Lecture 8-a2
Time Series:
Stationarity, AR(p) & MA(q)

Brooks (4th edition): Chapter 6

 1

Review: Times Series

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

• 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**.

• 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.

• 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.

Time Series: Introduction – Dependence

• Given the sequential nature of the time series 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.

• We need new conditions in the DGP to make sure the sample moments are good estimators population moments.

• The new assumptions and tools are needed: **stationarity**, **ergodicity**, CLT for martingale difference sequences (**MDS CLT**).

• The most common measure of dependence: **Covariance**.

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

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

Time Series: Introduction – AR and MA models

• Two popular models for $E_t[y_t \,|\, I_t]$:

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

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

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

– A **moving average** (**MA**) **process** models $E_t[y_t|I_t]$ with lagged errors, ε_t:

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

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

• 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.

Time Series – Stationarity

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

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

To do statistical analysis with dependent observations, we need extra assumptions. 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. We focus on **Weak stationarity**, which considers means & covariances independent of time:

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

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

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

Time Series – Stationarity & Autocorrelations

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

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

- $\text{Corr}(y_t, y_{t+k}) = \rho(k)$ is called the **auto-correlation function** (**ACF**), –think of it as a function of k. The ACF is also symmetric.

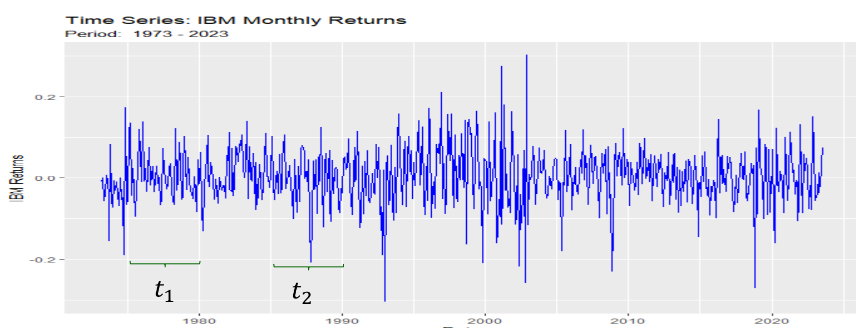
- Unlike autocovoriances, autocorrelations are not unit dependent. It is easier to compare dependencies across different time series.

- 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

Example: Informally, we check if in any two periods separated by k observations, we have similar means, variances and covariances. That is,

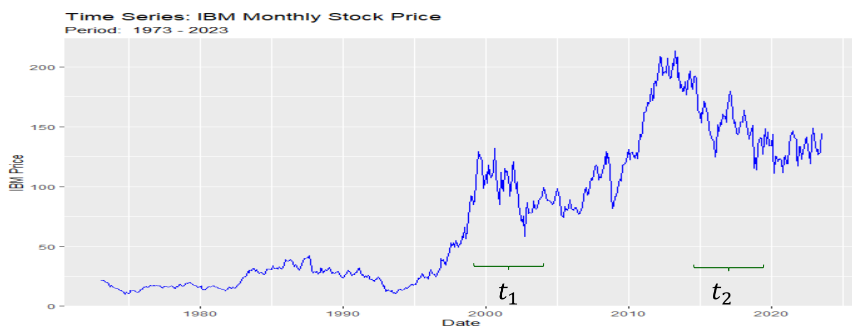
$$\mu_{t_1} = \mu_{t_{1+k}} = \mu$$
$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$
$$\text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$



Time Series – Stationarity & Constant Moments

Example: Informally, we check if in any two periods separated by k observations, we have similar means, variances and covariances. That is,

$$\mu_{t_1} = \mu_{t_{1+k}} = \mu$$
$$\sigma_{t_1} = \sigma_{t_{1+k}} = \sigma$$
$$\text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_{1+k}}, y_{t_{2+k}})$$



Time Series – Weak (Covariance) Stationary

- Z_t is covariance stationary if:

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

Remark: Covariance stationarity is only concerned with the covariance of a process, only the mean, variance and covariance are time-invariant.

