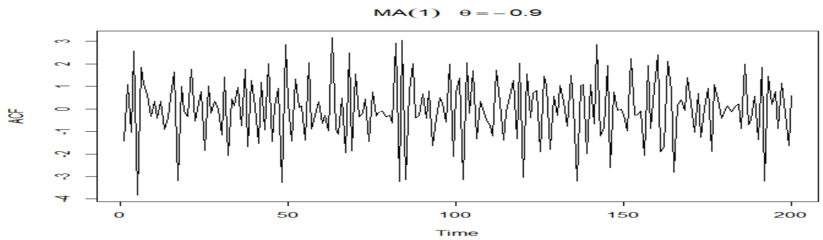
R script to plot $y_t = \varepsilon_t + \mathbf{0.5}\, \varepsilon_{t-1}$ with 200 simulations

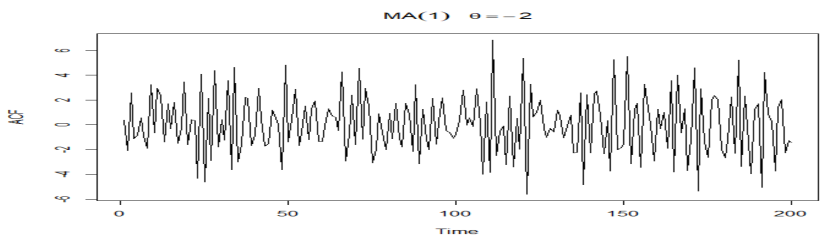
```
> plot(arima.sim(list(order=c(0,0,1), ma=0.5), n=200), ylab="ACF",
main=(expression(MA(1)~~~theta==+.5)))
```



## MA(1) Process – ACF: Simulations

**Example (continuation)**:




Note: The process $\theta_1 > 0$ is smoother than the ones with $\theta_1 < 0$.

## MA(1) Process – ACF: Simulations

**Example (continuation)**: Below, we compute and plot the ACF for the 3 simulated process.

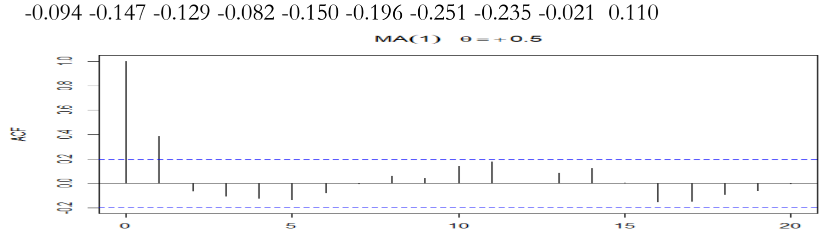
1) $y_t = \varepsilon_t + 0.5\, \varepsilon_{t-1}$

```
sim_ma1_5 <- arima.sim(list(order=c(0,0,1), ma=0.5), n=100)
acf_ma1_5 <- acf(sim_ma1_5, main=(expression(MA(1)~~~theta==+.5)))
> acf_ma1_5
```

Autocorrelations of series 'sim_ma1_5', by lag

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.000 | 0.438 | 0.069 | 0.014 | 0.103 | 0.173 | 0.107 | 0.015 | -0.080 | -0.054 | 0.011 | -0.006 | 0.041 | 0.000 |

| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|
| -0.094 | -0.147 | -0.129 | -0.082 | -0.150 | -0.196 | -0.251 | -0.235 | -0.021 | 0.110 |



# MA(1) Process – ACF: Simulations

**Example (continuation)**:

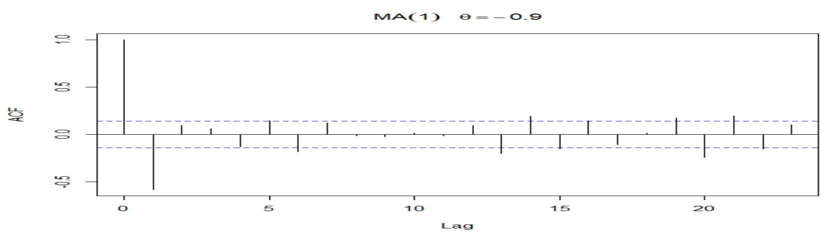
2) $y_t = \varepsilon_t$ **- 0.9** $\varepsilon_{t-1}$

```
sim_ma1_9 <- arima.sim(list(order=c(0,0,1), ma=-0.9), n=100)
acf_ma1_9 <- acf(sim_ma1_5, main=(expression(MA(1)~~~theta==+.5)))
> acf_ma1_9
```

