STATIONARY MODELS

Dr. Aric LaBarr
Institute for Advanced Analytics
MSA Class of 2014

STATIONARITY

Objectives

- Introduce the naïve model for time series random walk.
- Define stationarity.
- Discuss situations where stationarity fails.

Naïve Models

In regression, the naïve model was the following:

$$Y_i = \mu + e_i$$

• The best prediction is \overline{Y} when no other information is given.

Naïve Models

In time series, the naïve model is the following:

$$Y_i = Y_{t-1} + e_i$$

- The best prediction is Y_{t-1} when no other information is given.
- This model is called the random walk model.

Random Walk Model

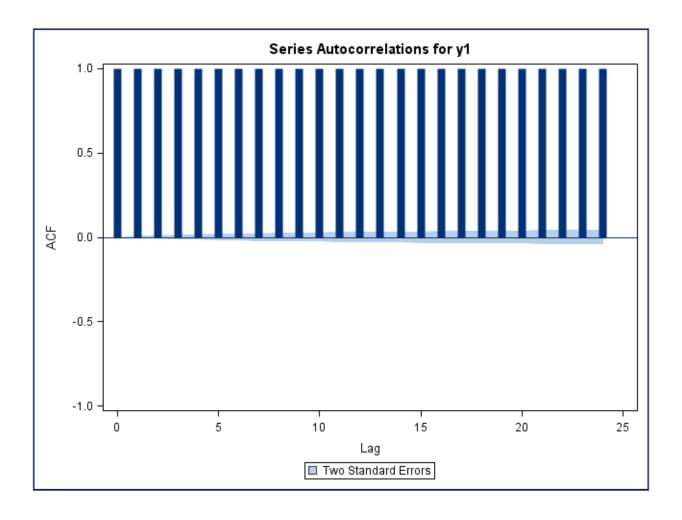
- There are two types of random walk models:
 - 1. Random Walk

$$Y_i = Y_{t-1} + e_i$$

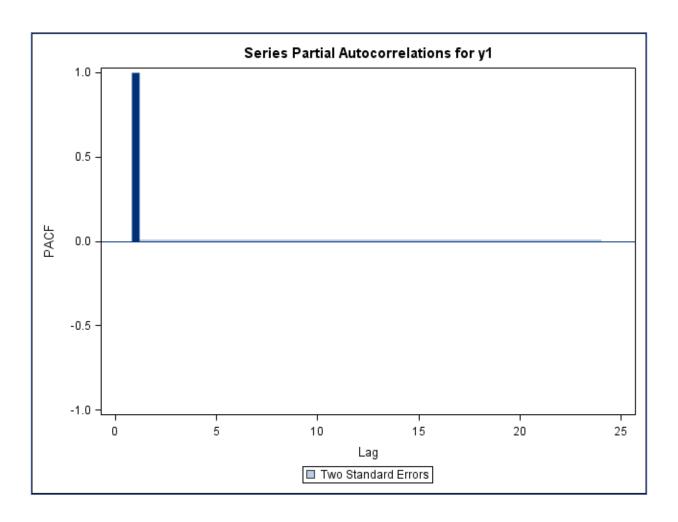
Random Walk with Drift

$$Y_i = \omega + Y_{t-1} + e_i$$

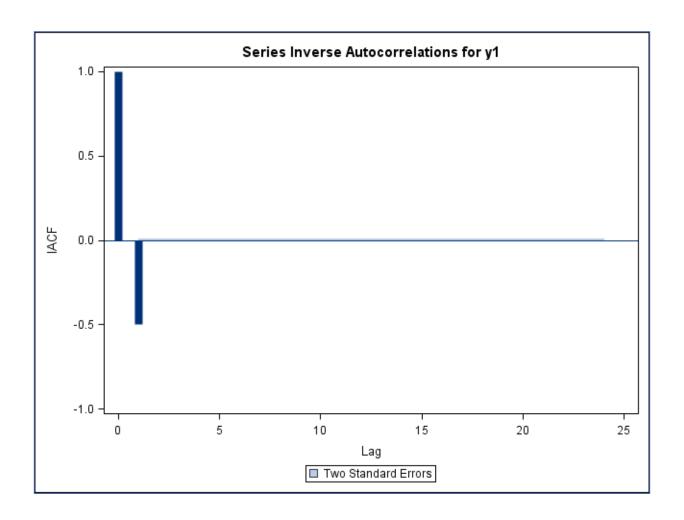
RW – Autocorrelation Function



RW— Partial Autocorrelation Function



RW – Inverse Autocorrelation Function



Stationarity

- The random walk model is an example of a model that lacks stationarity.
- We assumed that there exists a dependence structure to our times series.
- Does this dependence decrease across time?
- How quickly does the persistence of observations decrease?

Stationarity

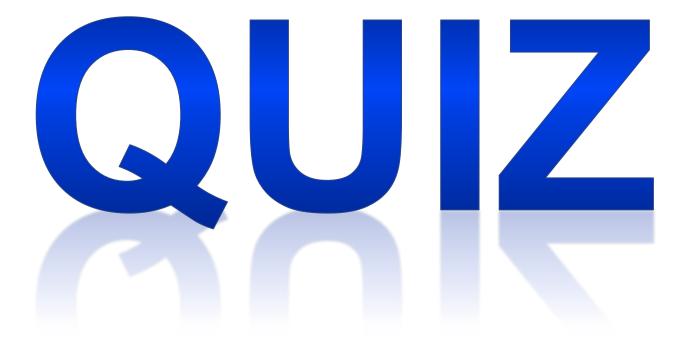
- A stationary time series has a constant mean and variance.
- A time series with long-term trend or seasonal components cannot be stationary because the mean of the series depends on the time that the value is observed.
 - Population gradually increasing over time.
 - Average sales in December are always higher than average sales in March.