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**

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

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

⋮

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

Time Series – Stationarity: Example

Example (continuation):

- **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.

Remark: To establish stationarity, we need to impose conditions on the AR parameters. (Conditions are not needed for MA processes.)

Note: From the autocovariance function, we derive ACF:

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

If $|\phi| < 1$, autocovariance function & 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**

  $\text{E}[y_t] = \mu\, t + y_0$

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

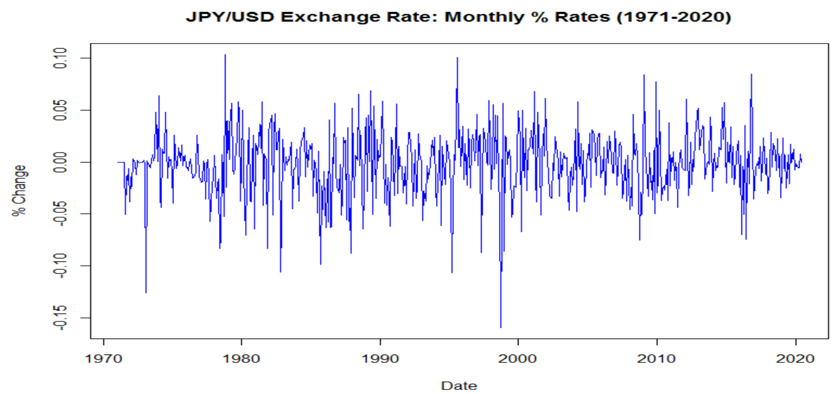
$\Rightarrow$ the process y_t is non-stationary: moments are time dependent.

Stationary Series: Examples

Examples: Assume ε_t ~ WN(0, σ^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

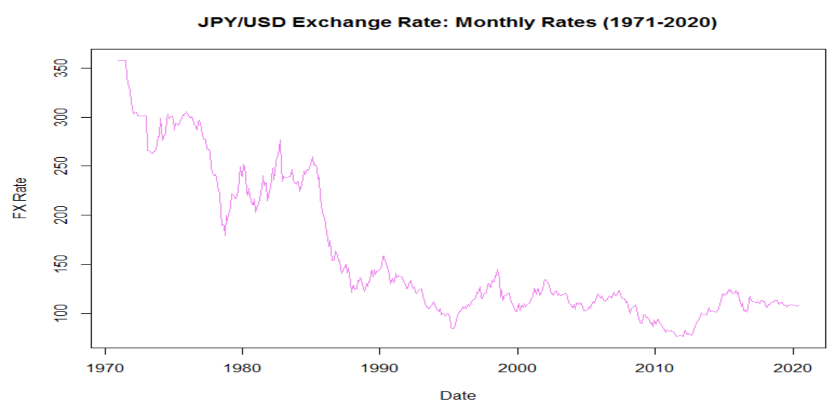


Non-Stationary Series: Examples

Examples: Assume ε_t ~ WN(0, σ^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

• Main characteristic of time series: Observations are **dependent**.

• If we have non-stationary series (say, mean or variance are changing with each observation), it is not possible to make inferences.

• Stationarity: 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-length time intervals will be more or less the same.

• In the long run, say 100-200 years, stationarity 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.

Ergodicity

• We want to estimate the mean of the process $\{Z_t\}$, $\mu(Z_t)$. 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 for a time series. 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!).

Result: 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 as we have done before with *i.i.d.* observations. 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

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

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

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

Theorem: A sufficient condition for ergodicity for the mean:

$$\rho_k \to 0 \quad \text{as} \quad k = t_i - t_j \to \infty$$

We need the correlation between (y_{t_i}, y_{t_j}) to decrease as they grow further apart in time.

- If the conditions of the *Ergodic Theorem* are met, we can use $\bar{z}$ instead of $\bar{\bar{z}}$.

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 $\quad\Rightarrow L\, c = c$.

