Moving Average Models

Jared Fisher

Lecture 5a

Announcements

- Homework 3 is due tomorrow/Wednesday March 3
- Project Checkpoint 2 is due Wednesday March 10. Specifics will be published later today. Bottom line: pursue stationarity two different ways (e.g. a polynomial+sinusoid model and a differencing approach)
- Midterm 1 grades will be done soon

Recap

Big Picture

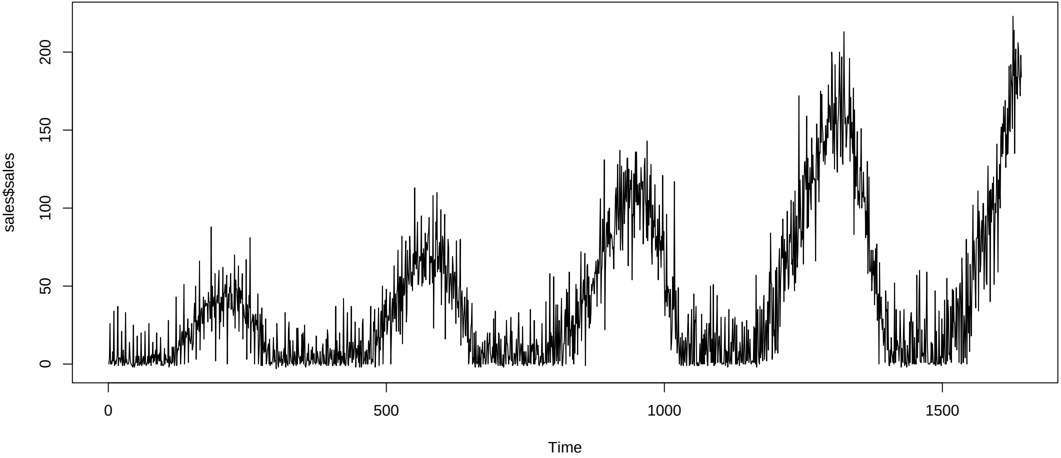
raw time series $\rightarrow$ stationary process $\rightarrow$ white noise

Pursuing Stationarity

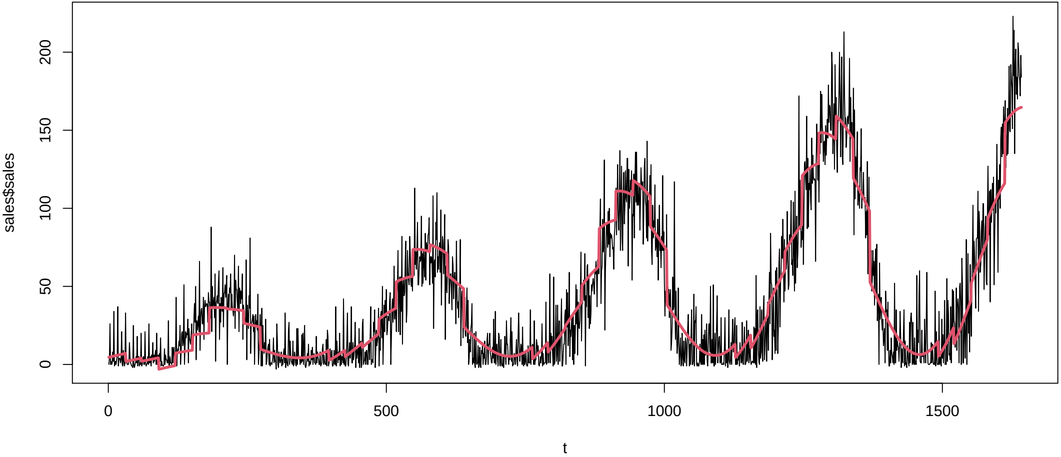
$$f(X_t) = m_t + s_t + X_t$$

- $f()$ Variance stabilizing transform (e.g. $\sqrt{x}$ or $log(x)$)
- m_t deterministic trend (e.g., approximately linear or quadratic)
- s_t deterministic periodic function of know period d, $s_{t+d} = s_t$
- X_t stationary process, e.g. white noise
- Idea: Remove both trend and seasonality so that what remains exhibit stable behavior over time (stability vs stationarity?)
- Instead of deterministic functions, we can also use filters like smoothing and differencing

