MA(q) | Properties and Characteristics

Moving Average (MA) Models

- Used to model stationary data
  - But not as useful as AR models for modeling actual stationary data
- They are useful in combination with the AR model to create the ARMA models, which are a useful extension of the AR models

Remember George Box's quote:

All models are wrong… but some are useful.

- We first briefly study the moving average (MA) model

Moving Average Model of Order *q*: MA(*q*)

$$\boxed{X_t = \quad + a_t \quad {}_1 a_{t \ 1} \ \ldots \quad {}_q a_{t \ q}}$$

Where the θ_j's are real constants and a_t's are white

Notes:

- X_t at time t is a linear combination of present and past noise components
- An MA(*q*) is a finite GLP and is always stationary
- GLP: $X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}$ so for an MA(q)

$$\psi_0 = 1, \psi_1 = -\theta_1, \ldots, \psi_q = -\theta_q, \psi_k = 0, k > q$$

Already in GLP form!

$$Example:\ MA(1):\ X_t = 4 + a_t + -.9a_{t-1} + 0a_{t-2} \ldots$$

Moving Average Model of Order *q*: MA(*q*)

$$\boxed{X_t = \quad + a_t \quad {}_1 a_{t \; 1} \; \ldots \quad {}_q a_{t \; q}}$$

Zero mean form

$$X_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Operator zero mean form

$$X_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t$$

MA-characteristic equation

$$1 - \theta_1 z - \cdots - \theta_q z^q = 0$$

Moving Average Model of Order 1: MA(1)

$$\boxed{X_t - \mu = a_t - \theta_1 a_{t-1}}$$

where θ_1 is a nonzero real, finite constant and a_t is white noise

Operator form

$$\boxed{X_t - \mu = (1 - \theta_1 B)a_t}$$

Zero mean operator form

$$\boxed{X_t = (1 \qquad {}_1 B)a_t}$$

MA(1) results: $\boxed{X_t - \mu = a_t - \theta_1 a_{t-1}}$

Mean

$$E(X_t) = \mu$$

Variance

$$\sigma_X^2 = \sigma_a^2(1+\theta_1^2)$$

Autocorrelations

$$\rho_0 = 1$$

$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$$

$$\rho_k = 0, k > 1$$

Spectral density

$$S_X(f) = \frac{\sigma_a^2}{\sigma_X^2}\left|1-\theta_1 e^{-2\pi i f}\right|^2$$

$$= \frac{\sigma_a^2}{\sigma_X^2}\left|1-\theta_1(\cos 2\pi f - i \sin 2\pi f)\right|^2$$

DataScience@SMU

Example

MA(1)

Example: $X_t = a_t - .99a_{t-1}$

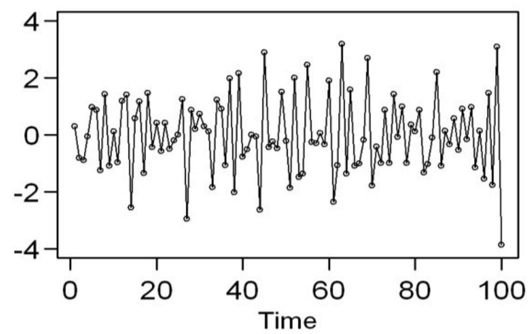
$\theta_1 = .99$ (watch signs!)

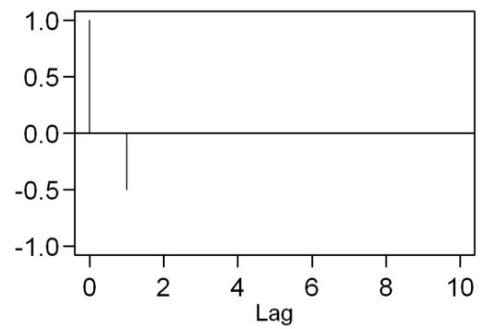
$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2} = \frac{-.99}{1+.99^2} = -.49997$$

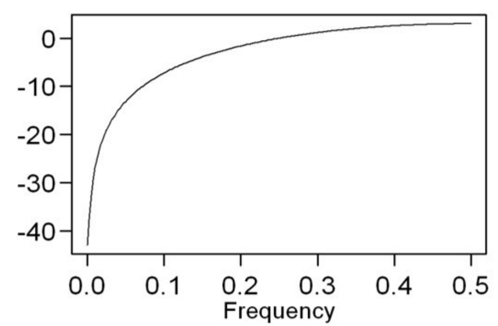
$$\rho_k = 0, k > 1$$

Note: For an MA(1), max $|\rho_1| = .5$.

Note: MA(1) spectral densities do not have "peaks." Instead, they have "dips."







Example: $X_t = a_t + .99a_{t-1}$

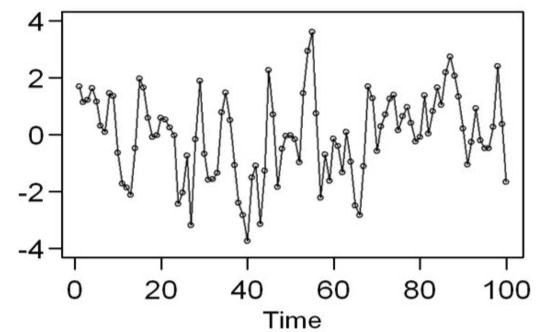
$$\theta_1 = -.99\ (watch\ signs!)$$

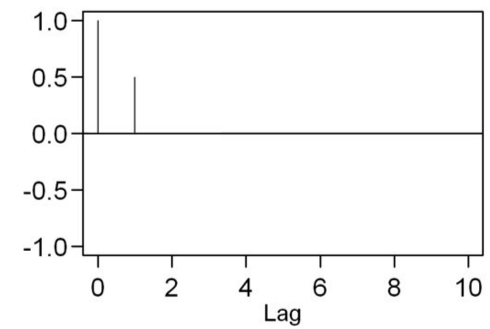
$$\rho_1 = \frac{-\theta_1}{1+\theta_1^2} = \frac{.9_1}{1+.9^2} = .49997$$

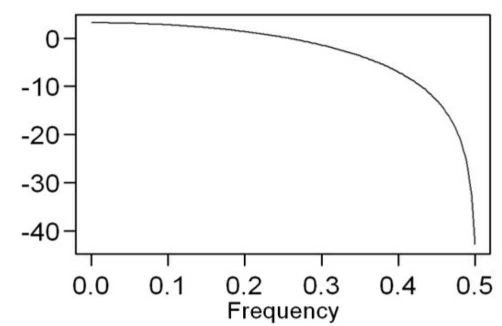
$$\rho_k = 0, k > 1$$

Note: For an MA(1), max $|\rho_1| = .5$.