Autocorrelations of series 'sim_ma1_9', by lag

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.000 | -0.584 | 0.093 | 0.061 | -0.132 | 0.147 | -0.181 | 0.122 | -0.013 | -0.023 | 0.014 | -0.012 | 0.092 | -0.199 |

| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|
| 0.193 | -0.155 | 0.143 | -0.107 | 0.014 | 0.174 | -0.244 | 0.196 | -0.154 | 0.105 |




# MA(1) Process – ACF: Simulations

**Example (continuation)**:

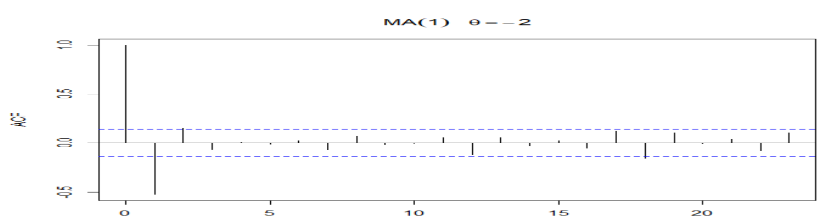
3) $y_t = \varepsilon_t$ **- 2** $\varepsilon_{t-1}$

```
sim_ma1_2 <- arima.sim(list(order=c(0,0,1), ma=-2), n=100)
acf_ma1_2 <- acf(sim_ma1_2, main=(expression(MA(1)~~~theta==-2)))
> acf_ma1_2
```

Autocorrelations of series 'sim_ma1_2', by lag

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.000 | -0.524 | 0.150 | -0.064 | 0.006 | -0.014 | 0.022 | -0.070 | 0.068 | -0.015 | -0.002 | 0.054 | -0.121 | 0.055 |

| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|
| -0.029 | 0.026 | -0.054 | 0.121 | -0.156 | 0.106 | -0.009 | 0.037 | -0.080 | 0.104 |



## MA Process – Example: MA(1)

**Example (continuation)**:

– Invertibility: If $|\theta_1| < 1$, we can write $(1 + \theta_1 L)^{-1}\, y_t + \mu^* = \varepsilon_t$

$$\Rightarrow \quad (1 - \theta_1 L + \theta_1{}^2 L^2 + \ldots + \theta_1{}^j L^j + \ldots)\, y_t + \mu^* = \mu^* + \sum_{i=1}^{\infty} \pi_i(L) y_t = \varepsilon_t$$

That is, $\pi_i = \theta_1{}^i$.

The simulated process with $\theta_1$= **-2** is non-invertible, the infinite sum of $\pi_i$ would explode. We would select the MA(1) with $\theta_1$ = **-.5**.

## MA Process – MA(2)

**Example**: MA(2) process:

$$y_t = \mu + \theta_2\, \varepsilon_{t-2} + \theta_1\, \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L)\, \varepsilon_t,$$

with $\theta(L) = (1 + \theta_1\, L + \theta_2\, L^2)$.

• **Moments**

$$E(Y_t) = \mu$$

$$\gamma_k = \begin{cases} \sigma^2\left(1 + \theta_1^2 + \theta_2^2\right), & k = 0 \\ -\theta_1 \sigma^2 (1 - \theta_2), & |k| = 1 \\ -\theta_2 \sigma^2, & |k| = 2 \\ 0, & |k| > 2 \end{cases}$$

<u>Remark</u>: The autocovariance function is **zero** after lag 2. Similarly, the ACF is also **zero** after lag 2.

## MA Process – Example: MA(2)

– Invertibility: The roots of $\lambda^2 - \theta_1 \lambda - \theta_2 = 0$ all lie inside the <u>unit circle</u>. It can be shown the invertibility condition for an MA(2) process is:

$$\theta_1 + \theta_2 < 1$$
$$\theta_1 - \theta_2 < 1$$
$$-1 < \theta_2 < 1.$$

## MA Process – Estimation

- MA processes are more complicated to estimate. Consider an MA(1):

$$y_t = \varepsilon_t + \theta\, \varepsilon_{t-1}$$

We cannot do OLS, since we do not observe $\varepsilon_{t-1}$. But, based on the ACF, we estimate $\theta$.

