# Introduction to ARMA Models

Richard Penny

StatsNZ ARIMA Course
2021

# Introduction

- ARMA models are a class of models for stationary time series.
- It is probably the most important class of time series models.
- There exist many extensions to ARMA models like ARIMA and GARCH for non-stationary series.
- To understand ARMA models we need to take the perspective of random variables and use some (very) basic probability theory.
- We start by repeating some basic definitions.

# Definitions

**Definition**

- A time series is a family of random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ indexed in time with $t_1 < t_2, \ldots < t_n$. We also write $\{X_t | t = t_1, t_2, \ldots t_n\}$. I will usually write this as $\{X_t\}$
- Observations of the time series are denoted by $x_{t_1},, x_{t_2}, \ldots, x_{t_n}$
- The value of $x_t$ is called the state of the time series at time $t$.

**Remark**

- I use capital letters for random variables and lowercase letters for realisations (observations) of random variables.
- Hence, $x_{t_1},, x_{t_2}, \ldots, x_{t_n}$ – $\{x_t\}$ – is the time series data while the time series is a probabilistic object.
- We introduce a new random variable for every time $t_i$

## Definitions

**Remember**

- Random variables are the mathematical model for processes that are not entirely predictable.
- We can assign probabilities to certain outcomes, e.g. $P(X_t \leq x) = p$ with $p \in [0, 1]$.
- Random variables can have an expectation $E(X_t)$ and variance $Var(X_t)$. These describe aspects of the probabilistic behaviour of $X_t$.
- When we observe a random variable, we collect data $x_t$.
- We cannot assign probabilities to $x_t$. Once the random variable has been observed the outcome is certain.
- For data only sample average and sample variance exist. These are different form expectation and variance of $X_t$.

# Definitions

**Time series are stochastic processes.**

**Definition**

A stochastic process is a collection of random variables

$$\{X_t | t = M\}$$

indexed by some index set M.

**Remark**

- Stochastic process are more general than univariate time series.
- $M$ can be a complicated set like $M = \mathbb{R}^3$.
- Also the state space can be a complicated object.

# Measures of dependency

We describe properties of random variables $X, Y$ by

- Expectation

$$\mathbb{E}(X) = \sum xP(X = x)(discrete),$$

$$\mathbb{E}(X) = \int xf_x(x)dx(continuous).$$

- Variance $Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$
- Covariance $Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)]$
- Correlation $Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(x)Var(Y)}}$

It is straight forward to generalise these concepts to time series.

# Measures of dependency

Expectation function

In a time series the expectation depends on time. We define the expectation function $\mu_X(t)$ for $X_t$

$$\mu_X(t) = \mathbb{E}(X_t)$$

# Autocovariance function

- The autocovariance function $\gamma_x(s,t)$ reports the covariance of observations at two points in time $s$ and $t$

$$\gamma_x(s,t) = Cov(X_s, X_t) = \mathbb{E}[(X_s - \mu_X(s))(X_t - \mu_x(t))]$$

- measures linear dependencies between observations.
- is positive if $X_t$. and $X_s$ tend to go into the same direction.
- is negative if $X_t$. and $X_s$ tend to go into opposite directions.
- typically takes larger values when $s$ and $t$ are close.
- is usually close or equal to 0 when $s$ and $t$ are farther apart.
- $\gamma_x(s,t) = \gamma_x(t,s)$
- $\gamma_x(t,t) = Var(X_t)$.

# Autocorrelation function (ACF)

Similar to the autocovariance function $\gamma_x(s,t)$, we define the autocorrelation function $\rho(s,t)$ for two points in time $s$ and $t$

$$\rho_X(s,t) = Corr(X_s, X_t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

- $\rho_x(s,t) = \rho_x(t,s)$
- $\rho_x(t,t) = 1$
- The ACF is always $-1 \leq \rho_x(s,t) \leq 1$.
- The ACF measures how perfect a linear dependency between observations in the series is.
- If $X_s = a + bX_t$, then $\rho_x(s,t) = \pm 1$.
- If no linear dependency exists, $\rho_x(s,t) = 0$.

Note

Expectation, autocovariance, and autocorrelation function are defined for random variables and not for data. There are sample counterparts, but they only work for stationary time series.

# Stationarity

A stationary time series does not change its probabilistic behaviour (too much) over time.

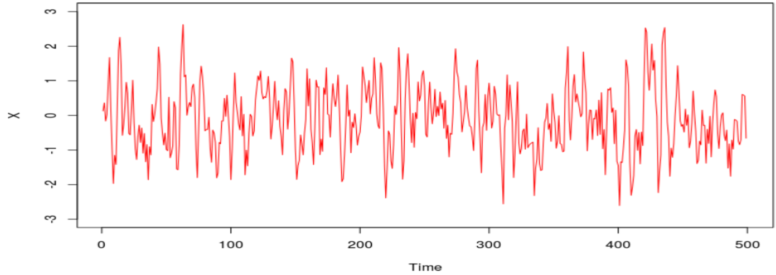
Definition

A time series $X_t$ is <u>strictly stationary</u> if for any collection of values $(X_1, X_2 \ldots, X_k)$ the joint probability distribution is the same as for the shifted values $(X_{1+h}, X_{2+h} \ldots, X_{k+h})$ for any shift $h$. That is

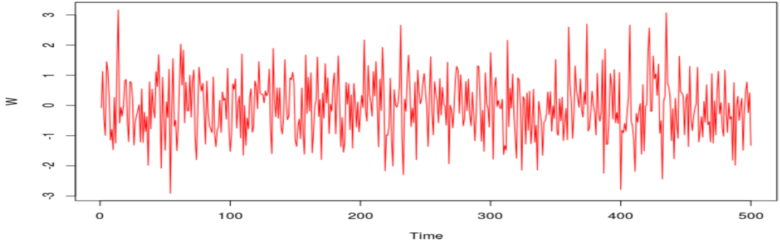
$$P(X_1 \leq c_1, \ldots, X_k \leq c_k) = P(X_{1+h} \leq c_1, \ldots, X_{k=h} \leq c_k)$$

for all $c_l, c_2, \ldots c_k \in \mathbb{R}$

# Two strictly stationary time series




Moving average $X_t = 1/\sqrt{3}(W_t + W_{t-1} + W_{t-2})$.




Strong Gaussian white noise $W_t$ with $\sigma_W = 1$.

# Strong and Weak Stationarity

One consequence of strong stationarity is that all observations have the same distribution. We can write $X_t \in X_1$

This is a very strong assumption and usually more than we need. We will use the concept of weak stationarity. A time series $X_t$ is weakly stationary if it has finite variance and

- $\mu_X(t)$ is the same for all $t$.
- $\gamma_x(t, t+h)$ only depends on $h$ but not on $t$.

Note

- All observations $X_t$ have the same mean and variance but not necessarily the same distribution.
- A strictly stationary series with finite variance is also weakly stationary but not the other way around.

# Properties of Weak Stationarity

I will write stationary instead of weakly stationary

- $\mu_x$ instead of $\mu_X(t)$
- $\gamma_X(h)$ instead of $\gamma_x(t, t+h)$
- $\rho_x(h)$ instead of $\rho_x(t, t+h)$

Properties

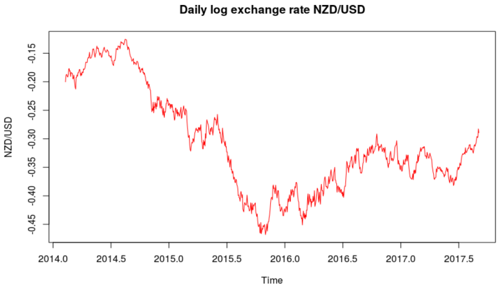
- The autocovariance and autocorrelation functions $\gamma_x(t, t+h)$ and $\rho_x(t, t+h)$ only depend on $h$ but not on $t$.
- $\gamma_x(h) = \gamma_x(-h)$
- $\rho_X(h) = \rho_X(-h)$

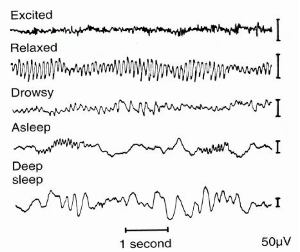
Stationarity is a very important concept in time series analysis. It only makes sense for random variables and not for data. Many time series methods are only valid for stationary time series.

# Stationary Series

Examples for approximately stationary time series are

- Differences of log stock prices, i.e. returns.
- Log FX rates.
- Electroencephalogram.




Non-stationary observations can often be transformed to approximately stationary series by detrending, log, or differencing.

## Building Blocks

ARMA models are build from three simpler type of models

1. White noise model
2. Auto regressive models
3. Moving average models

The acronym ARMA stands for Auto Regressive Moving Average model. I will cover these model step by step, then combine them

# White noise

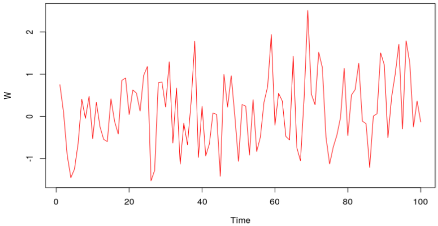
White noise is our model for measurement errors and uninformative noise and is a building block for several more complex models. Like stationarity there is weak and strong white noise.

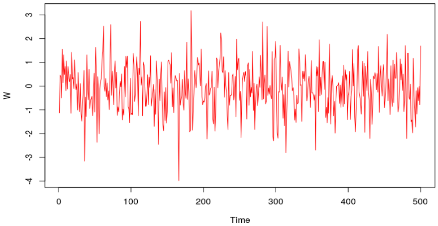
For weak white noise we have random variables $W_t, t = t_1, t_2, \dots, t_n$ with

- A mean of zero, i.e. $\mu_w(t) = \mathbb{E}(W_t) = 0 \; \forall t$.
- finite variance, i.e. $\gamma_w(t,t) = Var(W_t) < \infty \; \forall t$.
- homoscedasticity, i.e. all have the same variance. We write $Var(W_t) = \sigma_W^2$
- uncorrelated, i.e. $\rho_w(t,t') = Corr(W_t, W_{t'}) = 0 \; \forall t \neq t'$

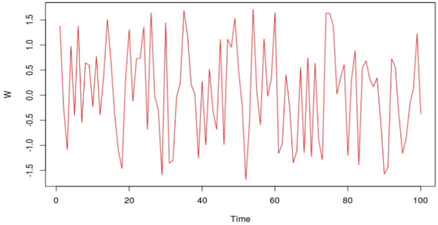
We write $W_t \in WN(0, \sigma_w^2)$. For strong white noise in addition, all $W_t$ are independent. For this course I assume weak white noise.

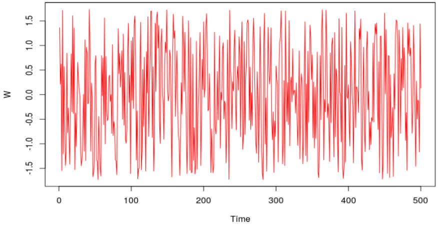
# White noise Examples




Strong Gaussian white noise $\sigma_W = 1$ , i.e. $W_t \sim \mathcal{N}(0, \sigma_W^2)$ i.i.d.
100 observations (left), 500 observations (right)




Strong uniform white noise $\sigma_W = 1$
100 observations (left), 500 observations (right)

# Autoregression

One objective of a time series model is to capture the dependency or correlation of present and past observations. A very successful model for correlations among random variables in general is, of course, linear regression. In time series we regress $X_t$ on the past values, hence **auto**regression.

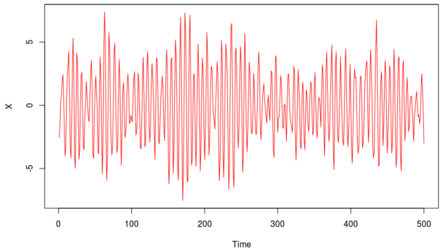
The Autoregressive model is

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdot + \phi_p X_{t-p} + W_t$$

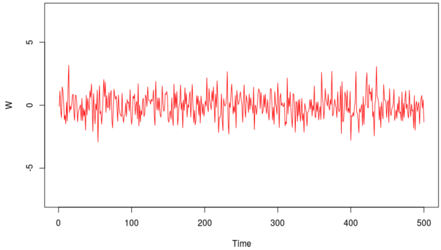
with $\phi_0, \phi_1, \phi_2, \ldots \phi_p \in \mathbb{R}$, $\phi_p \neq 0$, and $W_t$ white noise is an autoregressive model of order $p$, denoted AR(p).

The AR(1) model with $\phi_1 = 1$ ($X_t = \phi_0 + X_{t-1}$) is called random walk. It is an important model for stock prices but it is not stationary.

# AR example




Autoregression $X_t = X_{t-1} - 0.9X_{t-2} + W_t$.




Underlying strong Gaussian white noise $W_t$.

# Moving average

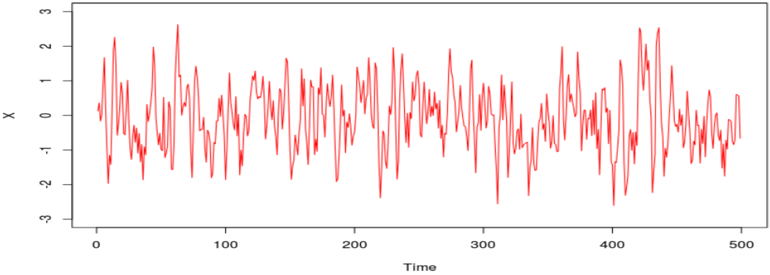
Sometimes we want the correlation to go back to 0 for large lags. This can be modelled by a moving average.

$$X_t = \mu + W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \ldots + \theta_q W_{t-q}$$

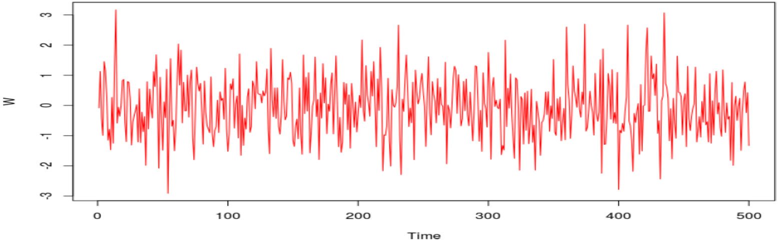
with $\mu, \theta_1, \theta_2, \ldots, \theta_q \in \mathbb{R}$, $\theta_q \neq 0$, and $W_t$ white noise – can be written as $\epsilon_t$ – is a moving average model of order $q$, denoted MA(q).
Concept

- $X_t$ and $X_s$ are correlated if and only if $|t - s| \leq q$.
- If $t \geq s \geq t - s$, then $X_t$ and $X_s$ are both correlated with $W_s, W_{s-1}, \ldots, W_{t-q}$.
- The current timepoint is just a weighted sum of recent errors. The sequence $(1, \theta_1, \ldots, \theta_q)$ determines the profile.
- The model is often stated without $\mu$. However, $\mu$ can been added to allow for a non-zero mean level.

# Two strictly stationary time series




Moving average $X_t = 1/\sqrt{3}(W_t + W_{t-1} + W_{t-2})$.




Strong Gaussian white noise $W_t$ with $\sigma_W = 1$.

# Wold decomposition theorem

Another reason for MA models is given by the Wold decomposition theorem.

**Theorem**

Every weakly stationary time series can be represented as a linear combination of a sequence of uncorrelated random variables

$$X_t = \mu + W_t + \psi_1 W_{t-1} + \psi_2 W_{t-2} + \ldots + \psi_q W_{t-q}$$

where the white noise $W_t \in WN(0, \sigma_W^2)$ has finite variance $\sigma_W^2 > 0$, and the coefficients are a convergent sum

$$\sum_{j=0}^{\infty} |\psi_j| < \infty.$$

# Consequence of the Wold decomposition

$$X_t = \mu + W_t + \psi_1 W_{t-1} + \psi_2 W_{t-2} + \dots$$

- Note that the sequence can be infinite. We can say that every stationary time series is an $MA(\infty)$ process.
- Almost all $|\psi_j|$ must be small and (preferably) only a small number can be large. Otherwise, the sum $\sum_{j=0}^{\infty} |\psi_j|$ would not converge.
- Hence, for every stationary $X_t$ there is an MA(q) with sufficiently large q and with suitable white noise, that provides a good approximation.
- This is a strong motivation to use MA models for stationary time series.

# ARMA Models

An AR(p) and an MA(q) model can be combined to an ARMA(p,q) model

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

or equivalently

$$X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$$

**Problem**

Although AR and MA are rather simple models it turns out to be quite tricky to combine them – $(p + 1 \times q + 1)$ space. Challenges are to choose the right p and q, and to fit the model to data – OLS works for AR but not for ARMA. We need to study some specific and some different properties of AR(p) and MA(q) to make this work.