Note: MA(1) spectral densities do not have "peaks." Instead they have "dips."







DataScience@SMU

Example

MA(2)

Next, We Briefly Consider the MA(2) Model:

$$\boxed{X_t \quad = a_t \quad {}_1 a_{t\ 1} \quad {}_2 a_{t\ 2}}$$

Autocorrelations

$$_0 = 1 \qquad _1 = \frac{_1 + \ _1\ _2}{1 + \ _1^2 + \ _2^2} \qquad _2 = \frac{_2}{1 + \ _1^2 + \ _2^2} \qquad _k = 0, k > 2$$

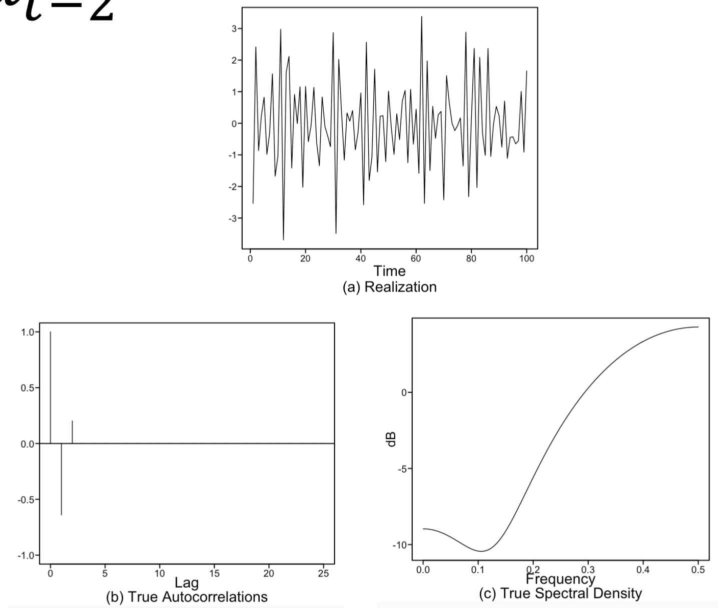
Example: $X_t = a_t - .9a_{t-1} + .4a_{t-2}$

$\rho_0 = 1$

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} = \frac{-.9 + (.9)(-.4)}{1 + .9^2 + (-.4)^2} = -.6396$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} = \frac{.4}{1 + .9^2 + (-.4)^2} = .2030$$

$\rho_k = 0, k > 2$



Key Feature of MA(q)

For an MA(q), $\rho_{\mathrm{k}} = 0, k > q$

Other notes about MA(q) processes

- As previously mentioned, MA processes are not appropriate for modeling many real-world processes.
  - For example, the property of an MA(2) that X_t is uncorrelated with $X_{t+3,}$ $X_{t+4,}$ … is unrealistic in most cases.
- **MA behavior is most useful in the formulation of ARMA(*p,q*) models.**

DataScience@SMU

MA(q) | tswge and MA Models

tswge demo

The following command generates and plots a realization from an MA, AR, or ARMA model.

`gen.arma.wge(n,phi,theta,vara,sn)`

Notes:

- For MA models, theta is a constant for q = 1 and a vector for $q > 1$. For an MA(q), phi takes on its default value of 0.
- + `sn=0` (default) generates a new (randomly obtained) realization each time. Setting sn > 0 allows you to generate the same realization each time you apply the command with the same `sn`.

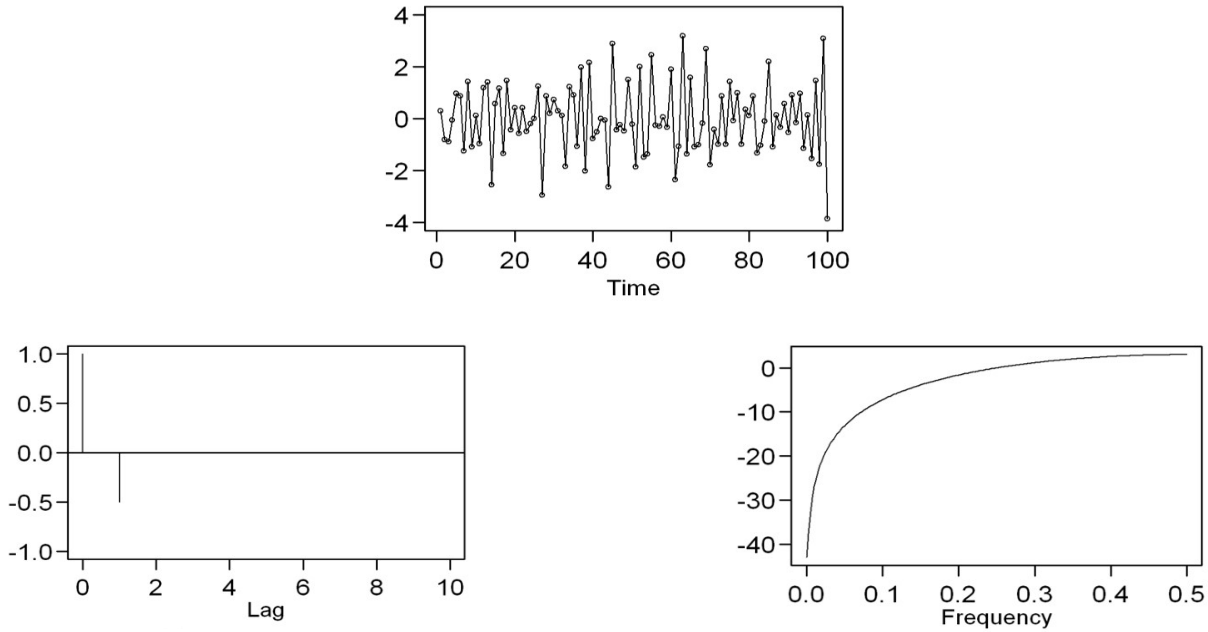
```
gen.arma.wge(n=100,theta=-.99)
gen.arma.wge(n=100,theta=-.99)
gen.arma.wge(n=100,theta=-.99,sn=5)
gen.arma.wge(n=100,theta=-.99,sn=5)
gen.arma.wge(n=100,theta=c(.9,-.4))
```

tswge demo

The following command plots a realization of length n (default $= 100$), the true autocorrelations, and the spectral density for an MA, AR, or ARMA model.

```
plotts.true.wge(n,phi,theta,lag.max)
```