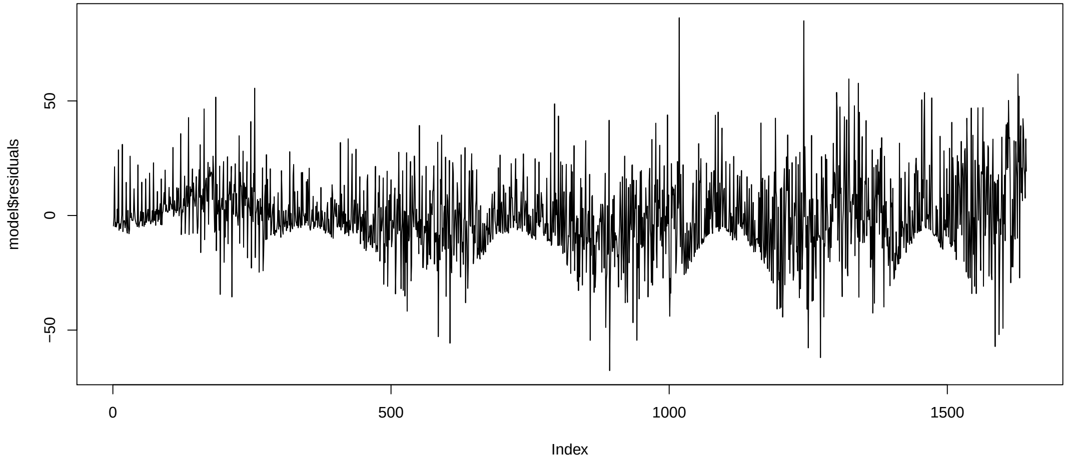
How this works in practice: back to sales data



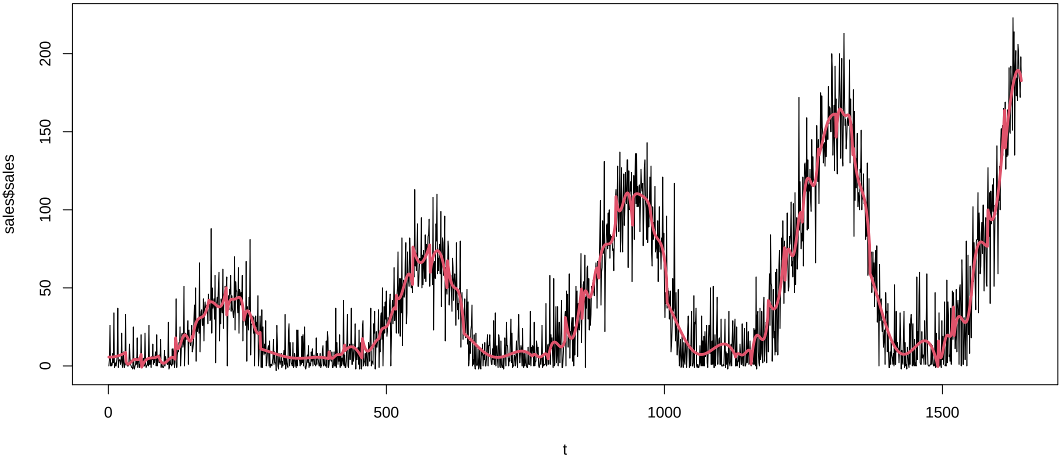
Model from previous lecture



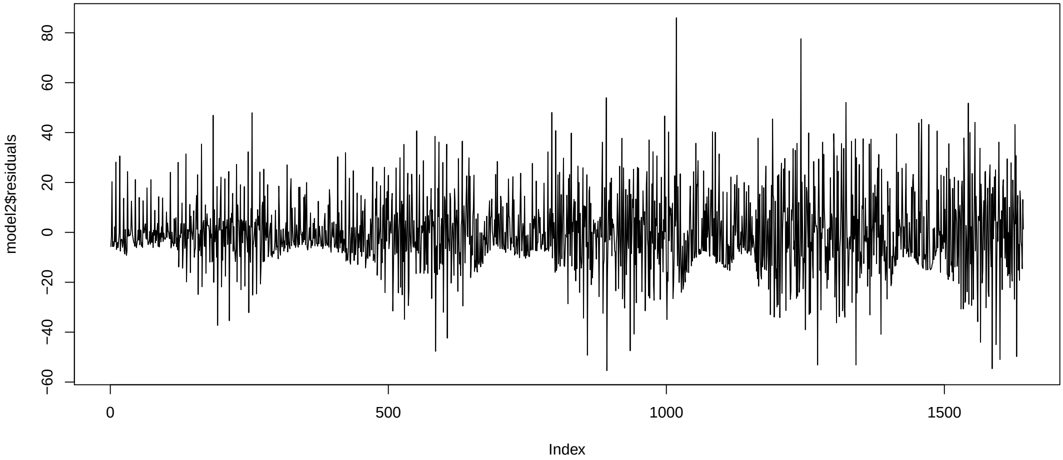
Have we reached stability?