- 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:

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

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: Application

- 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: Application

Example (continuation):

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

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

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

Then,

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

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

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(∞)!

Time Series – Useful Result: Invertibility

Example (continuation):

Then, solving for y_t:

$$y_t = \mu* + \theta_1 y_{t-1} - \theta_1^2\,y_{t-2} + \theta_1^3\,y_{t-3} - \theta_1^4\,y_{t-4} + \cdots + \varepsilon_t$$
$$= \mu* + \pi_1 y_{t-1} + \pi_2\,y_{t-2} + \pi_3\,y_{t-3} + \pi_4\,y_{t-4} + \cdots + \varepsilon_t$$

That is, $\pi_j = (-1) * (-\theta_1)^j$

- Now, $\quad y_t = \mu* + \sum_{j=1}^{\infty}\pi_j\,y_{t-j} + \varepsilon_t$

We express y_t as infinite AR process. We have an infinite sum of $\pi_i y_{t-i}$! To be useful for forecasting purposes, we need to make sure that this infinite sum is finite.

Restriction: Make sure the π_i's do not explode –i.e., $|\theta_1| < 1$. Under this condition, we will call the polynomial $\theta(L)$ **invertible**.

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

Example: A linear MA(q) model:

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

where

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

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

MA Process – MA(1): Stationarity

Example: MA(1) process ($\mu = 0$):

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

- **Mean**

$E[y_t] = 0$

- **Variance**

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

- **Covariance**

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

$\text{Cov}[y_t, y_{t-2}] = \gamma(2) = E[y_t\, y_{t-2}]$
$= E[(\theta_1 \varepsilon_{t-1} + \varepsilon_t) * (\theta_1 \varepsilon_{t-3} + \varepsilon_{t-2})] = 0$

MA Process – MA(1): Stationarity

Example (continuation): MA(1) process:

- **Covariance**

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

$\text{Cov}[y_t, y_{t-2}] = \gamma(2) = \text{E}[y_t\, y_{t-2}]$
$= \text{E}[(\theta_1\varepsilon_{t-1} + \varepsilon_t) * (\theta_1\, \varepsilon_{t-3} + \varepsilon_{t-2})] = 0$

⋮

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

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

⇒ MA(1) is always stationary –i.e., independent of values of θ_1.

Remark: The MA(q=1) process has $\gamma(q) = 0$, for $q > 1$. This result generalizes to MA(q) process: after lag q, the autocovariances are 0.

MA(1) Process – ACF

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

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

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

⋮

$\rho(k) = \gamma(k)/\gamma(0) = 0$ (for $k > 1$)

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, the **autocorrelogram**. 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 $\theta_1 = \mathbf{0.5}$ $\Rightarrow \rho(1) = \mathbf{0.4}$.

$\theta_1 = -0.9$ $\Rightarrow \rho(1) = -0.497238$.

$\theta_1 = -2$ $\Rightarrow \rho(1) = -0.4$.

$\theta_1 = \mathbf{2}$ $\Rightarrow \rho(1) = \mathbf{0.4}$. (same $\rho(1)$ for θ_1 & $\frac{1}{\theta_1}$.)

Note: Both MA(1) processes, with $\theta_1 = \mathbf{0.5}$ and $\theta_1 = \mathbf{2}$, have the same ACF. That is, ACFs are not unique. This is a problem: we deduce the order and the coefficients through the ACF, which is what we observe.

MA Process – MA(q): Stationarity

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

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

- **Mean**

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

- **Variance**

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

To get a positive variance, we require

$(1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2) > 0.$ (always positive)

- **Covariance**

It can shown (check book) for the k *autocovariance:*

$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j\, \theta_{j-k}$ for $|\, k\,| \leq q$ (where $\theta_0 = 1$)

$\gamma(k) = 0$ for $|\, k\,| > q$

MA Process – MA(q): Stationarity

• **Covariance**

$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \, \theta_{j-k}$ for $|\, k \,| \leq q$ (where $\theta_0 = 1$)

$\gamma(k) = 0$ for $|\, k \,| > q$

Remark: After lag q, the autocovariances are 0.