Stationarity

- The foundation of time series modeling is based in stationary time series models.
- A stationary time series can be short, moderate, or long term memory.
 - How long does the effect of an observation persist in the model?

Stationarity Time Series Models

- All stationary models in time series revert to the mean of the series.
- A long memory stationary time series generates forecasts that revert to the mean much more slowly than a short memory time series.
- To obtain accurate forecasts for many steps into the future, you want a time series to be non-stationary, but if it is stationary, you want it to be long memory.
- Mean reversion can be a "potential issue" for really longterm forecasts where the average of the series is a rather boring forecast.



Quiz 1

- Which of the following statements is true?
 - 1. Time series forecasting methods require that the time series of interest be stationary.
 - 2. A trigonometric sine wave is stationary because it cycles about a constant value.
 - 3. Forecasts for stationary time series always converge to the mean of the series.
 - 4. Long-term forecasts are more accurate for stationary time series than for non-stationary time series.

Quiz 1

- Which of the following statements is true?
 - 1. Time series forecasting methods require that the time series of interest be stationary.
 - A trigonometric sine wave is stationary because it cycles about a constant value.
 - 3.) Forecasts for stationary time series always converge to the mean of the series.
 - 4. Long-term forecasts are more accurate for stationary time series than for non-stationary time series.



AUTOREGRESSIVE MODELS

Objectives

- Define the autoregressive model with lag 1.
- Extend the model to include p lags.
- Visualize the ACF, PACF, and IACF for the autoregressive class of models.

- Often you can forecast a series based solely on the past values.
- We are going to focus on the most basic case only one lag value of Y_t called an AR(1) model:

$$Y_t = \omega + \phi Y_{t-1} + e_t$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega + \phi Y_{t-1} + e_{t}$$
$$Y_{t-1} = \omega + \phi Y_{t-2} + e_{t-1}$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega + \phi Y_{t-1} + e_{t}$$

$$Y_{t-1} = \omega + \phi Y_{t-2} + e_{t-1}$$

$$Y_{t} = \omega + \phi(\omega + \phi Y_{t-2} + e_{t-1}) + e_{t}$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega + \phi Y_{t-1} + e_{t}$$

$$Y_{t-1} = \omega + \phi Y_{t-2} + e_{t-1}$$

$$Y_{t} = \omega + \phi(\omega + \phi Y_{t-2} + e_{t-1}) + e_{t}$$

$$Y_{t} = \omega^{*} + \phi^{2} Y_{t-2} + \phi e_{t-1} + e_{t}$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega + \phi Y_{t-1} + e_{t}$$

$$Y_{t-1} = \omega + \phi Y_{t-2} + e_{t-1}$$

$$Y_{t} = \omega + \phi(\omega + \phi Y_{t-2} + e_{t-1}) + e_{t}$$

$$Y_{t} = \omega^{*} + \phi^{2}Y_{t-2} + \phi e_{t-1} + e_{t}$$

REPEAT!!!

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega^* + \phi^2 Y_{t-2} + \phi e_{t-1} + e_t$$
$$Y_{t-2} = \omega + \phi Y_{t-3} + e_{t-2}$$

- This relationship between *t* and *t-1* exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega^{*} + \phi^{2}Y_{t-2} + \phi e_{t-1} + e_{t}$$

$$Y_{t-2} = \omega + \phi Y_{t-3} + e_{t-2}$$

$$Y_{t} = \omega^{*} + \phi^{2}(\omega + \phi Y_{t-3} + e_{t-2}) + \phi e_{t-1} + e_{t}$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_{t} = \omega^{*} + \phi^{2}Y_{t-2} + \phi e_{t-1} + e_{t}$$

$$Y_{t-2} = \omega + \phi Y_{t-3} + e_{t-2}$$

$$Y_{t} = \omega^{*} + \phi^{2}(\omega + \phi Y_{t-3} + e_{t-2}) + \phi e_{t-1} + e_{t}$$

$$Y_{t} = \omega^{*} + \phi^{3}Y_{t-3} + \phi^{2}e_{t-2} + \phi e_{t-1} + e_{t}$$

REPEAT until we return to beginning!