Example:
```
plotts.true.wge(theta=c(.99))
plotts.true.wge(theta=c(.9,-.4))
```



DataScience@SMU

MA(q) | tswge

MA and Dip in Spectral Density

Demo: Dip in Spectral Density

Consider the MA(2) model:

$X_t = a_t \quad 1.1a_{t\;\;1} + .9a_{t\;\;2}$

First, recall the spectral density of

$X_t \quad 1.1X_{t\;\;1} + .9X_{t\;\;2} = a_t$

```
plotts.true.wge(phi=c(1.1,-.9))
```

Now, what does the spectral density of

$X_t = a_t \quad 1.1a_{t\;\;1} + .9a_{t\;\;2}$

look like?

```
plotts.true.wge(theta=c(1.1,-.9))
```

DataScience@SMU

MA(q) Comments

Key Feature of MA(q)

For an MA(q), $\rho_{\mathrm{k}} = 0,\ k > q$

Other notes about MA(q) processes

- As previously mentioned, MA processes are not appropriate for modeling many real-world processes.
  - For example, the property of an MA(2) that X_t is uncorrelated with $X_{t+3,}\ X_{t+4,}$,… is unrealistic in most cases.
- **MA behavior is most useful in the formulation of ARMA(p,q) models.**

DataScience@SMU

MA(q) | Invertibility

Invertibility

Light board?

Consider the MA(1) processes:

(a) $X_t = a_t - .8a_{t-1}$ (b) $X_t = a_t - 1.25a_{t-1}$

For (a), $\rho_1(\text{a}) = \dfrac{-\theta_1}{1+\theta_1^2} = \dfrac{-.8}{1.64}$

For (b), $\rho_1(\text{b}) = \dfrac{-\theta_1}{1+\theta_1^2} = \dfrac{-1.25}{1+1.25^2} = \dfrac{-1/.8}{1+(1/.8)^2}$

$$= \left(\frac{-1/.8}{1+(1/.8)^2}\right)\frac{.8^2}{.8^2}$$

$$= \frac{-.8}{.8^2+1}$$

$$= \rho_1(\text{a})$$

DataScience@SMU

Invertibility

- The previous example shows that two different MA(1) models can have the same autocorrelations
  - ρ_1 is the same for both models, and in each case, $\rho_0 = 1$ and $\rho_k = 0, k > 1$
- Undesirable situation
  - Called model multiplicity
- By restricting our models to ***invertible models,*** we can avoid ***model multiplicity***

Which MA Models Are Invertible

An MA(q) model is invertible if and only if the roots of the MA-characteristic equation, $\theta(z) = 0$, are greater than 1 in absolute value.

For the previous example:

Light board?

Model (a): $1 - .8z = 0$ has root $1/.8 = 1.25>1$

Model (b): $1 - 1.25z = 0$ has root $1/1.25 = .8 < 1$

Results:

- **Both** models are stationary (all MA models are).
- Model (a) is invertible.
- We do not consider non-invertible models such as model (b).

DataScience@SMU

MA(q) | tswge and Invertibility

tswge demo

Check for Invertibility

Use **`factor.wge`**

Example 1: Is $X_t = a_t \quad 1.6a_{t\ \ 1} + .9a_{t\text{-}2}$ invertible?

```
factor.wge(phi=c(1.6,-.9))
```

Example 2: Is $X_t = a_t \quad 1.6a_{t\ \ 1} \quad .9a_{t\text{-}2}$ invertible?

```
factor.wge(phi=c(1.6,.9))
```

Note: The frequency f_0 shown in the factor table is **not** a system frequency.

- In fact, it is a frequency at which there is a "dip" in the spectrum instead of a peak.

tswge demo

1. $X_t = a_t - .9a_{t-1}$

2. $X_t = a_t - 1.111a_{t-1}$

Note: 1/.9 = 1.1111….

Use `plotts.true.wge` to compare the characteristics of these two models.

DataScience@SMU

ARMA(p,q) | Properties and Characteristics

Autoregressive: Moving Average Process of Orders *p* and *q* ARMA(*p*,*q*)

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = \beta + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Where

- $\beta = (1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p)\mu$
- a_t is white noise
- φ_i's and θ_j's are real constants
- $\varphi_p \neq 0$ and $\theta_q \neq 0$
- $\varphi(z)$ and $\theta(z)$ have no common factors

Zero mean form

We will typically use the “zero mean” form of the model.

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Operator notation

$$(1 - \varphi_1 B - \cdots - \varphi_p B^p) X_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t$$

or

$$\varphi(B) X_t = \theta(B) a_t$$

DataScience@SMU

ARMA(p,q) | Cancellation

ARMA(p,q) Model $\varphi(B)X_t = \theta(B)a_t$

- $\varphi(B)$ and $\theta(B)$ introduce characteristics of the same type as in the AR(p) and MA(q) cases.
- "Near cancellation" may "hide" some characteristics.

Light board

Example: Cancellation

Light board

$$(1-1.3B+.4B^2)X_t=(1-.8B)a_t$$

What model is this? ARMA(2,1)?

k	ρ_k	
0	1	1
1	.5	$.5^1$
2	.25	$.5^2$
3	.125	$.5^3$
$\vdots$	$\vdots$	$\vdots$

That is, $\rho^k = .5^k$,

which are the autocorrelations for the AR(1) model

$(1-.5B)X_t = a_t$.

Note:

$(1-1.3B+.4B^2)X_t=(1-.8B)a_t$. Factoring, you get

$(1-.8B)(1-.5B)X_t=(1-.8B)a_t$. By cancellation, we get

$(1-.5B)X_t=a_t$ (i.e., the model above is an AR(1)).