Applying formula:

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

$= \sigma^2 \theta_1 + \sigma^2 \theta_2 \, \theta_1 + \sigma^2 \theta_3 \, \theta_2 + \ldots + \sigma^2 \theta_q \, \theta_{q-1}$

$\gamma(2) = \sigma^2 \sum_{j=2}^{q} \theta_j \, \theta_{j-2}$

$= \sigma^2 \theta_2 + \sigma^2 \theta_3 \, \theta_1 + \sigma^2 \theta_4 \, \theta_2 + \ldots + \sigma^2 \theta_q \, \theta_{q-2}$

$\vdots$

$\gamma(q) = \sigma^2 \sum_{j=q}^{q} \theta_j \, \theta_{j-q} = \sigma^2 \theta_q$

MA Process – MA(q): Stationarity for MA(1)

• The k *autocovariance:*

$\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \, \theta_{j-k}$ for $|\, k \,| \leq q$ (where $\theta_0 = 1$)

$\gamma(k) = 0$ for $|\, k \,| > q$

• 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 –i.e., independent of values of θ_j's.

• Check for MA(1):

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

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

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

Remark: After lag $q = 1$, the autocovariances of an MA(1) are 0.

MA Process – Invertibility

- As mentioned above, the autocovariances are non-unique.

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

$$y_t = \varepsilon_t + 0.2\,\varepsilon_{t-1}, \qquad \varepsilon_t \sim i.i.d.\ N(0, 25)$$
$$z_t = \upsilon_t + 5\,\upsilon_{t-1}, \qquad \upsilon_t \sim i.i.d.\ N(0; 1)$$

We only observe the time series, y_t or z_t, and not the noise, ε_t or υ_t. We cannot distinguish between the models using the autocovariances.

We want to select one process to forecast: We select the model with an AR(∞) representation that does not explode: That is, we select the process that is **invertible**.

- Assuming $\theta(L) \neq 1$, we invert $\theta(L)$:

$$y_t = \mu + \theta(L)\,\varepsilon_t \qquad \Rightarrow \theta(L)^{-1} y_t = \Pi(L)\,y_t = \mu^* + \varepsilon_t.$$
$$\Rightarrow y_t = \mu* + \textstyle\sum_{j=1}^{\infty} \pi_j\, y_{t-j} + \varepsilon_t$$

MA Process – Invertibility

- We convert an MA(*q*) into an AR(∞):

$$y_t = \mu* + \textstyle\sum_{j=1}^{\infty} \pi_j\, y_{t-j} + \varepsilon_t$$

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 ε_t as a causal function of y_t. We say the MA is *invertible*. For this to hold, we require:

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

<u>Technical note</u>: 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,

$$\theta(z) = (1 + \theta_1 z) = 0 \qquad \Rightarrow \text{root: } z = -\frac{1}{\theta_1} \ (\Rightarrow |\theta_1| < 1)$$

In the previous example, we select the model with $\theta_1 = 0.2$.

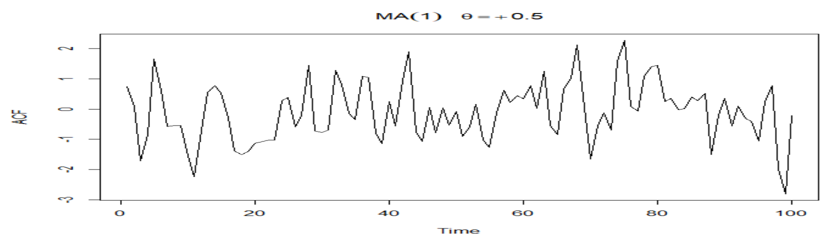
MA(1) Process: Simulations

Simulated Example: We simulate with R function *arima.sim* (& plot) three MA(1) processes, with standard normal ε_t -i.e., $\mu = 0$ & $\sigma = 1$:

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

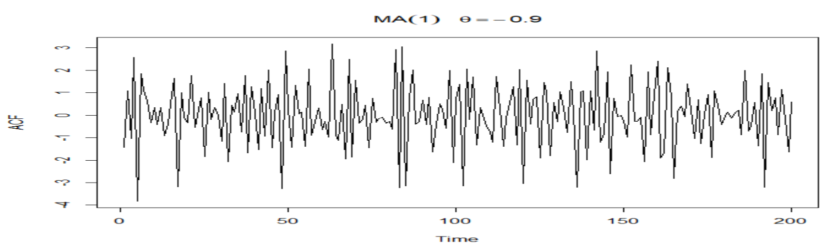
R script to plot $y_t = \varepsilon_t + \mathbf{0.5}\,\varepsilon_{t-1}$ with **200** simulations

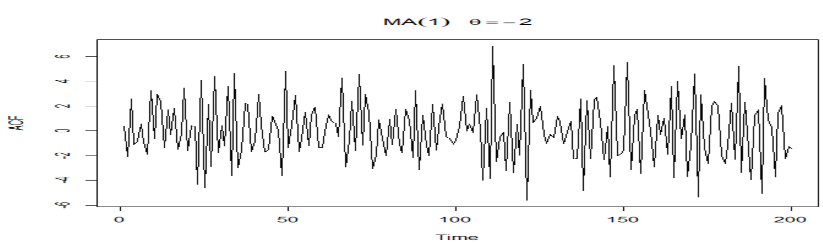
```
> plot(arima.sim(list(order=c(0,0,1), ma = 0.5), n = 200), ylab="ACF", main=(expression(MA(1)~~~theta==+.5)))
```



MA(1) Process: Simulations

Simulated Example (continuation):





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

MA(1) Process: Simulations (ACF)

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

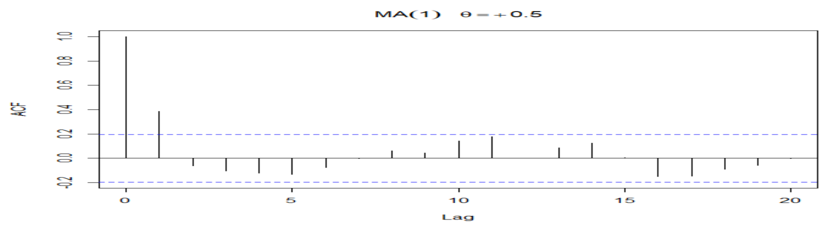
1) $y_t = \varepsilon_t + 0.5\, \varepsilon_{t-1}$

```
sim_ma1_5 <- arima.sim(list(order=c(0,0,1), ma = 0.5), n = 200)
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: Simulations (ACF)

Simulated Example (continuation):

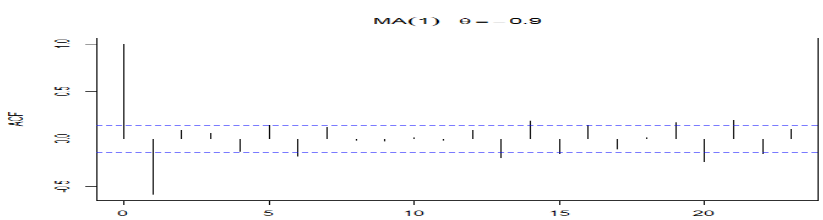
2) $y_t = \varepsilon_t - 0.9\, \varepsilon_{t-1}$

```
sim_ma1_9 <- arima.sim(list(order=c(0,0,1), ma = -0.9), n = 200)
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: Simulations (ACF)

Simulated Example (continuation):