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_t = \omega + \phi Y_{t-1} + e_t$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_t = \omega + \phi Y_{t-1} + e_t$$

MATH!!!

$$Y_t = \frac{\omega}{1 - \phi} + \phi^t Y_1 + \phi^{t-1} e_2 + \phi^{t-2} e_3 + \dots + e_t$$

- This relationship between t and t-1 exists for all one time period differences across the data set.
- Therefore, we can recursively solve for Y_t :

$$Y_t = \omega + \phi Y_{t-1} + e_t$$

MATH!!!

$$Y_t = \frac{\omega}{1 - \phi} + \phi^t Y_1 + \phi^{t-1} e_2 + \phi^{t-2} e_3 + \dots + e_t$$

$$Y_t = \frac{\omega}{1 - \phi} + \phi^t Y_1 + \phi^{t-1} e_2 + \phi^{t-2} e_3 + \dots + e_t$$

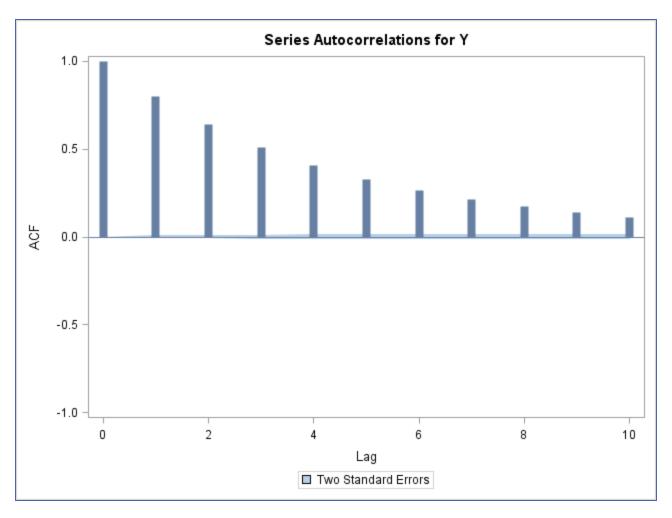
- So the effect of shocks that happened long ago has little effect on the present IF the value for $|\phi| < 1$.
- This goes back to our idea of stationarity the dependence of previous observations declines over time.

Correlation Functions for AR(1)

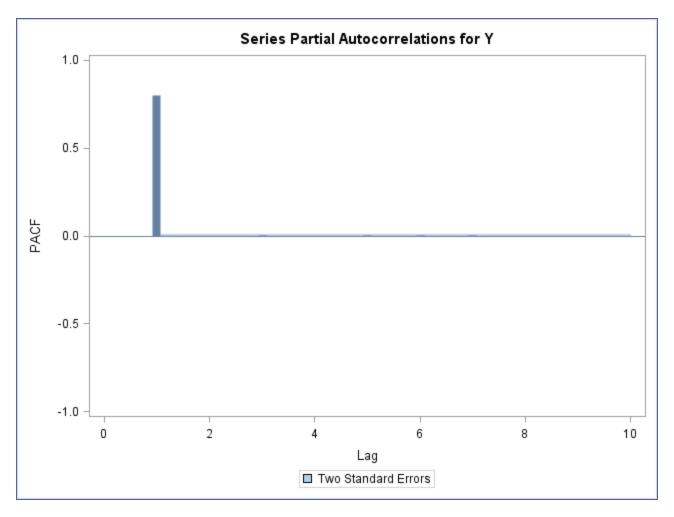
- The ACF decreases exponentially as the number of lags increases.
- The PACF has a significant spike at the first lag, followed by nothing after.
- The IACF has a significant spike at the first lag, followed by nothing after.
- Let's examine the following AR(1) model:

$$Y_t = 0 + 0.8Y_{t-1} + e_t$$

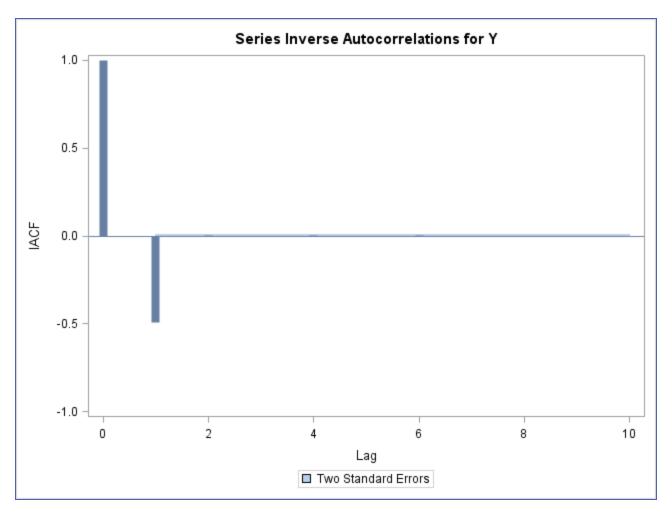
AR(1) - ACF



AR(1) - PACF



AR(1) - IACF



AR(p) Model

 A time series that is a linear function of p past values plus error is called an autoregressive process of order p – AR(p).

$$Y_t = \omega + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

Correlation Functions for AR(p)

- The ACF can have a variety of patterns.
- The PACF has a significant spike at the significant lags up to p lags, followed by nothing after.
- The IACF has a significant spike at the significant lags up to p lags, followed by nothing after.

Autoregressive Models

```
proc arima data=Time.AR2 plot=all;
   identify var=y nlag=10;
   estimate p=2 method=ML;
run;
quit;
```



MOVING AVERAGE MODELS

Objectives

- Define the moving average model with lag 1.
- Extend the model to include q lags.
- Visualize the ACF, PACF, and IACF for the moving average class of models.