DataScience@SMU

ARMA(p,q) | Cancellation

Example: Cancellation

$$(1-1.3B+.4B^2)X_t=(1-.8B)a_t$$

What model is this? ARMA(2,1)?

k	ρ_k	
0	1	1
1	.5	$.5^1$
2	.25	$.5^2$
3	.125	$.5^3$
:	:	:

That is, $\rho^k = .5^k$,

which are the autocorrelations for the AR(1) model

$(1-.5B)X_t = a_t$.

Note:

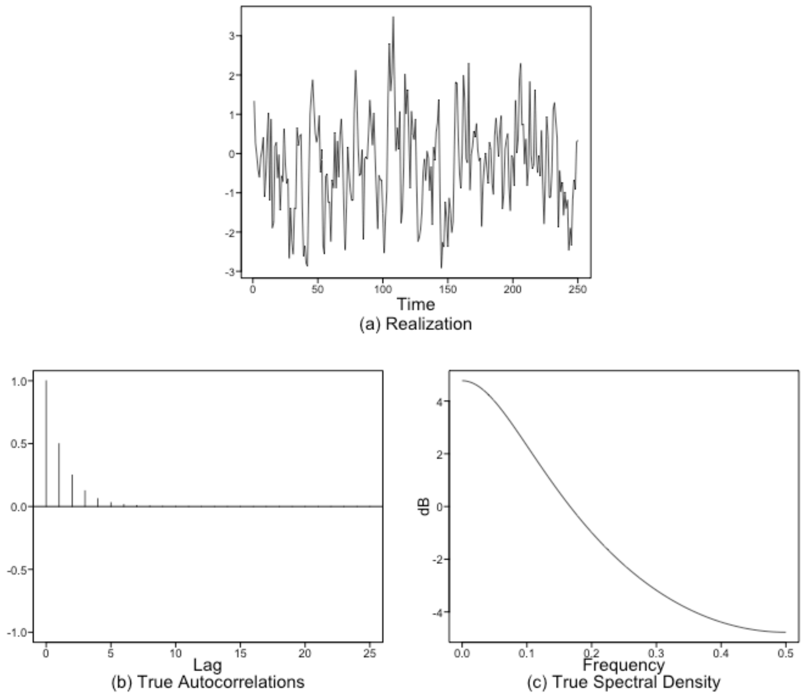
$(1-1.3B+.4B^2)X_t=(1-.8B)a_t$. Factoring, you get

$(1-.8B)(1-.5B)X_t=(1-.8B)a_t$. By cancellation, we get

$(1-.5B)X_t=a_t$ (i.e., the model above is an AR(1)).

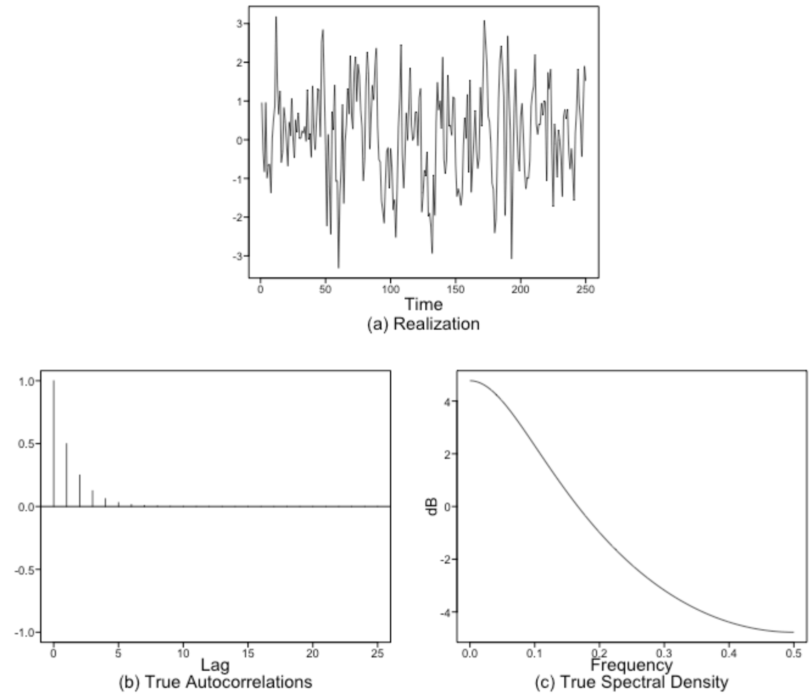
Cancellation

$X_t - 1.3X_{t-1} + .4X_{t-2} = a_t - .8a_{t-1}$



plotts.true.wge(250,phi = c(1.3,-.4),theta = c(.8))

$X_t - .5X_{t-1} = a_t$



plotts.true.wge(250,phi = c(.5))

The above two models are equivalent due to cancelation of the MA term and a factor in the AR term.

ARMA(*p*,*q*) | Blend of AR and MA

ARMA(p,q) Model $\varphi(B)X_t = \theta(B)a_t$

- $\varphi(B)$ and $\theta(B)$ introduce characteristics of the same type as in the AR(p) and MA(q) cases.

Key result:

$\varphi(B)X_t = \theta(B)a_t$ is **a stationary and invertible** ARMA(p,q) process iff:

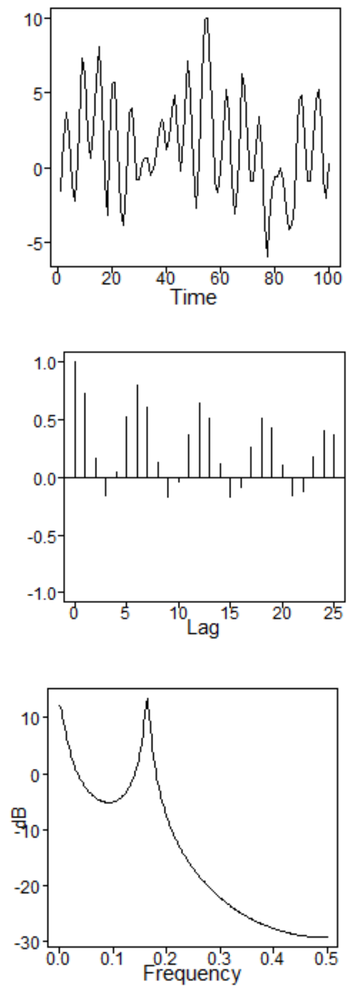
i. Roots of $\varphi(z) = 0$ are all outside the unit circle