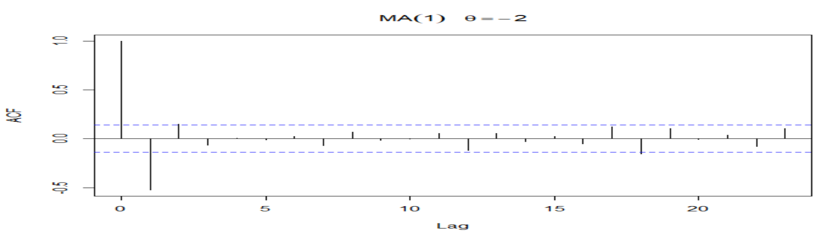
3) $y_t = \varepsilon_t$ - 2 ε_{t-1}

```
sim_ma1_2 <- arima.sim(list(order=c(0,0,1), ma = -2), n = 200)
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)

Simulated Example (continuation):

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

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

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

The simulated process with θ_1= -2 is non-invertible, the infinite sum of π_i would explode. We would select the MA(1) with θ_1 = -.5.

MA Process – Estimation

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

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

We cannot do OLS, since we do not observe ε_{t-1}. But, based on the ACF, we estimate θ_1.

• 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, θ_1:

$$r_1 = \frac{\hat{\theta}_1}{(1+\hat{\theta}_1^{\,2})} \quad \Rightarrow \quad \hat{\theta}_1 = \frac{1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

• A nonlinear solution and difficult to solve.

MA Process – Estimation

• Alternatively, if $|\theta_1| < 1$, 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

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

where $a^t = \theta_1^{\,t}$.

Autoregressive (AR) Process

• We model the conditional expectation of y_t, $E_t[y_t | I_{t-1}]$, as a function of its past history. We assume ε_t ~ WN(0, σ^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 $\qquad \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$

• We can look at an AR(p) process as a *stochastic (linear) difference equation* (SDE). We want to work with a stable y_t process (not explosive).

AR Process – AR(1): Stability