Moving Average (MA) Models

- You can also forecast a series based solely on the past error values.
- This kind of model is better for describing events whose effect only lasts for short periods of time.
- We are going to focus on the most basic case only one error lag value of e_t , called an MA(1) model:

$$Y_t = \omega + e_t + \theta e_{t-1}$$

• Therefore, for an MA(1) model, individual "shocks" only last for a short time.

$$Y_{t-1} = \omega + e_{t-1} + \theta e_{t-2}$$

 Therefore, for an MA(1) model, individual "shocks" only last for a short time.

$$Y_{t-1} = \omega + e_{t-1} + \theta e_{t-2}$$

$$Y_t = \omega + e_t + \theta e_{t-1}$$

 Therefore, for an MA(1) model, individual "shocks" only last for a short time.

$$Y_{t-1} = \omega + e_{t-1} + \theta e_{t-2}$$

$$Y_t = \omega + e_t + \theta e_{t-1}$$

$$Y_{t+1} = \omega + e_{t+1} + \theta e_t \longrightarrow \text{No more } e_{t-1}!!$$

 Therefore, for an MA(1) model, individual "shocks" only last for a short time.

$$Y_{t-1} = \omega + e_{t-1} + \theta e_{t-2}$$

$$Y_t = \omega + e_t + \theta e_{t-1}$$

$$Y_{t+1} = \omega + e_{t+1} + \theta e_t \longrightarrow \text{No more } e_{t-1}!!$$

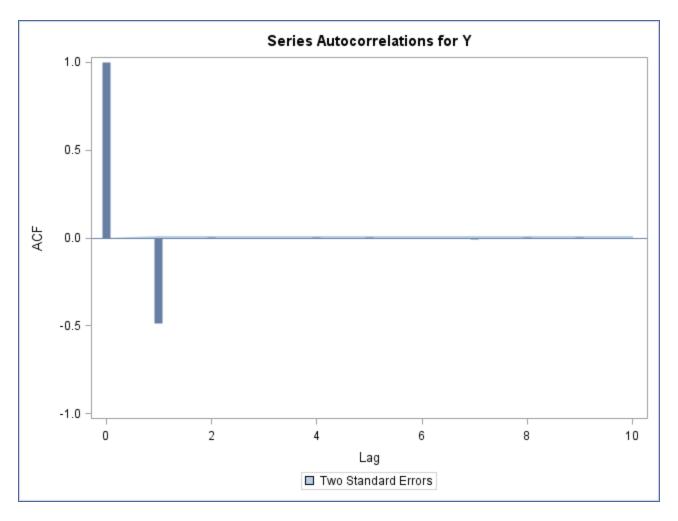
 This again refers back to stationarity – the dependence of previous "shocks" eventually disappear.

Correlation Functions for MA(1)

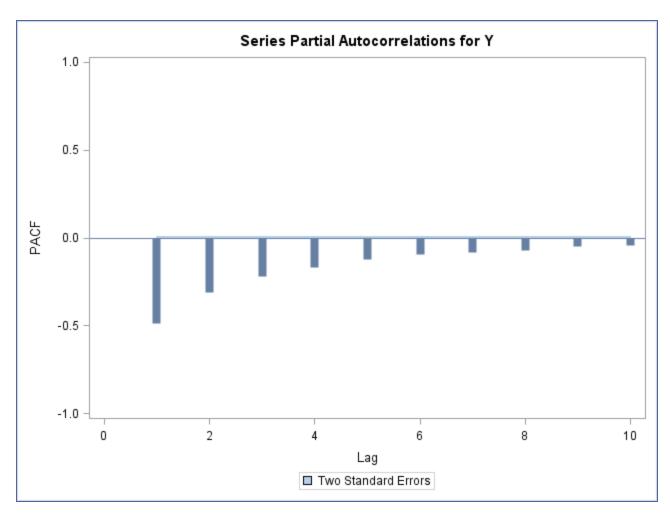
- The ACF has a significant spike at the first lag, followed by nothing after.
- The PACF decreases exponentially as the number of lags increases.
- The IACF decreases exponentially as the number of lags increases.
- Let's examine the following MA(1) model:

$$Y_t = 0 + e_t - 0.8e_{t-1}$$

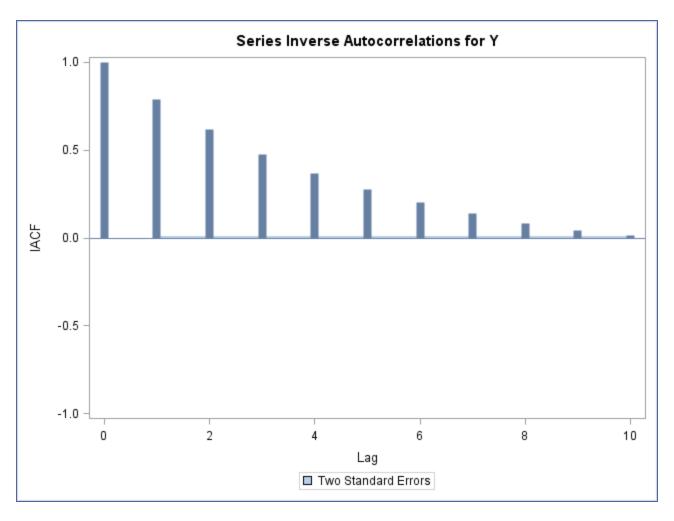
MA(1) - ACF



MA(1) - PACF



MA(1) - IACF



 A time series that is a linear function of q past errors is called a moving average process of order q – called an MA(q).

$$Y_t = \omega + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Correlation Functions for MA(q)

- The ACF has a significant spike at the significant lags up to lag q, followed by nothing after.
- The PACF can have a variety of patterns.
- The IACF can have a variety of patterns.