Add more interactions

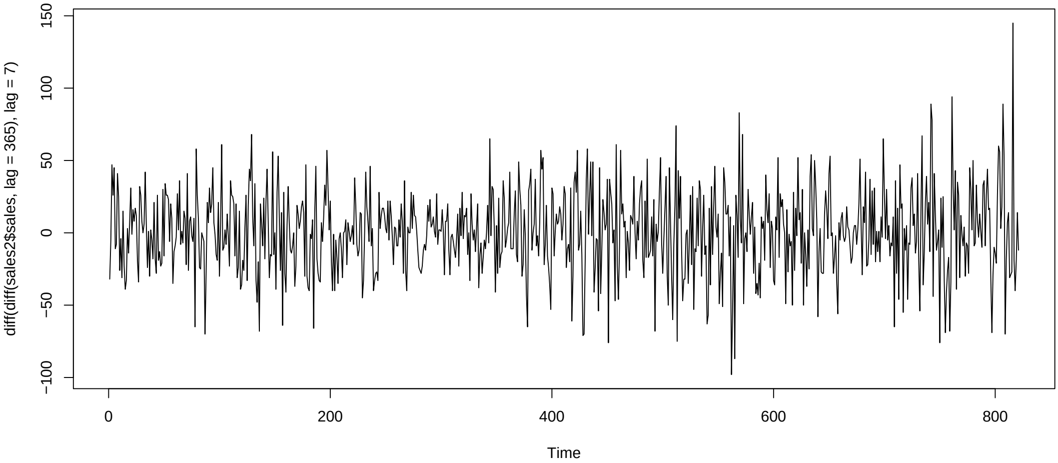


Are the residuals stable?

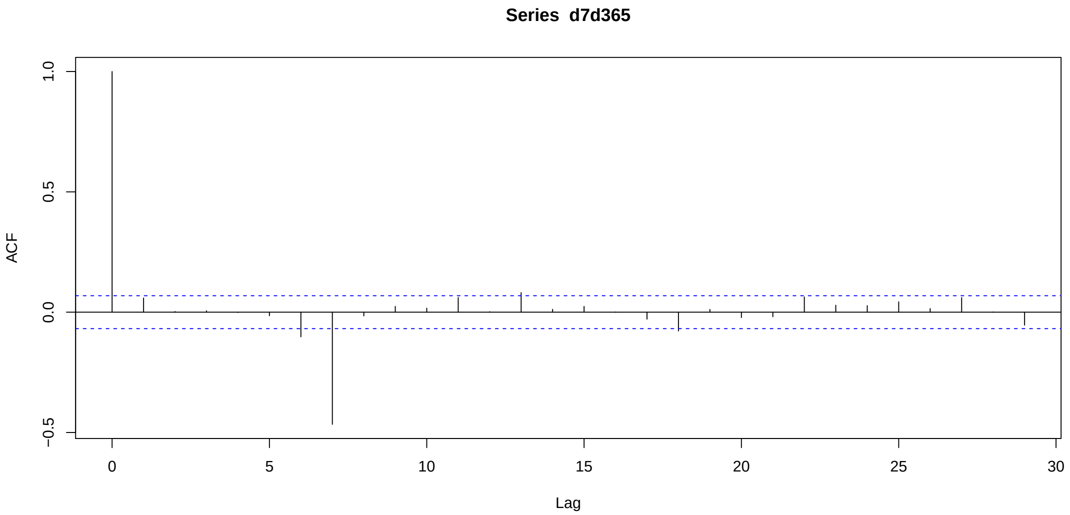


Differencing

```
sales2 = sales[450:nrow(sales),]
plot.ts(diff(diff(sales2$sales,lag=365),lag=7))
```



ACF



NEXT!

- We have pursued stationarity (and achieved stability, meaning approximately constant mean and variance)
- Now we can model the autocorrelation structure in this stable series!
- First we will discuss some theory of modeling stationary processes ($\sim$2+ weeks)
- Then we'll implement the ideas from theory into applied modeling

NEXT!