• We analyze the stability of an AR(p) process 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^{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** = "**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 mean, variance and auto-covariance function:

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

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

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

- We also derived the autocorrelations:

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

Remark: When $|\phi_1| < 1$, the autocorrelations do not explode as k increases. There is an exponential decay towards zero.

AR(1) Process – Stationarity & ACF

- ACF for an AR(1) process:

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

Then, the autocorrelogram –i.e., plot of $\rho(k)$ against k– shows

– when $0 < \phi_1 < 1$ ⇒ All autocorrelations are positive.

– when $-1 < \phi_1 < 0$ ⇒ The sign of $\rho(k)$ shows an alternating pattern beginning with a negative value.

– when $\phi_1 = 1$ ⇒ AR(1) is non-stationary, $\rho(k) = 1$, for all k. Present & past are always correlated!

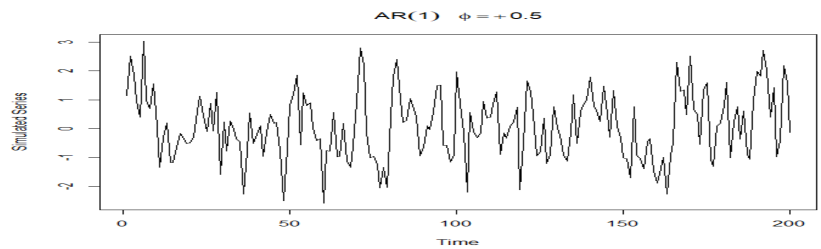
AR(1) Process – Stationarity & ACF: Simulations

Simulated Example: We simulate (& plot) three AR(1) processes, with standard normal ε_t -i.e., σ = 1:

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

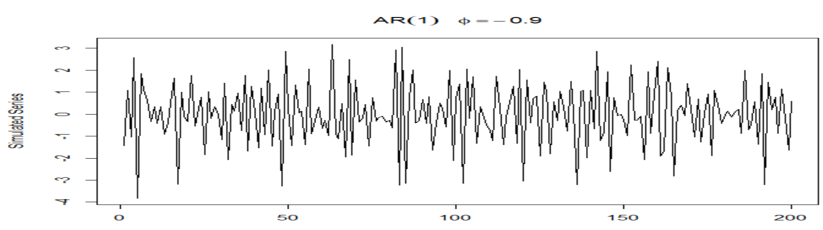
R script to plot $y_t = \mathbf{0.5}\, y_{t-1} + \varepsilon_t$ with **200** simulations

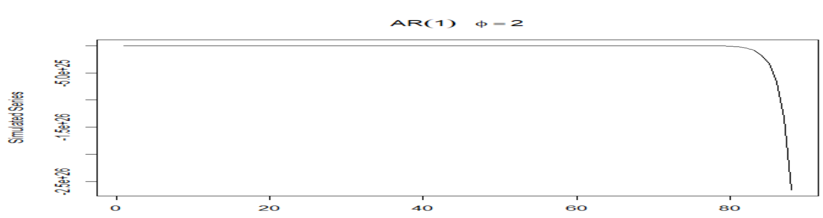
```
> plot(arima.sim(list(order=c(1,0,0), ar = 0.5), n = 200), ylab="ACF",
main=(expression(AR(1)~~~phi==+.5)))
```



AR(1) Process – Stationarity & ACF: Simulations

Simulated 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

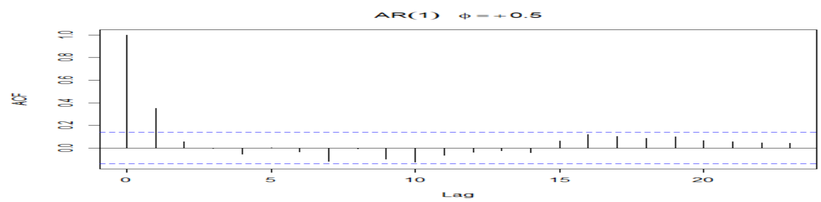