Moving Average Models

```
proc arima data=Time.SimMA1 plot=all;
   identify var=Y nlag=12;
   estimate q=1 method=ML;
run;
quit;
```



ARMA MODELS

Objectives

- Describe the relationship between AR and MA models.
- Construct the autoregressive moving average model (ARMA) of order (1, 1).
- Extend the ARMA(1,1) model to the ARMA(p,q) model.
- Visualize the ACF, PACF, and IACF for the autoregressive moving average class of models.

- The best part about AR models and MA models is that they are the same thing – approximately.
- In certain situations (stationarity), AR models can be represented as an infinite MA model.
- In certain situations, MA models can be represented as an infinite AR model.

Remember the recursive calculation for the AR model:

$$\begin{split} Y_t &= \omega + \phi Y_{t-1} + e_t \\ Y_t &= \omega^* + \phi^2 Y_{t-2} + \phi e_{t-1} + e_t \\ Y_t &= \omega^* + \phi^3 Y_{t-3} + \phi^2 e_{t-2} + \phi e_{t-1} + e_t \\ &\vdots \\ Y_t &= \omega^* + \phi^t Y_1 + \phi^{t-1} e_2 + \phi^{t-2} e_3 + \dots + e_t \end{split}$$

Remember the recursive calculation for the AR model:

$$Y_{t} = \omega + \phi Y_{t-1} + e_{t}$$

$$Y_{t} = \omega^{*} + \phi^{2} Y_{t-2} + \phi e_{t-1} + e_{t}$$

$$Y_{t} = \omega^{*} + \phi^{3} Y_{t-3} + \phi^{2} e_{t-2} + \phi e_{t-1} + e_{t}$$

$$\vdots$$

$$Y_{t} = \omega^{*} + \phi^{t} Y_{1} + \phi^{t-1} e_{2} + \phi^{t-2} e_{3} + \dots + e_{t}$$

Large moving average term!!

- A similar calculation can be done to get a large number of AR lags for an MA(1) model.
- This is why the PACF and IACF in an MA(1) have an exponentially decreasing correlations as the lags increase.

ARMA Model

- There is nothing to limit both an AR process and an MA process to be in the model simultaneously.
- These "mixed" models are typically used to help reduce the number of parameters needed for good estimation in the model.
- We are going to focus on the most basic model with only one lag of each piece – the ARMA(1,1) model.

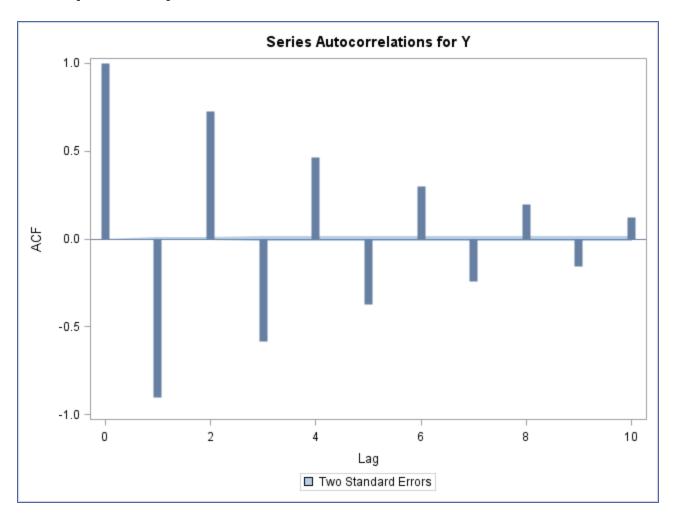
$$Y_t = \omega + \phi Y_{t-1} + e_t - \theta e_{t-1}$$

ARMA(1, 1) – ACF, PACF, IACF

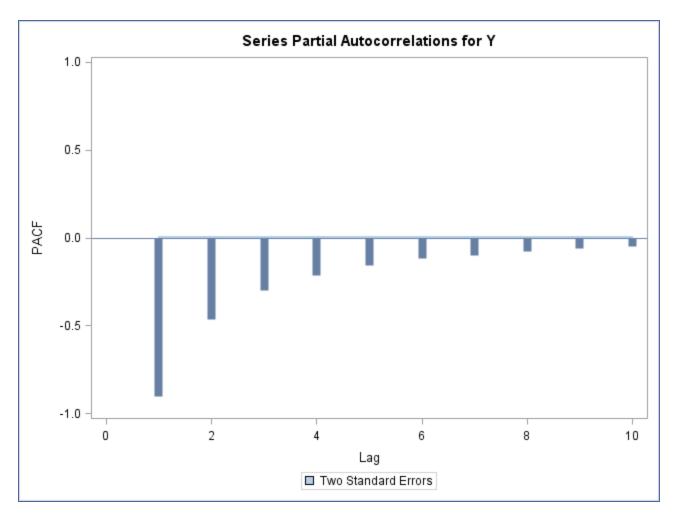
- Although these terms can be calculated for the ARMA(1,1) process, they become very complicated.
- There are some important things to note:
 - Characteristics from both are in the correlation functions.
 - All of the functions tail off exponentially as the lags increase.
- Let's plot the ACF, PACF, and IACF for the following:

$$Y_t = 0 - 0.8Y_{t-1} + e_t - 0.8e_{t-1}$$

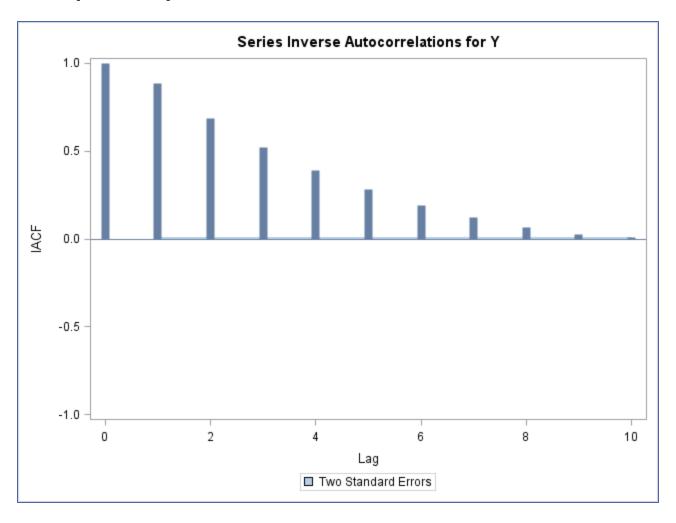
ARMA(1, 1) - ACF



ARMA(1, 1) - PACF



ARMA(1, 1) - IACF



ARMA Models

```
proc arima data=Time.SimARMA plot=all;
   identify var=Y nlag=12;
   estimate p=1 q=1 method=ML;
run;
quit;
```



MODEL SELECTION

Objectives

- Review visual inspection of ACF, PACF, and IACF for model selection.
- Introduce the automatic selection methods of MINIC, SCAN, and ESACF.
- Work through an example.

Identification

- There are a couple of different sets of techniques used for model identification for stationary models.
 - Plotting Patterns ACF, PACF, IACF
 - 2. Automatic Selection Techniques:
 - Minimum Information Criterion MINIC
 - Smallest Canonical Correlation SCAN
 - Extended Sample Autocorrelation Function ESACF

Plotting Patterns – Box-Jenkins

 Here is a review of the patterns we have previously talked about as we introduced the models:

	MA(q)	AR(p)	ARMA(p,q)	WN
ACF	D(q)	Т	Т	0
PACF	Т	D(p)	Т	0
IACF	Т	D(p)	Т	0

- T the function tails off exponentially
- D(i) the function drops off after lag i
- 0 the function is 0 at all nonzero lags

Plotting Patterns – Box-Jenkins

 How do we handle the situation where both functions are suppose to tail off?

	MA(q)	AR(p)	ARMA(p,q)	WN
ACF	D(q)	Т	Т	0
PACF	Т	D(p)	Т	0
IACF		D(p)	T	0

- 1. Start by identifying the *significant* correlations use those as starting point for *p* and *q*.
- 2. Trial and error around that starting point.

```
proc arima data=Time.Hurricanes plot(unpack);
    identify var=MeanVMax nlag=12 minic
    P=(0:12) Q=(0:12);
run;
quit;
```

```
proc arima data=Time.Hurricanes plot(unpack);
    identify var=MeanVMax nlag=12 scan
    P=(0:12) Q=(0:12);
run;
quit;
```

```
proc arima data=Time.Hurricanes plot(unpack);
    identify var=MeanVMax nlag=12 esacf
    P=(0:12) Q=(0:12);
run;
quit;
```



FORECASTING

Objectives

- Describe two of the estimation methods available in PROC ARIMA.
- List some of the issues related to the optimization algorithms used to obtain parameter estimates.
- Learn how to forecast values for ARMA models.
- Combine identification, estimation, and forecasting into one example.

Estimation Methods – CLS

- Conditional Least Squares estimators are the following:
 - Generally inferior to MLE for small samples
 - More computationally efficient than MLE
 - Are the DEFAULT in PROC ARIMA
- "Conditional" least squares comes from the fact that estimation of the parameter estimates is conditioned on unobserved past values being equal to the sample mean.

Estimation Methods – MLE

- Maximum Likelihood estimators are the following:
 - Superior to other estimates, especially in small samples.
 - Least efficient computationally, therefore not made the default when PROC ARIMA was created in 1980.
 - To help keep upward compatibility in SAS, the defaults have stayed the same in PROC ARIMA.
 - Method of choice by most forecasting professionals as well as SAS.

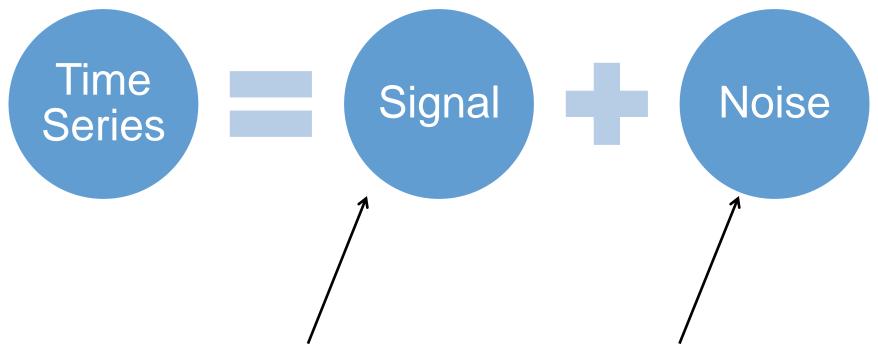
Optimization Algorithms

- CLS and ML algorithms are not guaranteed to find an optimal solution.
- Problems:
 - Local Maxima/Minima
 - Ridges (no improvement in any direction, but stopping rule not satisfied)
 - Stability Problems
 - Others

Forecasting

- The best part about time series models using information from the past is that you could forecast the future.
- If we know the pattern on how an observation at one time period is related to an observation at another time period, then we can recursively forecast the future using this pattern (assuming it stays the same).