ii. Roots of $\theta(z) = 0$ are all outside the unit circle

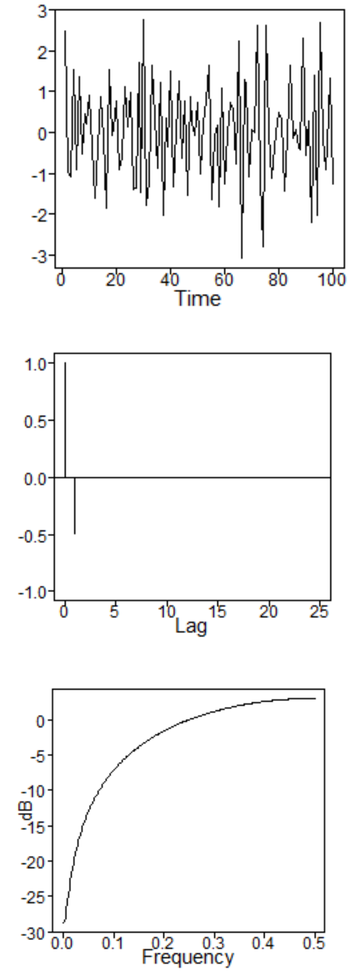
Note: In order to refer to a process as "ARMA," we will require it to be **stationary and invertible**.

Example $(1 \quad 1.95B + 1.9B^2 \quad .9025B^3)X_t = (1 \quad .95B)a_t$

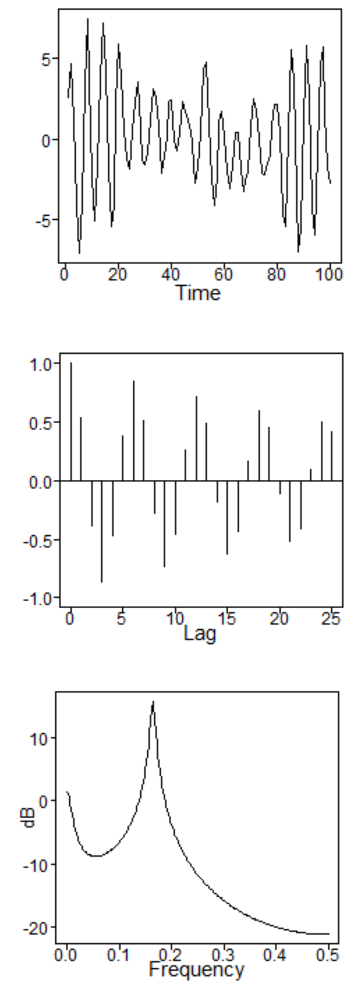
$(1 - 1.95B + 1.9B^2 - .9025B^3)X_t = a_t$



$X_t = (1 - .95B)\, a_t$



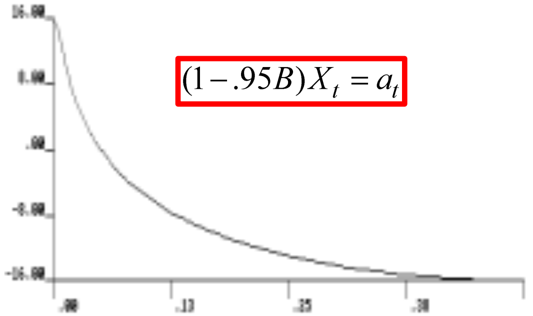
$(1 - 1.95B + 1.9B^2 - .9025B^3)X_t = (1 - .95B)a_t$

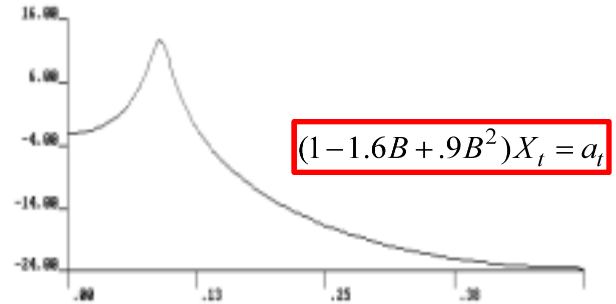


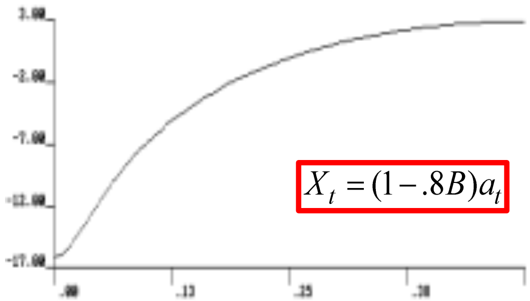
Notes

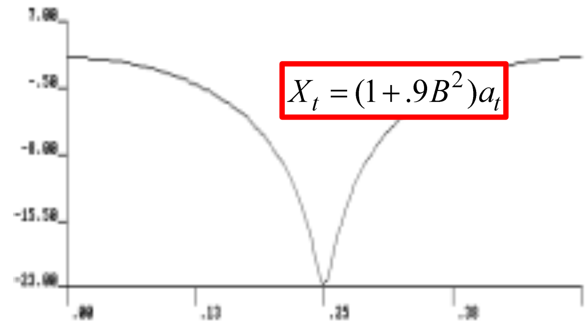
- The AR(3) part of the model (associated with the second column of graphs) displays:
  - A periodic behavior of about $f = .16$
  - A wandering behavior associated with $f = 0$; this is shown best in the realization and the spectral density
- The MA(1) part of the model is associated with a dip in the spectral density at $f = 0$.
  - It tends to "remove" behavior related to $f = 0$ (i.e., the aimless, aperiodic wandering)
- The ARMA(3,1) model retains the cyclic behavior, but the wandering behavior associated with $f = 0$ has been minimized.