- The auto-correlation of order one is:

$$\rho(1) = \theta_1/(1+\theta_1^2)$$

Then, we can use the method of moments (MM), which sets the theoretical moment equal to the estimated sample moment $\rho_1$, $r_1$. Then, we solve for the parameter of interest, $\theta$:

$$r_1 = \frac{\hat{\theta}}{(1+\hat{\theta}^2)} \quad \Rightarrow \quad \hat{\theta} = \frac{1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

- A nonlinear solution and difficult to solve.

## MA Process – Estimation

• Alternatively, if $|\theta| < 1$, we can invert the MA(1) process. Then, based on the AR representation, we can try finding $a \in (-1; 1)$,

$$\varepsilon_t(a) = y_t + a\, y_{t-1} + a^2\, y_{t-2} + a^3\, y_{t-3} + \ldots.$$

and look (numerically) for the least-square estimator

$$\hat{\boldsymbol{\theta}} = \arg\min_\theta \{S(\boldsymbol{y}; \boldsymbol{\theta}) = \sum_{i=1}^{T} \varepsilon_i(a)^2\}$$

where $a^i = \theta_1{}^i$.

## Autoregressive (AR) Process

• We model the conditional expectation of $y_t$, $E[y_t | I_{t-1}]$, as a function of its past history. We assume $\varepsilon_t$ follows a WN(0, $\sigma^2$).

• The most common models are AR models. An AR(1) model involves a single lag, while an AR($p$) model involves $p$ lags. Then, the AR($p$) process is given by:

$$y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \ldots + \phi_p\, y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

Using the lag operator we write the AR($p$) process: $\phi(L)\, y_t = \varepsilon_t$

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

## AR Process: SDE

- We can look at an AR($p$) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t,$$

as a *stochastic (linear) difference equation* (SDE). With difference equations we try to get a solution –i.e., given some initial conditions/history, we know the value of $y_t$ for any $t$– and, then, we study its characteristics (stability, long-run value, etc.).

The solution to a DE can be written as a sum of two solutions:

1) Homogeneous equation (the part that only depends on the $y_t$'s):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} \qquad (\text{set } \mu + \varepsilon_t = 0)$$

2) A particular solution to the equation.

- Once we get a solution, we study its stability. We want a stable one.

## AR Process – AR(1): Solution

- We get a solution to the simple case, the AR(1) process.

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim WN.$$

We use the backward substitution method:

$$\begin{aligned} y_t &= \mu + \phi_1 (\mu + \phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \mu (1 + \phi_1) + \phi_1^2 y_{t-2} + \varepsilon_t + \phi_1 \varepsilon_{t-1} \\ &= \mu (1 + \phi_1) + \phi_1^2 (\mu + \phi_1 y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \phi_1 \varepsilon_{t-1} \\ &= \mu (1 + \phi_1 + \phi_1^2) + \phi_1^3 y_{t-3} + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} \\ &\vdots \end{aligned}$$

$$\Rightarrow y_t = \mu (1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{t-1}) + \sum_{j=0}^{t-1} \phi_1^j \varepsilon_{t-j} + \phi_1^t y_0$$

The solution is a function of $t$, the whole sequence $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1$ and the initial condition $y_0$. The effect of $y_0$ "dies out" if $|\phi_1| < 1$.

## AR Process – AR(1): Solution & Stability

• The stability of the solution is crucial. With a stable solution, $Y_t$ does not explode. This is good: We need well defined moments.

It turns out that the stability of the equation depends on the solution to the homogenous equation. In the AR(1) case:

$$y_t = \phi_1 \, y_{t-1}$$

with solution $y_t = \phi_1^t \, y_0$

If $|\phi_1| < 1$, $y_t$ never explodes, as $t \to \infty$. In this case, in the solution to the AR(1) process, the effect of $y_0$ "dies out" as $t \to \infty$.

We can analyze the stability from the point of view of the roots of the characteristic equation of the process or the lag polynomial.