Statistical Forecasting



Forecasts extrapolate signal portion of model.

Confidence intervals account for uncertainty.

Forecasting -AR(1)

 Autoregressive models are easy to forecast because of the recursive nature of the model:

$$\begin{split} Y_t &= \widehat{\omega} + \widehat{\phi} Y_{t-1} \\ \widehat{Y}_{t+1} &= \widehat{\omega} + \widehat{\phi} Y_t \\ \widehat{Y}_{t+2} &= \widehat{\omega} + \widehat{\phi} \widehat{Y}_{t+1} = \widehat{\omega} + \widehat{\phi}^2 Y_t \\ &\vdots \\ \widehat{Y}_{t+k} &= \widehat{\omega} + \widehat{\phi} \widehat{Y}_{t+k-1} = \widehat{\omega} + \widehat{\phi}^k Y_t \end{split}$$

Forecasting -AR(1)

 Autoregressive models are easy to forecast because of the recursive nature of the model:

$$\begin{aligned} Y_t &= \widehat{\omega} + \widehat{\phi} Y_{t-1} \\ \widehat{Y}_{t+1} &= \widehat{\omega} + \widehat{\phi} Y_t \\ \widehat{Y}_{t+2} &= \widehat{\omega} + \widehat{\phi} \widehat{Y}_{t+1} = \widehat{\omega} + \widehat{\phi}^2 Y_t \\ &\vdots \\ \widehat{Y}_{t+k} &= \widehat{\omega} + \widehat{\phi} \widehat{Y}_{t+k-1} = \widehat{\omega} + \widehat{\phi}^k Y_t \end{aligned}$$

The same idea can be extended to the AR(p) model.

Forecasting – MA(1)

- The MA(1) model is a little harder to forecast because you need an estimate of the current error, not just the current value.
- So we need to start from the beginning:

$$\begin{split} \hat{Y}_1 &= \bar{Y} \rightarrow \hat{e}_1 = Y_1 - \hat{Y}_1 \\ \hat{Y}_2 &= \hat{\omega} - \hat{\theta} \hat{e}_1 \rightarrow \hat{e}_2 = Y_2 - \hat{Y}_2 \\ \hat{Y}_3 &= \hat{\omega} - \hat{\theta} \hat{e}_2 \rightarrow \hat{e}_3 = Y_3 - \hat{Y}_3 \\ &\vdots \\ \hat{Y}_{t+1} &= \hat{\omega} - \hat{\theta} \hat{e}_t \\ \hat{Y}_{t+2} &= \hat{\omega} \end{split}$$

Forecasting Summary for AR and MA

- Forecasts for autoregressive models revert to the mean slowly over time.
- Forecasts for moving average models revert to the mean quickly over time.

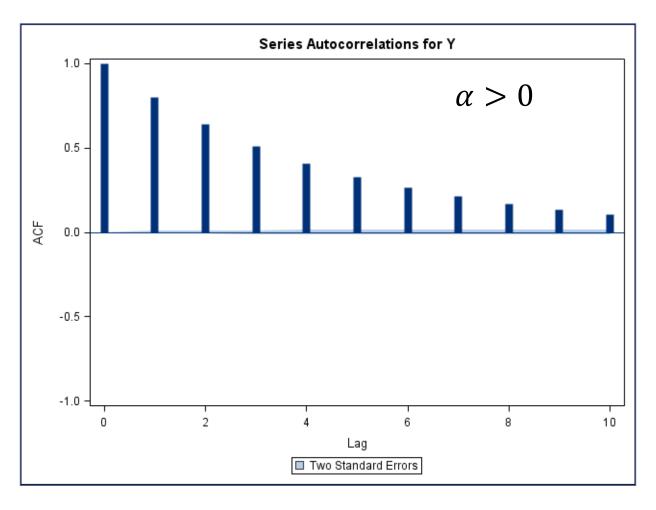
Forecasting ARMA(p, q)

- A combined approach between the AR(p) and MA(q) models is used for an ARMA(p, q) model.
- We will not go through the details here because it is just a combination of the two previous cases.