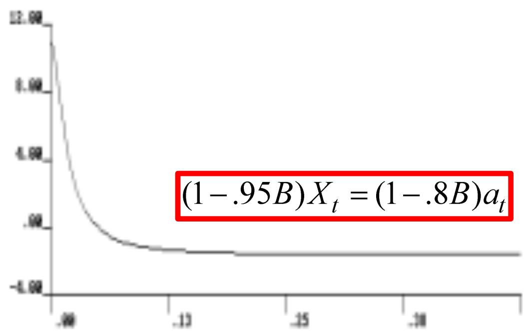
Spectral Densities

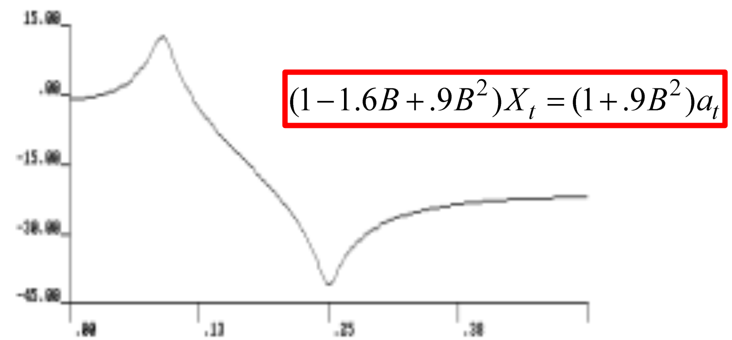










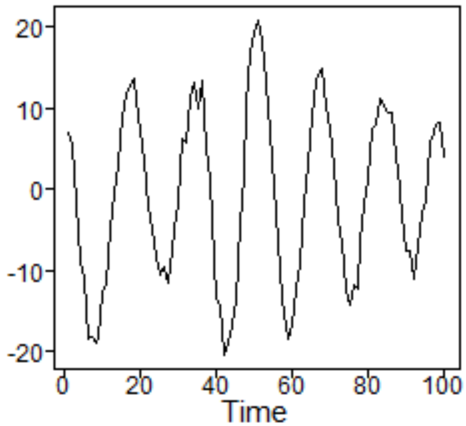


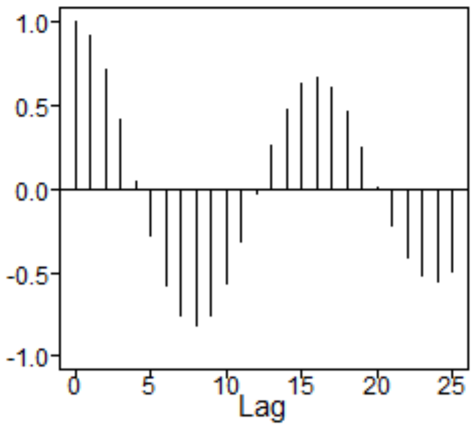
DataScience@SMU

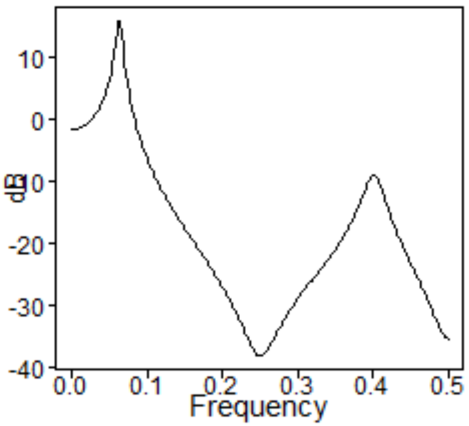
ARMA(*p*,*q*) | tswge and Illustrating the Blend of AR and MA

What Can You Tell About the ARMA Model from These Plots?

$$(1-.3B-.9B^2-.1B^3+.8075B^4)X_t = (1+.9B+.8B^2+.72B^3)a_t$$







tswge

```
# AR factors
factor.wge(c(.3,.9,.1,-.8075))
# MA factors
factor.wge(c(-.9,-.8,-.72))
```

DataScience@SMU

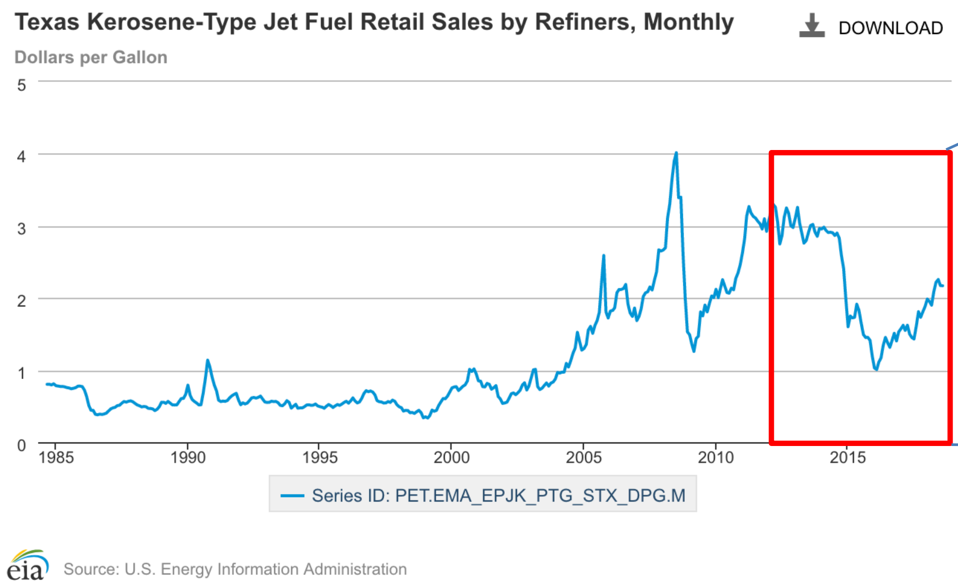
ARMA(p,q) | Example 1

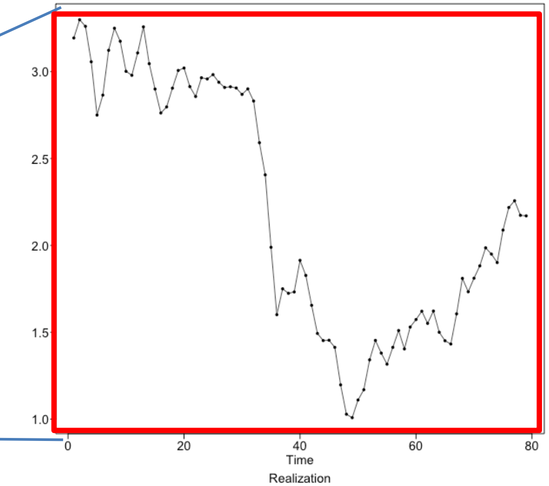
Jet Fuel A

Review: AIC

- Akaike’s Information Criterion
- Given a set of data, the AIC is used to evaluate and compare the quality of models
- Given a set of models, the model with the **lowest** AIC is thought to have the most quality