## AR Process – AR(1): Solution & Stability

• We can analyze the stability from the point of view of the roots of the lag polynomial. For the AR(1) process

$$\phi(z) = 1 - \phi_1 \, z = 0 \qquad \Rightarrow |z| = \frac{1}{|\phi_1|} > 1$$

That is, the AR(1) process is stable if the root of $\phi(z)$ is greater than one (also said as "**the roots lie outside the unit circle**").

This result generalizes to AR($p$) process:

**Theorem**

A necessary and sufficient condition for global asymptotical stability of a $p^{\text{th}}$ order deterministic difference equation with constant coefficients is that *all roots* of the associated lag polynomial equation $\phi(z)$=0 have *moduli* strictly more than 1.

(For the case of real roots, moduli means "absolute values.")

## AR(1) Process – Stationarity & ACF

- An AR(1) model:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN.$$

Recall that in a previous example, under the stationarity condition $|\phi_1| < 1$, we derived the moments:

$E[y_t] = \mu = 0$ (assuming $\phi_1 \neq 1$)

$Var[y_t] = \gamma(0) = \sigma^2/(1 - \phi_1^2)$ (assuming $|\phi_1| < 1$)

$\gamma(1) = \phi_1 \gamma(0)$

$\gamma(2) = \phi_1^2 \gamma(0)$

$\gamma(3) = \phi_1^3 \gamma(0)$

⋮

$\gamma(k) = \phi_1^k \gamma(0)$

## AR(1) Process – Stationarity & ACF

- We want to derive the autocorrelation: $\rho(t_1, t_2) = \frac{\gamma(t_1 - t_2)}{\sigma_{t_1}\sigma_{t_2}}$

If the process is stationary ($\sigma_t = \sigma_{t-1} = \sqrt{\gamma(0)}$)

$$\rho(1) = \rho(t, t-1) = \frac{\gamma(1)}{\sigma_t \sigma_{t-1}} = \frac{\gamma(1)}{\gamma(0)} = \phi_1$$

$$\rho(2) = \frac{\gamma(2)}{\gamma(0)} = \phi_1^2$$

⋮

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k$$

Remark: The ACF decays with $k$.

When we plot $\rho(k)$ against $k$, we plot also $\rho(0)$ which is 1.

- Note that when $\phi_1 = 1$, the AR(1) is non-stationary, $\rho(k) = 1$, for all $k$. The present and the past are always correlated!

## AR(1) Process – Stationarity & ACF

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k$$

• Again, when $|\phi_1| < 1$, the autocorrelations do not explode as $k$ increases. There is an exponential decay towards zero.

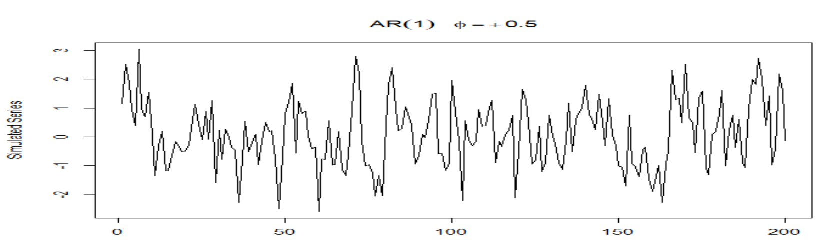
Note:
– when $0 < \phi_1 < 1$ ⇒ All autocorrelations are positive.
– when $-1 < \phi_1 < 0$ ⇒ The sign of $\rho(k)$ shows an alternating pattern beginning a negative value.

## AR(1) Process – Stationarity & ACF: Simulations

**Example**: We simulate and plot three MA(1) processes, with standard normal $\varepsilon_t$ -i.e., $\sigma = 1$:

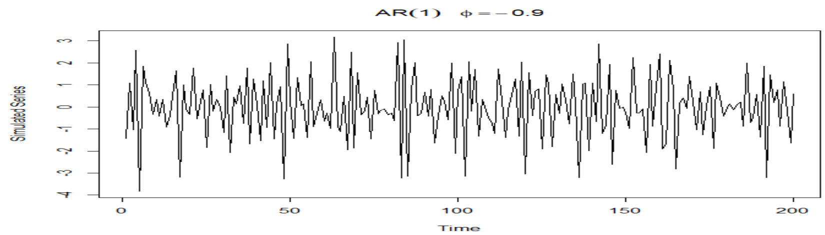
$y_t = 0.5\, y_{t-1} + \varepsilon_t$
$y_t = -0.9\, y_{t-1} + \varepsilon_t$
$y_t = 2\, y_{t-1} + \varepsilon_t$