One way to think of this next step:

$$\text{stationary process} \rightarrow \text{white noise}$$

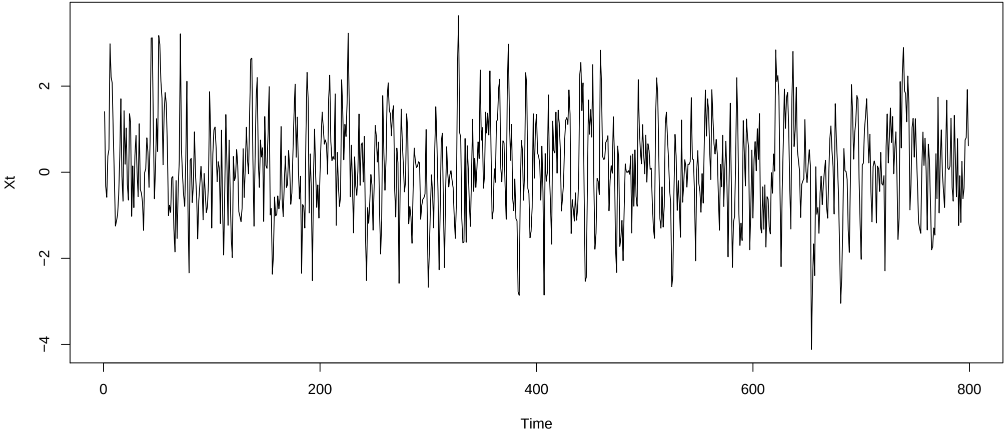
means

$$X_t = \sum_j a_j W_{t-j}$$

Moving Average models

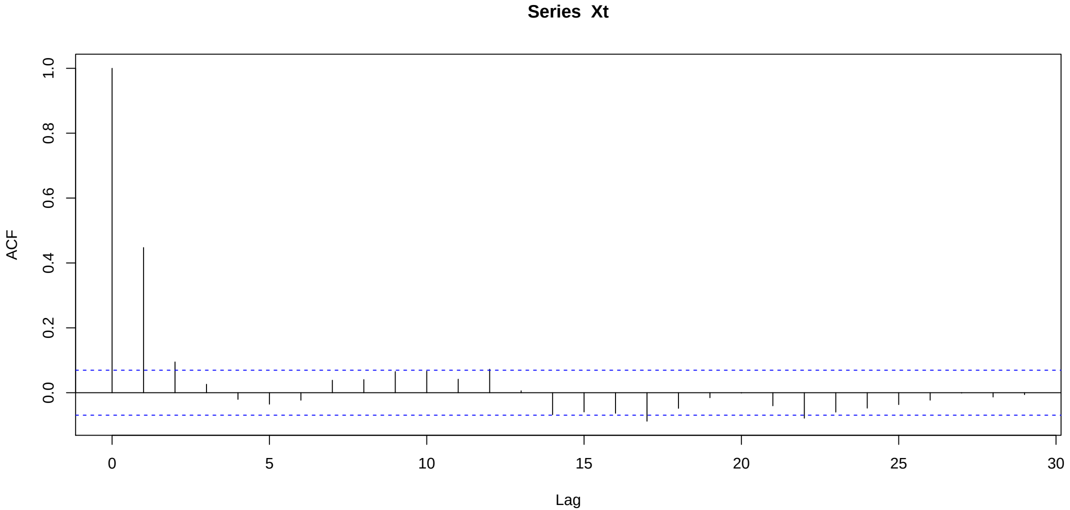
Motivating Example

- Does this look like White Noise? (does it look stable?)



- More or less, yes! But what does the Correlogram/ACF say?

Motivating Example



- Here, $X_t = Z_t + \frac{1}{2} Z_{t-1}$ and $Z_t \sim N(0,1)$ (so Gaussian noise!).

So what?

- Given a white noise series $\{W_t\}$ with variance σ^2 and a number $\theta \in R$, set

$$X_t = W_t + \theta W_{t-1}.$$

- This is called a **moving average** of order 1, or MA(1).
- What is the mean? Covariance? Is it stationary?
- Try by yourselves for a few minutes, then we'll derive together

Derivations (handwritten during lecture)

Moving Average Process of Order 1

The series is stationary with mean zero and auto-covariance function (ACVF)

$$\gamma_X(h) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \theta\sigma^2 & h = 1 \\ 0 & \text{otherwise} \end{cases}$$

As a consequence, X_s and X_t are uncorrelated whenever s and t are two or more time points apart. This time series has *short memory*.

MA(1)

- The autocorrelation function (ACF), for $\{X_t\}$ is given by

$$\rho_X(h) = \frac{\theta}{1+\theta^2}$$

  for $h = 1$ and 0 for $h > 1$.

- What is the maximum value that $\rho_X(1)$ can take?
- This is our first type of non-white-noise stationary process that we'll explore
- This gives us a tool for modeling noise that has autocorrelation

Definition

Let $\ldots, W_{-2}, W_{-1}, W_0, W_1, W_2, \ldots$ be a double infinite white noise sequence. The **moving average model** of order q or **MA(q)** model is defined as

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

where $\theta_1, \ldots, \theta_q$ are parameters, with $\theta_q \neq 0$.

Autocovariance function of an MA(q) time series:

- The MA(q) model can be concisely written as $X_t = \sum_{j=0}^{q} \theta_j W_{t-j}$ where we take $\theta_0 = 1$.
- The mean of X_t is clearly 0.
- For $h \geq 0$, the covariance between X_t and X_{t+h} is given by

$$\begin{aligned} \mathrm{cov}(X_t, X_{t+h}) &= \mathrm{cov}\left(\sum_{j=0}^{q} \theta_j W_{t-j}, \sum_{k=0}^{q} \theta_k W_{t+h-k}\right) \\ &= \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_j \theta_k \mathrm{cov}(W_{t-j}, W_{t+h-k}). \end{aligned}$$

Autocovariance function of an MA(q) time series:

- Note that because $\{W_t\}$ is white noise, the

$$cov(W_{t-j}, W_{t+h-k}) = \sigma^2 \neq 0$$

if and only if $t - j = t + h - k$ i.e., if and only if $k = j + h$.

- But because k has to lie between 0 and q, we must have that j has to lie between 0 and $q - h$.
- We thus get:

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h} & h = 0, 1, \ldots, q \\ 0 & if \text{h > q.} \end{cases}$$

Autocorrelation function of an MA(q) time series:

For the autocorrelation function we thus get

$$\rho_X(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{\sum_{j=0}^{q} \theta_j^2} & h = 0, 1, \ldots, q \\ 0 & h > q \end{cases}$$

Note that the autocovariance and the autocorrelation functions *cut off* after lag q.

Theorem: Stationarity of MA(q)

- Theorem: Let $\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$ be a time series which follows an MA(q) model. Then $\{X_t\}$ is weakly stationary.
- Why? Because the mean is always 0 and
- $\mathrm{cov}(X_t, X_{t+h})$ does not depend on t, only h.

Backshift Notation

Backshift Notation

- A convenient piece of notation avoids the trouble of writing huge expressions!
- Let B denote the **backshift operator** defined by

$$BX_t = X_{t-1}, B^2 X_t = X_{t-2}, B^3 X_t = X_{t-3}, \ldots$$

  and similarly

$$BW_t = W_{t-1}, B^2 W_t = W_{t-2}, B^3 W_t = W_{t-3}, \ldots.$$

- Also let I (or 1) denote the identity operator: $IX_t = X_t$.

Backshift Notation Examples

- Polynomial functions of the backshift operator:

$$\begin{aligned}(I + B + 3B^2)X_t &= IX_t + BX_t + 3B^2X_t \\ &= X_t + X_{t-1} + 3X_{t-2}\end{aligned}$$

- In general, for every polynomial $f(z)$, we can define $f(B)$.
- Negative powers of B correspond forward shifts.
  - $B^{-1}X_t = X_{t+1}$
  - $B^{-5}X_t = X_{t+5}$
  - $(B^3 + 9B^{-2})X_t = X_{t-3} + 9X_{t+2}$

Moving Average Operator

- The MA(1) process $X_t = W_t + \theta W_{t-1}$ can be written as

$$X_t = \theta(B) W_t$$

for the polynomial $\theta(z) = 1 + \theta_1 z$.

- Definition: for parameters $\theta_1, \dots, \theta_q$ with $\theta_q \neq 0$ define the **moving average operator** of order q as

$$\theta(B) = 1 + \theta_1 B + \dots \theta_q B^q.$$

- Then we can write the MA(q) model as

$$X_t = \theta(B) W_t,$$

for a white noise process $\{W_t\}$.