Jet Fuel Prices





plotts.wge(jetA$Price)

aic.wge(jetA$Price,p = 1, q = 0)$value

```
[1] -4.012713
```

aic.wge(jet$Price,p = 2, q = 0)$value

```
[1] -4.144347
```

aic.wge(jetA$Price,p = 1, q = 1)$value

```
[1] -4.195998
```

aic5.wge(jetA$Price)

```
Five Smallest Values of  aic
      p    q          aic
5     1    1 -4.195998155
10    3    0 -4.190446477
6     1    2 -4.172063529
8     2    1 -4.171285161
13    4    0 -4.165614321
```

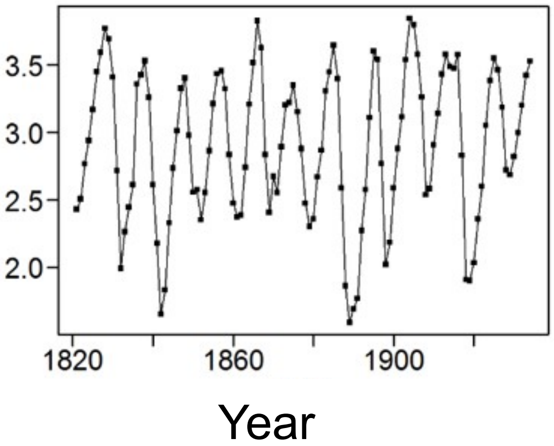
DataScience@SMU

ARMA(*p*,*q*) | Example 2

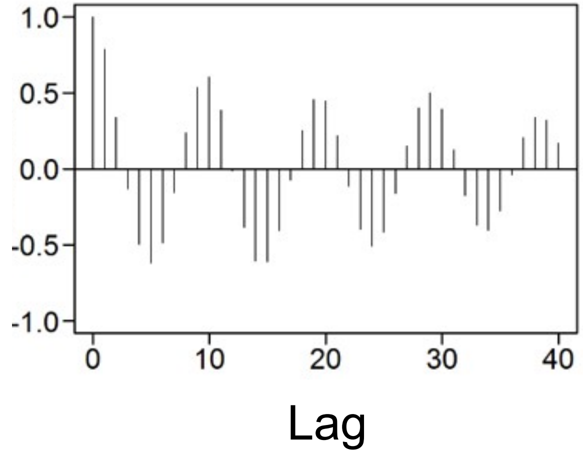
Canadian Lynx

Lynx Data

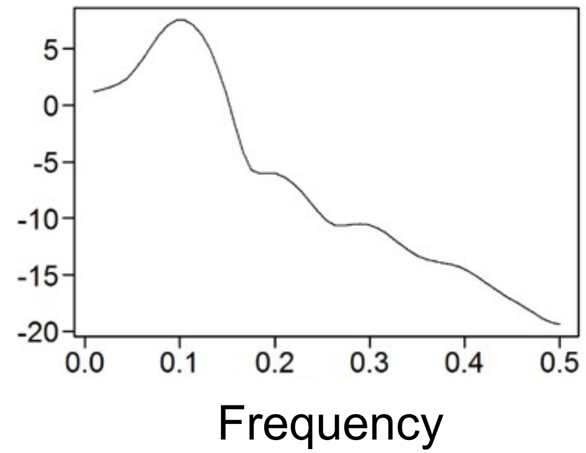
Log lynx data



Autocorrelations



Spectral density



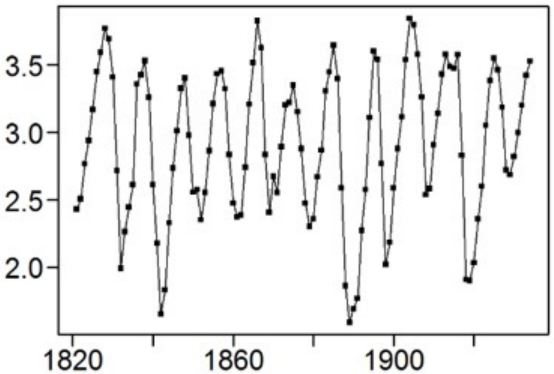
Notes: The above plots are typical of those produced by AR models for which the characteristic equation has complex conjugate roots.

- The log lynx data have a pseudo-cyclic appearance.
- The autocorrelations show damped sinusoidal behavior.
- The spectral density has a peak at some f_0 between 0 and $.5$ (at approximately $.1$ in this case).

DataScience@SMU

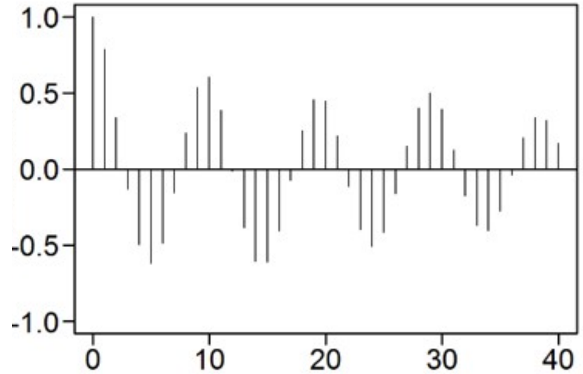
Lynx Data

Log lynx data



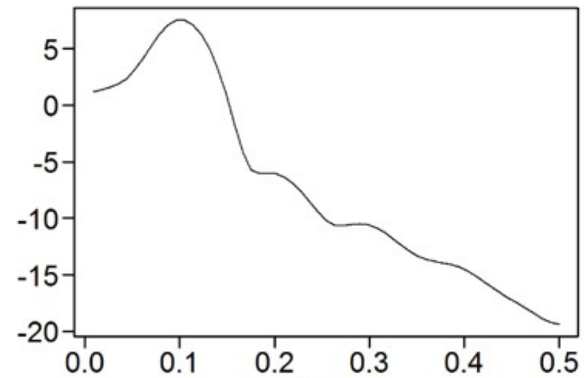
Year

Autocorrelations



Lag

Spectral density



Frequency

```
> aic5.wge(llynx)
---------WORKING... PLEASE WAIT...


Five Smallest Values of  aic
      p    q         aic
14    4    1   -2.951518
16    5    0   -2.951301
13    4    0   -2.949515
12    3    2   -2.942965
17    5    1   -2.936873
```