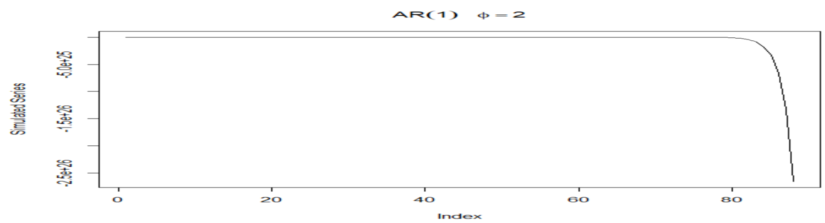
R script to plot $y_t$ = **0.5** $y_{t-1} + \varepsilon_t$ with 200 simulations
> plot(arima.sim(list(order=c(1,0,0), ar=**0.5**), n=200), ylab="ACF", main=(expression(AR(1)~~~phi==+.5)))



## AR(1) Process – Stationarity & ACF: Simulations

**Example (continuation)**:




Note: The process $\theta_1 > 0$ is smoother than the ones with $\theta_1 < 0$. The process with $|\theta_1| > 1$, explodes!

## AR(1) Process – Stationarity & ACF: Simulations

**Example (continuation)**: Below, we compute and plot the ACF for the the two stable simulated process.

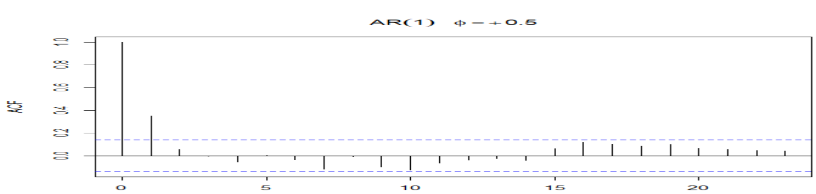
1) $y_t = \mathbf{0.5}\, y_{t-1} + \varepsilon_t$

```
sim_ar1_5 <- arima.sim(list(order=c(1,0,0), ar=0.5), n=200)
acf_ar1_5 <- acf(sim_ar1_5, main=(expression(AR(1)~~~phi==+.5)))
acf_ar1_5
```

Autocorrelations of series 'sim_ma1_5', by lag

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.000 | 0.351 | 0.055 | -0.005 | -0.054 | 0.002 | -0.036 | -0.119 | -0.008 | -0.099 | -0.125 | -0.066 | -0.036 | -0.023 |

| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|
| -0.042 | 0.062 | 0.119 | 0.102 | 0.087 | 0.099 | 0.065 | 0.056 | 0.047 | 0.044 |



## AR(1) Process – Stationarity & ACF: Simulations

**Example (continuation)**:

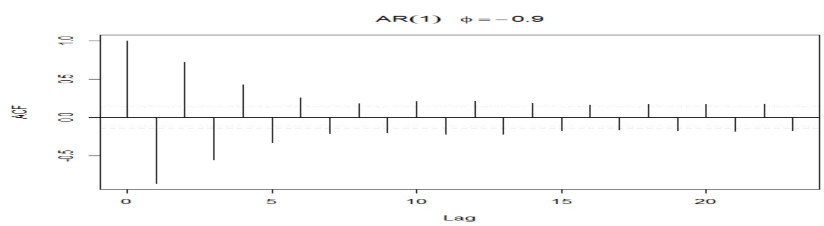
2) $y_t = -0.9\, y_{t-1} + \varepsilon_t$

sim_ar1_9 <- arima.sim(list(order=c(1,0,0), ar=-0.9), n=200) acf_ar1_9 <- acf(sim_ar1_9, main=(expression(AR(1)~~~phi==-.9)))