Moving Average Operator Example

- MA(2):

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2}$$

- Then we can write the MA(2) model as

$$X_t = (1 + \theta_1 B + \theta_2 B^2) W_t$$

- Such that $\theta(B) = (1 + \theta_1 B + \theta_2 B^2)$

Invertibility

Motivation

- Consider the case of the MA(1) model whose ACVF is given by

$$\begin{aligned}\gamma_X(0) &= \sigma_W^2(1+\theta^2)\\ \gamma_X(1) &= \theta\sigma_W^2\\ \gamma_X(h) &= 0 \text{ for all } h \geq 2.\end{aligned}$$

- Let's say $\theta = 5, \sigma_W^2 = 1$
- But we'd get the same ACVF as for $\theta = 1/5, \sigma_W^2 = 25$.
- In other words, there exist different parameter values that give the same ACVF.
- This implies that one **cannot uniquely** estimate the parameters of an MA(1) model from data.

Invertibility

$$X_t = W_t + \theta W_{t-1}$$

- A natural fix is to consider only those MA(1) for which $|\theta| < 1$:
- This condition is called **invertibility**.
- The condition $|\theta| < 1$ for the MA(1) model is equivalent to stating that the moving average polynomial $\theta(z) = 1 + \theta z$ has all roots of magnitude strictly larger than one.

Definition

An MA(q) model $X_t = \theta(B)W_t$ is said to be **invertible**, if $\theta(z) \neq 0$ for $|z| \leq 1$.

Alternate Definition via Theorem

An MA(q) model $X_t = \theta(B)W_t$ is invertible if and only if the time series $\{X_t\}$ and the white noise $\{W_t\}$ can be written as

$$W_t = \pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ and $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $\pi_0 = 1$.

$MA(\infty)$

Infinite Order Moving Average

- This is an $MA(\infty)$ model:

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

  with $\{W_t\}$ as white noise with mean zero and variance σ^2.

- We will write this expression succinctly via

$$X_t = \sum_{j=0}^{\infty} \theta_j W_{t-j}$$

  with θ_0 taken to be 1.

Infinite Order Moving Average

- Infinite sums have convergence issues
- A sufficient condition which ensures that the infinite sum is finite (almost surely) is $\sum_j |\theta_j| < \infty$.
- In this class, we will always assume this condition when talking about the infinite series $\sum_{j \geq 0} \theta_j W_{t-j}$.

Infinite Order Moving Average

It turns out that $X_t = \sum_{j=0}^{\infty} \theta_j W_{t-j}$ is a stationary process because

$$EX_t = E\left(\sum_{j=0}^{\infty} \theta_j W_{t-j}\right) = \sum_{j=0}^{\infty} \theta_j EW_{t-j} = 0$$

and

$$\begin{aligned} Cov(X_t, X_{t+h}) &= Cov\left(\sum_{j=0}^{\infty} \theta_j W_{t-j}, \sum_{k=0}^{\infty} \theta_k W_{t+h-k}\right) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j \theta_k Cov(W_{t-j}, W_{t+h-k}) = \sigma^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+h}. \end{aligned}$$

We could freely interchange the expectation and covariance operators above with the infinite sum because of the condition $\sum_j |\theta_j| < \infty$.

Infinite Order Moving Average

- Note that the expectation EX_t and the covariance $Cov(X_t, X_{t+h})$ do not depend on t and the autocovariance is given by

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+h}. \tag{1}$$

  In particular, we get the following

- Thoerem: Let $\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ be a time series which follows an MA(∞) model. Then $\{X_t\}$ is weakly stationary.

An Interesting $MA(\infty)$

- Fix ϕ with $|\phi| < 1$.
- Choose weights $\theta_j = \phi^j$ in $MA(\infty)$
- $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$
- ACVF:

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\phi^h \sigma^2}{1 - \phi^2} \text{for } h \geq 0$$

- ACF: $\rho(h) = \phi^h$ for $h \geq 0$.
- Unlike the MA(1), this ACF is strictly non-zero for all lags! But, since $\rho(h)$ drops exponentially as lag increases, this is effectively a stationary time series with short range dependence.
- Note that if ϕ is negative, the ACF $\rho(h)$ osillates as h increases.

An Interesting $MA(\infty)$

- Here is an important property of this process X_t:

$$\begin{aligned} X_t &= W_t + \phi W_{t-1} + \phi^2 W_{t-2} + \dots \\ &= W_t + \phi \left(W_{t-1} + \phi W_{t-2} + \phi^2 W_{t-3} + \dots \right) \\ &= W_t + \phi X_{t-1} \text{ for every } t = \dots, -1, 0, 1, \dots. \end{aligned}$$

- Thus X_t satisfies the following first order *difference equation*:

$$X_t = \phi X_{t-1} + W_t.$$

- For this reason, $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$ is called the **Stationary Autoregressive Process of order one**. More next time!