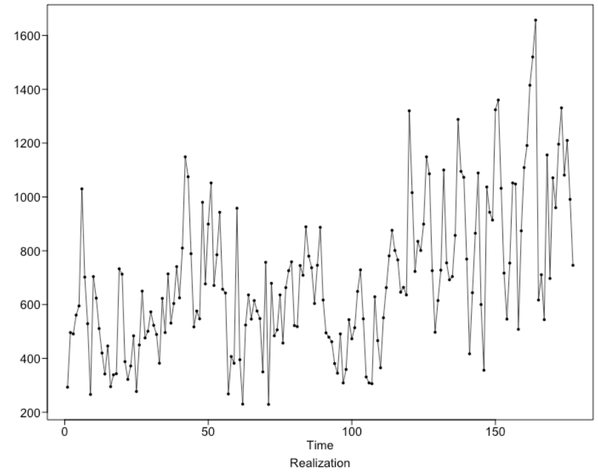
DataScience@SMU

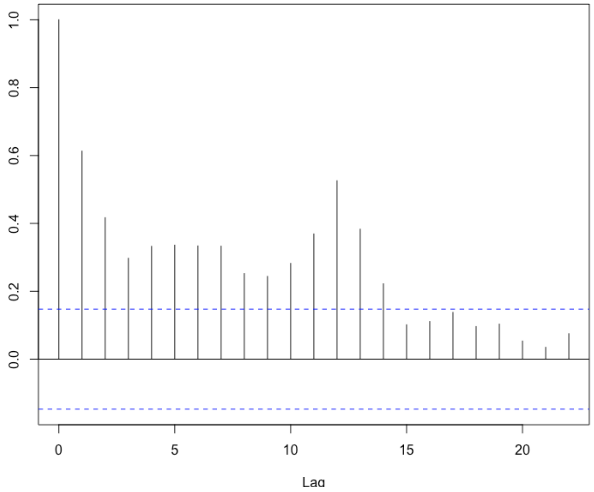
ARMA(p,q) | Example 3

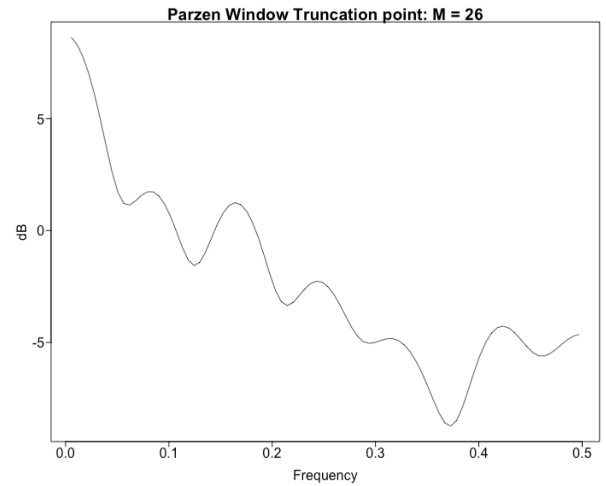
Airline Weather Delay

tswge Another Example: Airline Delays

SWA Monthly Delay Since 2004







Five Smallest Values of aic

	p	q	aic
14	4	1	19.00136
17	5	1	19.01266
18	5	2	19.02307
4	1	0	19.03069
8	2	1	19.04178

DataScience@SMU

ARMA(p,q) | PSI Weights

ψ-Weights

As mentioned earlier, the AR, MA, and ARMA processes can all be expressed as general linear processes (GLPs).

The form of the GLP is $X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$

Examples

MA

$X_t = a_t - .8a_{t-1}$ ψ-weights: $\psi_0 = 1, \psi_1 = -.8, \psi_j = 0,\ j > 1$

AR

$X_t - .8X_{t-1} = a_t$ ψ-weights: $\psi_j = .8^j,\ j \geq 0$

(See Woodward et al., 2017.)

$$X_t = a_t + .8a_{t-1} + .64a_{t-1} + .512a_{t-1} + \ldots$$

ψ-Weights

- Are used in establishing prediction limits on forecasts
- They are more difficult to obtain for AR(*p*) models, $p > 1$ and ARMA(*p*,*q*) models
  - See Woodward et al. (2017) for calculation techniques.
- In this course, we will calculate ψ-weights using tswge command psi.weights.wge

DataScience@SMU

ARMA(p,q) | tswge and PSI Weights

ψ-Weights and tswge

tswge gives $\psi_1, \psi_2, \ldots, \psi_{\text{lag.max}}$.

- In all cases, $\psi_0 = 1$.

```
# psi-weights for simple MA(1) model X(t)=(1-.8B)a(t)
psi.weights.wge(theta=.8,lag.max=5)
# psi-weights for simple AR(1) model (1-.8B)X(t)=a(t)
psi.weights.wge(phi=.8,lag.max=5)  #note that psi(j)=.8^j
# psi-weights for ARMA(2,1) model (1-1.2B+.6B^2)X(t)=(1-.5B)a(t)
psi.weights.wge(phi=c(1.2,-.6),theta=c(.5),lag.max=5)
```

DataScience@SMU

ARMA(p,q) | Summary

Comments about ARMA Models

- An MA(q) is an ARMA($0,q$) model.
- An AR(p) is an ARMA($p,0$) model.
- A stationary and invertible ARMA(p,q) model can be written in the two forms:

1. A GLP (infinite order MA): $X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$

2. Infinite order AR: $\underset{j=0}{} \quad {}_j X_t \quad {}_j = a_t$ (where ${}_0 = 1$)

(See Woodward et al., 2017)

Comments about ARMA Models

- An ARMA(p,q) can be written as an infinite order AR (form 2).
- That is, an ARMA(p,q) model may be able to be approximated by a high-order AR.
- Some time series analysts choose to do that and focus on AR models.
- Box and Jenkins favor an ARMA model fit over a high-order AR fit.
  - More "parsimonious" model
  - That is, fewer parameters

DataScience@SMU