> acf_ar1_9

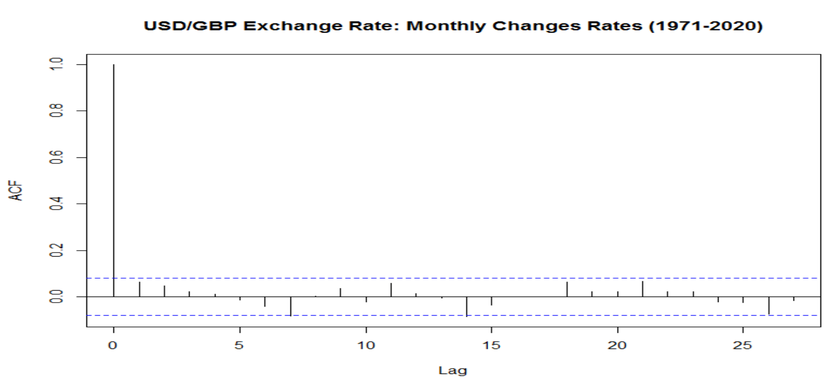
Autocorrelations of series 'sim_ma1_9', by lag

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.000 | -0.584 | 0.093 | 0.061 | -0.132 | 0.147 | -0.181 | 0.122 | -0.013 | -0.023 | 0.014 | -0.012 | 0.092 | -0.199 |
| 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | | | | |
| 0.193 | -0.155 | 0.143 | -0.107 | 0.014 | 0.174 | -0.244 | 0.196 | -0.154 | 0.105 | | | | |



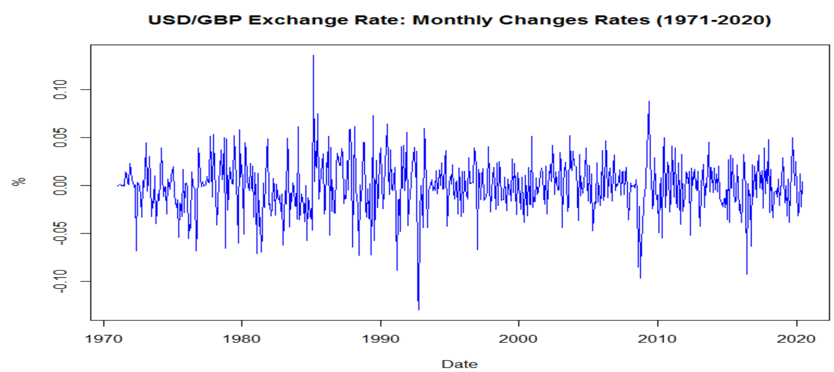

## AR(1) Process – Stationarity & ACF: Examples

**Example:** A process with $|\phi_1| < 1$ (actually, 0.065) is the monthly changes in the USD/GBP exchange rate. Below we plot its corresponding ACF:



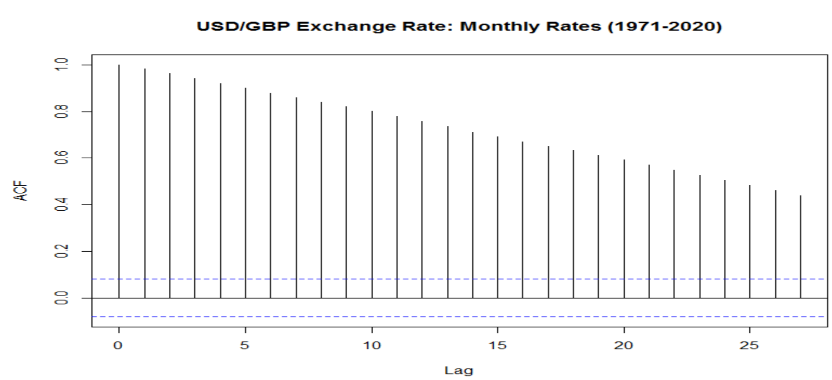
## AR(1) Process – Stationarity & ACF

**Example:** Below we plot the monthly changes in the USD/GBP exchange rate. Stationary series do not look smooth:




## AR(1) Process – Stationarity & ACF

**Example:** A process with $\phi_1 \approx 1$ (actually, 0.99) is the nominal USD/GBP exchange rate. Below, we plot the ACF, it is not 1 all the time, but its decay is very slow (after 30 months, it is still .40 correlated!):



## AR(1) Process – Stationarity & ACF

**Example:** Below we plot the nominal USD/GBP exchange rate. Stationary series look smooth, smooth enough that you can clearly spot trends: