Predictive Modeling

Mathematical Models for Time Series

Mirko Birbaumer

HSLU T&A

1. Introduction

Mathematical Models for Time Series

- In the last section, we have introduced
  - the concept of **time series** as observations that can be ordered chronically in a natural way
  - transformation, visualization and decomposition methods for time series
  - time series computations in `R` and `Python`
- Next step : **Modeling** of time series - a necessary step before we can **predict** future values of a series

Mathematical Concepts for Time Series

- Objectives of time series analysis :
  - exploratory data analysis, i.e. summary statistics and plots
  - to develop mathematical models that provide plausible descriptions for sample data
- Describing characteristics of data that seemingly fluctuate in a random fashion over time: we assume a time series can be considered as the **realization of random variables** that are **indexed over time**

Time Series and Discrete Stochastic Processes

Time Series and Discrete Stochastic Processes

Let T be a set of equidistant time points $T = \{t_1, t_2, \dots\}$.

1. A *discrete stochastic process* is a set of random variables $\{X_1, X_2, \dots\}$. Each single random variable X_i has a univariate distribution function F_i and can be observed at time t_i.
2. A *time series* $\{x_1, x_2, \dots\}$ is a realization of a discrete stochastic process $\{X_1, X_2, \dots\}$. In other words, the value x_i is a realization of the random variable X_i measured at time t_i.

Important distinction between a **time series**, i.e. a concrete observation of values, and a **discrete stochastic process** which is a theoretical construct to model the underlying mechanism that generates the values

Example: Random Walk

- A person starts walking from the coordinate center with constant speed in x-direction. At each step, however, the person decides **at random** either to walk 1m to the left or to the right.
- This is the simplest instance of a **random walk**
- Probabilistic model for this random walk
  1. Choose n independent Bernoulli random variables $D_1, \ldots, D_n$ that take on the values 1 and -1 with equal probability, i.e. $p = 0.5$
  2. Define the random variables $X_i = D_1 + \cdots + D_i$ for each $1 \leq i \leq n$. Then $\{X_1, X_2, \ldots\}$ is a discrete stochastic process modeling the random walk
- See example `1.1` of the `Mathematical Models for Time Series` chapter

Example : Random Walk

- From the definition of the process it is clear that the following recursive definition is equivalent

$$X_i = X_{i-1} + D_i, \quad X_0 = 0.$$

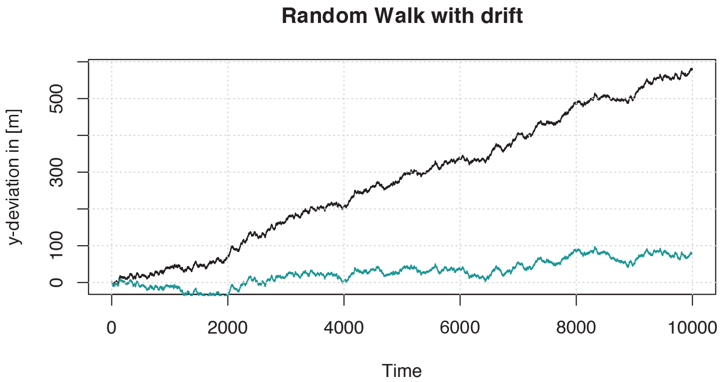
- If in each step a fixed constant δ is added to the series, i.e.

$$Y_i = \delta + Y_{i-1} + D_i, \quad Y_0 = 0.$$

  then we obtain a random walk **with drift**

- The random walk with drift models are used to model trends in time series data

- Random walk with drift (black) and random walk without drift (green):



- See example 1.1 of the Mathematical Models for Time Series chapter

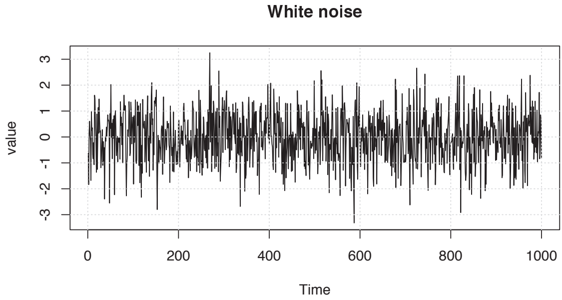
Stochastic Processes

- A time series
$$\{x_1, x_2, \ldots, x_n\}$$
can be understood as **one** realization of a multivariate random variable
$$\{X_1, X_2, \ldots, X_n\}$$

- Modeling and prediction for time series hence amounts to analyze a data set with *one* observation, which is impossible without further assumptions on the series.

- Example that lacks all these assumptions and thus is impredictible: **white noise**

- White noise plot:



- A white noise process consists of independent and identically distributed random variables $\{W_1, W_2, \ldots\}$ where each W_i has mean 0 and variance σ^2

- See example **1.2** of the **Mathematical Models for Time Series** chapter

- If in addition the individual random variables W_i are normally distributed $\quad \rightarrow \quad$ **Gaussian white noise**

- These models are used to describe **noise** in engineering applications

- The term **white** is chosen in analogy to white light and indicates that all possible periodic oscillations are present in a time series originating from a white noise process with equal strength.

- The observations of a white noise process are **uncorrelated** and could hence be treated with ordinary statistical methods.

Discrete Stochastic Processes Generated from White Noise Exhibiting Serial Correlation

- We apply a sliding window filter to the white noise process

$$\{W_1, W_2, \dots\}$$

- We obtain a **moving average** process

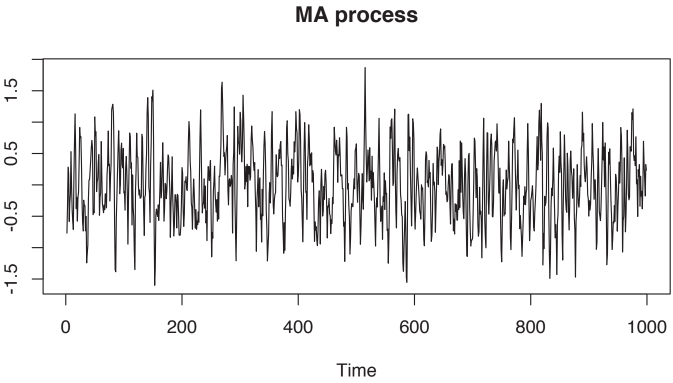
- If we choose the window length to be 3, we obtain

$$V_i = \frac{1}{3}(W_{i-1} + W_i + W_{i+1}).$$

- We choose

$$V_1 = W_1 \quad \text{and} \quad V_2 = 0.5(W_1 + W_2)$$

Moving Average Process



- The resulting process is **smoother**, i.e. the higher order oscillations are smoothed out
- See example `1.3` of the `Mathematical Models for Time Series` chapter

Autoregressive Time Series

- Many examples of real world time series, e.g. acoustic time series in speech analysis, contain dominant oscillatory components, producing a sinusoidal type of behaviour

- One possible model to generate such quasi-periodic data are **autoregressive series**.

Example: Autoregressive Time Series

- We consider again the white noise process

$$\{W_1, W_2, \dots, W_n\}$$

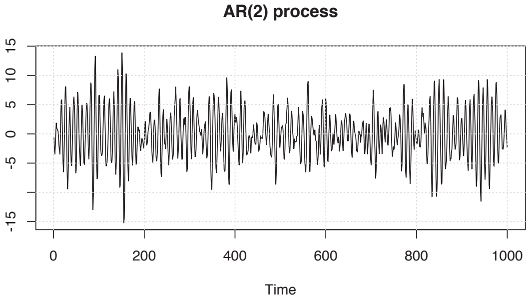
- We recursively define the following sequence:

$$X_i = 1.5X_{i-1} - 0.9X_{i-2} + W_i$$

- In other words, the value of the process at time instance i is modeled as a **linear combination** of the **past** two values plus some random component.

- Therefore this process is called **autoregressive**.

Autoregressive Model



- Realization of the autoregressive process

- The **oscillatory behaviour** becomes clearly visible

- Code : see example `1.4` of the `Mathematical Models for Time Series` chapter

Autoregressive Model

- Another interpretation of the autoregressive process above is via differential equations

- The finite difference scheme for the second order equation

$$\ddot{x} + 2\delta\dot{x} + \omega_0^2 x = W(t)$$

is given by

$$\frac{x_{i-2} - 2x_{i-1} + x_i}{\Delta t^2} + 2\delta\frac{x_{i-1} - x_i}{\Delta t} + \omega_0^2 x_i = W_i$$

- Here δ is the damping term and ω_0 the frequency of the homogeneous equation

- Setting $\Delta t = 1$, $\omega_0^2 = 0.4$ and $\delta = 0.05$ gives – after some rearrangements – the autoregressive process above

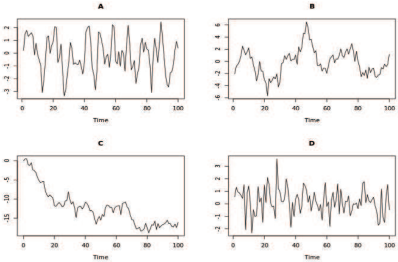
- Hence it can be seen as a **harmonic oscillator with random input**

- The wave length of the exact solution is

$$T = \frac{2\pi}{\omega_0} \approx 10.0$$

  which matches the situation in Figure above

Clicker Question



A time series from each of the four models we have considered in this section was simulated and they are shown in one of the four panels in the plot on the right. They include the white noise (WN), random walk (RW), autoregressive (AR), and simple moving average (MA) models. Match each time series plot with one of our models WN, RW, AR, MA.

- A) WN, (B) RW, (C) AR, (D) MA
- (A) MA, (B) AR, (C) RW, (D) WN
- (A) MA, (B) RW, (C) AR, (D) WN
- (A) RW, (B) MA, (C) WN, (D) AR

- In the definition of a discrete stochastic process $\{X_1, X_2, \dots\}$ we have postulated the existence of a distribution function $F_i(x)$ for each observation X_i in the process, i.e.

$$P(X_i \leq x) = F_i(x).$$

- Knowledge of individual distributions, however, is not sufficient to understand the **serial** behaviour of the process, because the observations are mutually **dependent**

Measures of Dependence

- Aside to the individual (also called **marginal**) distributions F_i of the random variables in the process we define **first** and **second order** moments to analyse the whole process.

- We start with the first order moments of the process, the **mean sequence**:

Mean sequence

The mean sequence $\{\mu(1), \mu(2), \dots\}$ (or mean function) of a discrete stochastic process $\{X_1, X_2, \dots\}$ is defined as the sequence of the means:

$$\mu(i) = \mathrm{E}[X_i].$$

Examples : Mean Sequence

- If W_i denotes a **white noise process**, then $\mathrm{E}[X_i] = 0$ for all $i \geq 1$.

- **Averaging** the values in the process does not change the mean and the mean sequence of a moving average process, and hence is also 0.

- If X_i is a **random walk with drift**, i.e. $X_0 = 0$ and

$$X_i = \delta + X_{i-1} + W_i$$

then we find that

$$\begin{aligned} \mathsf{E}[X_1] &= \delta + \mathsf{E}[X_0] + \mathsf{E}[W_1] = \delta \\ \mathsf{E}[X_2] &= \delta + \mathsf{E}[X_1] + \mathsf{E}[W_2] = 2\delta \\ \mathsf{E}[X_3] &= \delta + \mathsf{E}[X_2] + \mathsf{E}[W_3] = 3\delta \\ &\vdots \end{aligned}$$

- This means : $\mu(i) = i\delta$

Covariance

- As second order moments, we consider the covariance within a *single* process

Autocovariance and autocorrelation

Let $\{X_1, X_2, \dots\}$ be a discrete stochastic process.

1. The *autocovariance* γ_X is defined as

$$\gamma_X(i,j) = \mathrm{Cov}(X_i, X_j) = \mathrm{E}[(X_i - \mu(i))(X_j - \mu(j))].$$

2. The *autocorrelation* ρ_X is defined as

$$\rho_X(i,j) = \frac{\gamma_X(i,j)}{\sqrt{\gamma_X(i,i)\gamma_X(j,j)}}.$$

- For $i = j$ the autocovariance reduces to the **variance** of X_i
- Important property of autocovariance and autocorrelation: they are symmetric, i.e.

$$\gamma(i, j) = \gamma(j, i)$$

- The autocovariance measures the **linear dependence** of two points on the same process observed at different times
- If a series is very **smooth**, the autocovariance is **large**, even if i and j are far apart
- Note that if

$$\gamma(i, j) = 0$$

this only means that X_i and X_j are **not linearly dependent**, however, they still may depend in some nonlinear way

- The **autocorrelation** is the normalized version of autocovariance, i.e.

$$\rho(i,j) \in [-1,1]$$

- if there is a linear relationship between X_i and X_j, then

$$\rho(X_i, Y_j) = \pm 1$$

- To be more precise : if

$$X_i = \beta_0 + \beta_1 X_j$$

  the correlation will be 1 if $\beta_1 > 0$, and -1 else.

- Autocorrelation hence gives a rough measure how well the series at time i can be **forecast** by the value at time j

Example : Autocovariance and Autocorrelation

- A white noise process has the autocovariance function

$$\gamma(i,j) = \begin{cases} 0 & \text{if } i \neq j \\ \sigma^2 & \text{if } i = j \end{cases}$$

- Accordingly, the autocorrelation is 1 if $i = j$ and 0 else

- Autocovariance of the **three point moving average process**

- From the properties of the covariance it becomes clear that

$$\gamma(i,j) = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}\left(\frac{1}{3}(W_{i-1} + W_i + W_{i+1}), \frac{1}{3}(W_{j-1} + W_j + W_{j+1})\right)$$

- If $i = j$, then

$$\begin{aligned}
\mathrm{Cov}(X_i, X_i) &= \frac{1}{9}\mathrm{Cov}\left(W_{i-1} + W_i + W_{i+1}, W_{i-1} + W_i + W_{i+1}\right) \\
&= \frac{1}{9}\left(\mathrm{Cov}(W_{i-1}, W_{i-1}) + \mathrm{Cov}(W_i, W_i) + \mathrm{Cov}(W_{i+1}, W_{i+1})\right) \\
&= \frac{3\sigma^2}{9}
\end{aligned}$$

- This follows from the fact, that W_i, W_{i-1} and W_{i+1} are mutually uncorrelated

- Analogously for $i+1=j$:

$$
\begin{aligned}
\mathrm{Cov}(X_i, X_{i+1}) &= \frac{1}{9}\mathrm{Cov}\left(W_{i-1}+W_i+W_{i+1}, W_i+W_{i+1}+W_{i+2}\right) \\
&= \frac{1}{9}\left(\mathrm{Cov}(W_i, W_i)+\mathrm{Cov}(W_{i+1}, W_{i+1})\right) \\
&= \frac{2\sigma^2}{9}
\end{aligned}
$$

- To summarize, we find

$$
\gamma(i,j) = \begin{cases} \frac{3\sigma^2}{9} & \text{if } i=j \\ \frac{2\sigma^2}{9} & \text{if } |i-j|=1 \\ \frac{\sigma^2}{9} & \text{if } |i-j|=2 \\ 0 & \text{else} \end{cases}
$$

- Smoothing of the white noise introduces a **nontrivial autocovariance structure**
- Noteworthy : autocovariance only depends on the distance of the observations, but not on their value.

- Autocorrelation

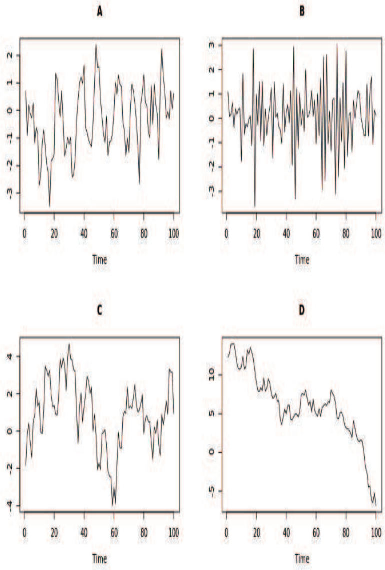
$$\rho(i,j) = \frac{\gamma(i,j)}{\sqrt{\gamma(i,i)\gamma(j,j)}} = \frac{\gamma(i,j)}{\gamma(i,i)}$$

- This gives

$$\rho(i,j) = \begin{cases} 1 & \text{falls } i = j \\ \frac{2}{3} & \text{if } |i-j| = 1 \\ \frac{1}{3} & \text{if } |i-j| = 2 \\ 0 & \text{else} \end{cases}$$

- Please solve exercises 1 and 2

Clicker Question



Autoregressive processes can exhibit varying levels of persistence as well as anti-persistence or oscillatory behavior. Persistence is defined by a high correlation between an observation and its lag, while anti-persistence is defined by a large amount of variation between an observation and its lag. The four plots on the left show varying degrees of persistence and anti-persistence. Which series exhibits the greatest persistence?

- A
- B
- C
- D

Stationarity

- A time series can be seen as a **single** realization of a multivariate random variable $\{X_1, X_2, \dots\}$ which we call in this special setting a stochastic process

- As we already know from elementary statistics, there is no way to do sound statistical analysis on a single observation

- New Concept of regularity that allows us to infer information from a single time series: $\rightarrow$ **stationarity**

Strict stationarity

Strict stationarity

A discrete stochastic process is called *strictly stationary* if for each finite collection $\{X_{i_1}, \dots, X_{i_n}\}$ and each lag $h \in \mathbb{Z}$ the shifted collection

$$\{X_{i_1+h}, \dots, X_{i_n+h}\}$$

has the same distribution. Put differently:

$$P(X_{i_1} \leq c_1, \dots, X_{i_n} \leq c_n) = P(X_{i_1+h} \leq c_1, \dots, X_{i_n+h} \leq c_n)$$

for any $c_1, \dots, c_n$.

In other words : the probabilistic character of the process does not change over time

Stationarity

- The definition of **strict stationarity** implies for the special case $n = 1$, i.e. the "collection" of time instances, that

$$P(X_i \leq c) = P(X_j \leq c)$$

for all i, j and for all c

- This amounts to say that the marginal distributions $F_i = F$ of the process coincide and in particular this implies that

$$\mu_X(i) = \mu_X(j)$$

- In other words, the **mean sequence is constant**

Stationarity

- We can consider the case $n = 2$
- Then the definition of strict stationarity implies that

$$P(X_i \leq c_1, X_j \leq c_2) = P(X_{i+h} \leq c_1, X_{j+h} \leq c_2).$$

- The joint distribution of each pair of time instances in the process depends only on the difference h and hence the autocovariance satisfies

$$\gamma(i, j) = \gamma(i + h, j + h).$$

Weak stationarity

- These two conclusions from strict stationarity are sufficient for most applications in order to come up with reasonable statistical models.

- Hence they give rise to the definition of a weaker form of stationarity

Weak stationarity

A discrete stochastic process X_i is called *weakly stationary* if

1. the mean sequence $\mu_X(i)$ is constant and does not depend on the time index i and
2. the autocovariance sequence $\gamma_X(i, j)$ depends on i and j only through their difference $|i - j|$.

Example : Weak Stationarity

- Reconsider the **three point moving average** process

- It is obvious that the mean function

$$\mu(i) = \mu = 0$$

  is constant

- The autocovariance only depends on the time lags

$$\gamma(h) = \begin{cases} \frac{3\sigma^2}{9} & \text{if } h = 0 \\ \frac{2\sigma^2}{9} & \text{if } |h| = 1 \\ \frac{\sigma^2}{9} & \text{if } |h| = 2 \\ 0 & \text{else.} \end{cases}$$

- Moving average process thus is **weakly stationary**

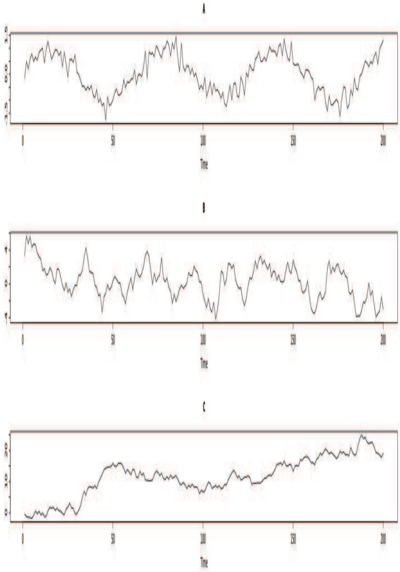
Testing Stationarity

- In practice, we are faced with time series, i.e. single observations of a discrete stochastic process.
- Stationarity is a property of stochastic processes
- **Example**: time series decomposition where trend and seasonality of a time series are estimated and subtracted which yields the so called **remainder sequence**
- Typical assumption for modeling the process is the (weak) stationarity of this remainder sequence
- So how to infer this from the data?

Testing Stationarity

- The first and simplest type of test is to **plot** the time series and look for evidence of trend in mean sequence or in the autocorrelation function
- If any such patterns are present then these are **signs of non-stationarity**
- A further possibility is to compute the mean and autocovariance sequences seperately for several windows and compare their behaviour
- When there is a dramatic change, then the hypothesis of stationarity can be **rejected**

Clicker Question



Which of the three time series plots on your right appear to be from a stationary model?

- A
- B
- C

Estimation of the Mean Sequence

- **Serial dependence measure**, such as **autocovariance**, are restricted to stochastic processes and hence make use of the probabilistic behaviour at each time instance

- In practice, however, we are confronted with a **single realization** of the process, i.e. at each time instance **only one** value is available

- Due to the **stationarity assumption** of the process we know that the mean sequence $\mu(k) = \mu$ is constant. A canonical estimator hence is

$$\widehat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Estimation of the autocovariance

The theorectical autocovariance function is estimated by the sample autocovariance sequence:

Sample autocovariance

1. The *sample autocovariance* is defined by

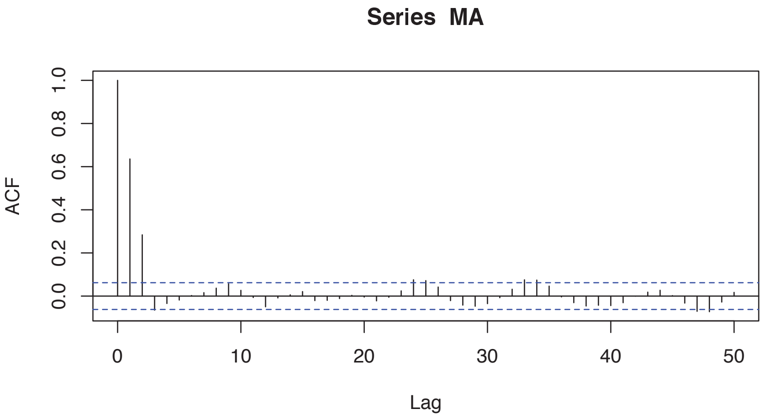
$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(x_{i+h} - \bar{x})(x_i - \bar{x})$$

with $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$

2. The *sample autocorrelation* is defined by

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}.$$

Repetition : Sample ACF of Simulated MA(5) Process



Repetition : ACF of White Noise

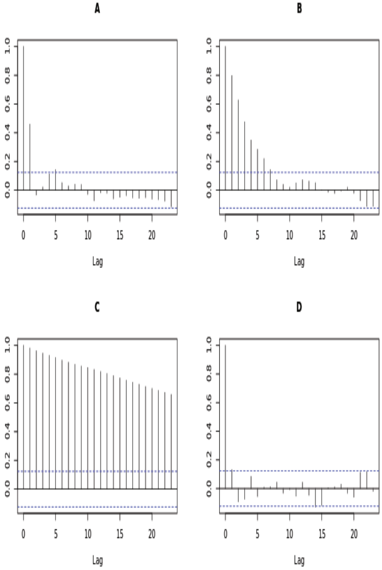
- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal with mean $-1/n$ and standard deviation

$$\sigma_{\widehat{\rho}} = \frac{1}{\sqrt{n}}.$$

- Used to check whether a time series stems from a white noise process or not: If the series is white, it should stay 95% of the time within the $-1/n \pm 2/\sqrt{n}$ limits.

- These limits are automatically drawn by `R` or `Pyton` when calling `acf` (**blue lines**)

Clicker Question



A time series from each of the four models you have considered in this course was simulated and their sample autocorrelation functions (ACF) are shown in one of the four panels in the adjoining figure. They include the white noise (WN), random walk (RW), autoregressive (AR), and simple moving average (MA) models. Match each sample ACF plot with one of our models WN, RW, AR, MA.

- (A) MA, (B) RW, (C) AR, (D) WN
- (A) RW, (B) MA, (C) WN, (D) AR
- (A) WN, (B) RW, (C) AR, (D) MA
- (A) MA, (B) AR, (C) RW, (D) WN