Simulated Example (continuation): Below, we compute and plot the ACF for the two stable simulated process.

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

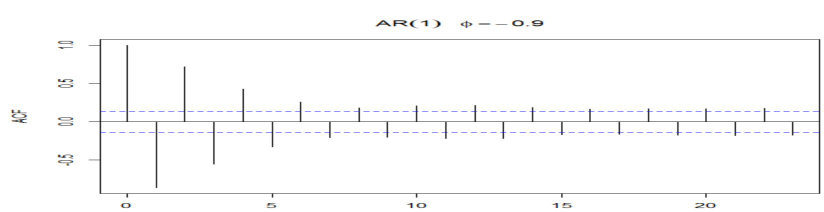
Simulated Example (continuation):

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

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

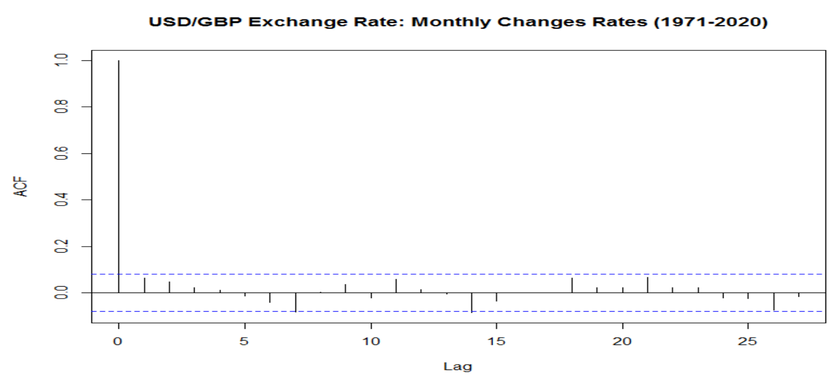
0	1	2	3	4	5	6	7	8	9	10	11	12	13
1.000	-0.584	0.093	0.061	-0.132	0.147	-0.181	0.122	-0.013	-0.023	0.014	-0.012	0.092	-0.199

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



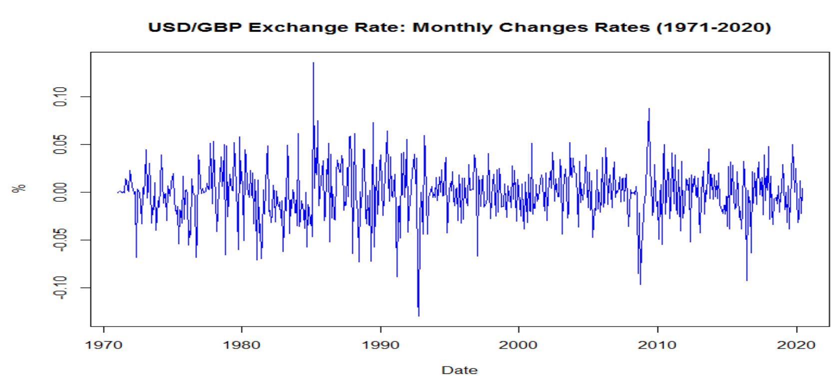
AR(1) Process – Stationarity & ACF: Examples

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:



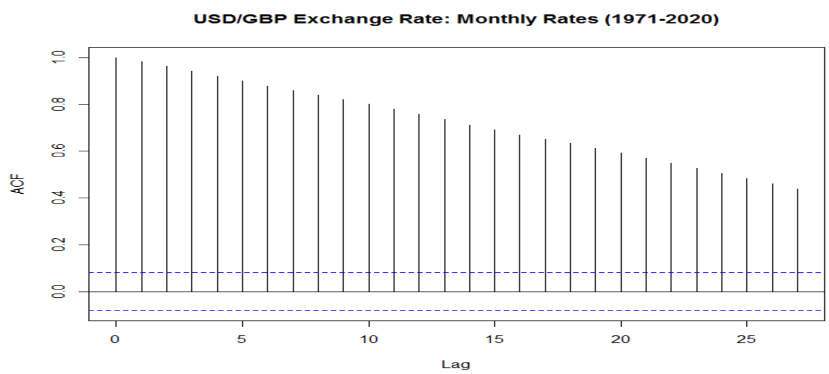
AR(1) Process – Stationarity & ACF: Examples

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



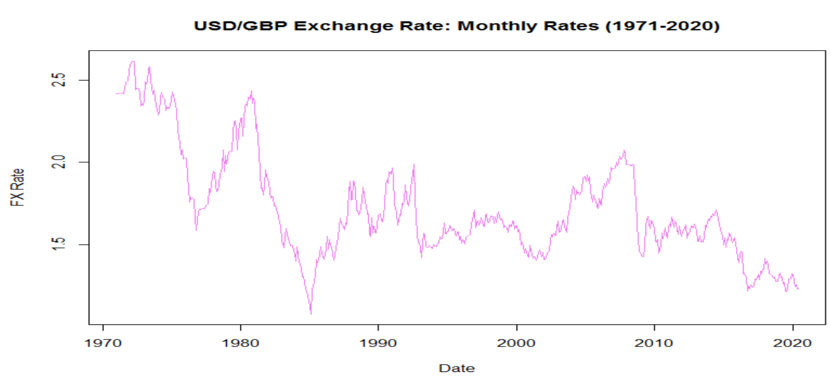
AR(1) Process – Stationarity & ACF: Examples

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: Examples

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



AR Process – Stationarity and Ergodicity

Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if the roots of $\phi(L)$ are $|z_j|$>1 for all j, where $|z_j|$ is the modulus of the complex number r_j.

Note: If one of the z_j's equals 1, ϕ(L) (& y_t) has a **unit root** –i.e., ϕ(1)=0. This is a special case of *non-stationarity*.

• Recall $\phi(L)^{-1}$ produces an infinite sum on the ε_{t-j}'s. If this sum does not explode, we say the process is **stable**.

• If the process is stable, the $\phi(L)$ polynomial can be inverted. It is possible to transform the AR(p) into an MA(∞). Then, we say the process y_t is **causal** (strictly speaking, a *causal function of* $\{\varepsilon_t\}$).