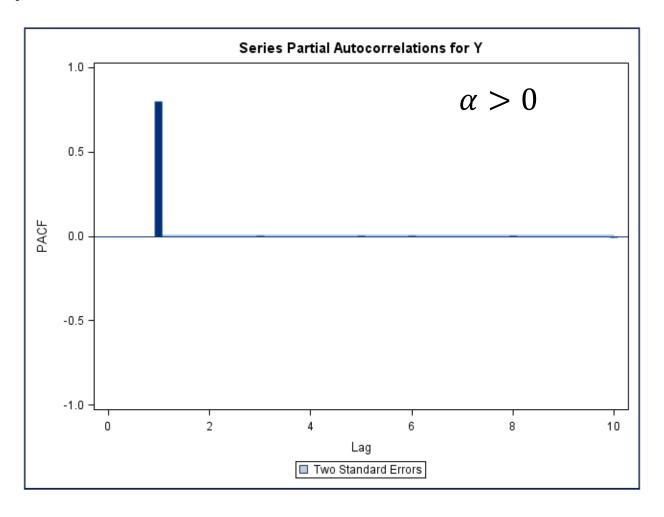
Forecasting

MODEL IDENTIFICATION CHARTS

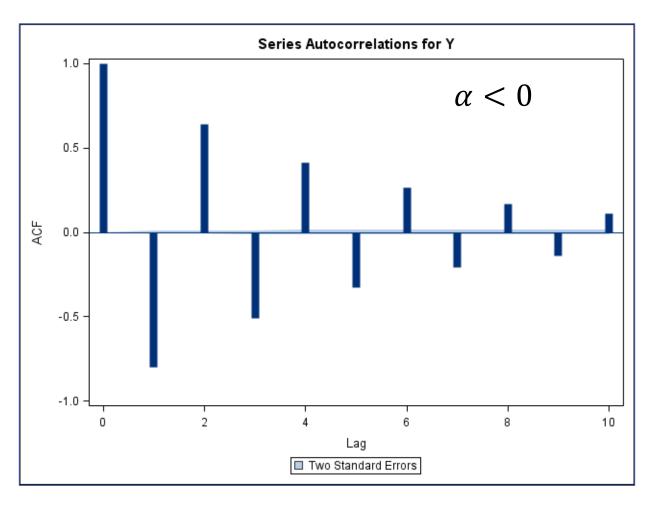
AR(1) – Autocorrelation Function



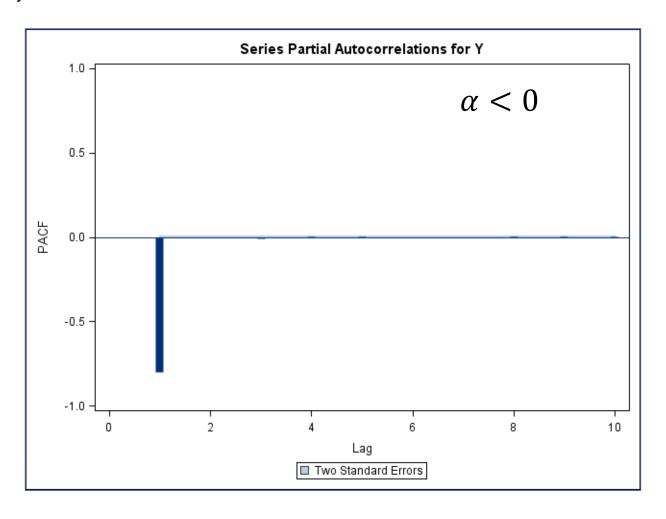
AR(1) – Partial Autocorrelation Function



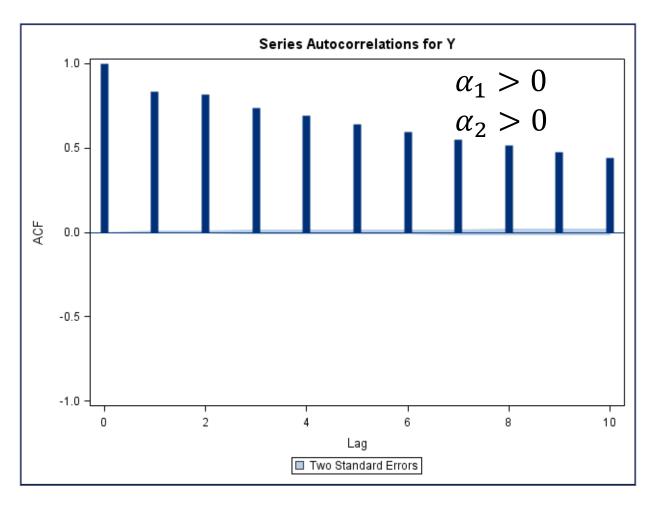
AR(1) – Autocorrelation Function



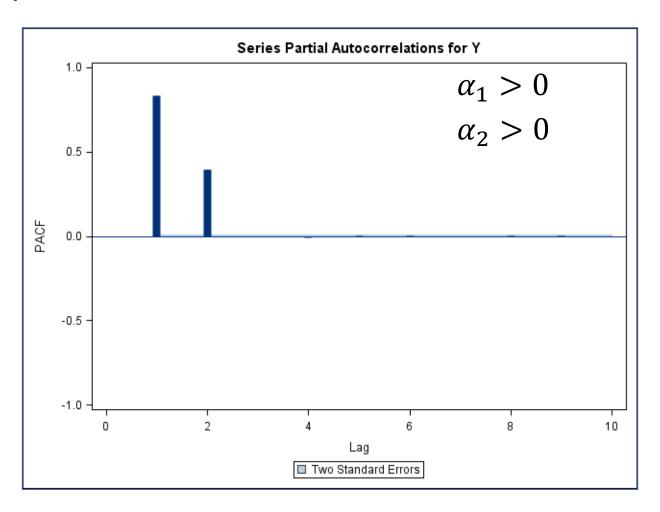
AR(1) – Partial Autocorrelation Function



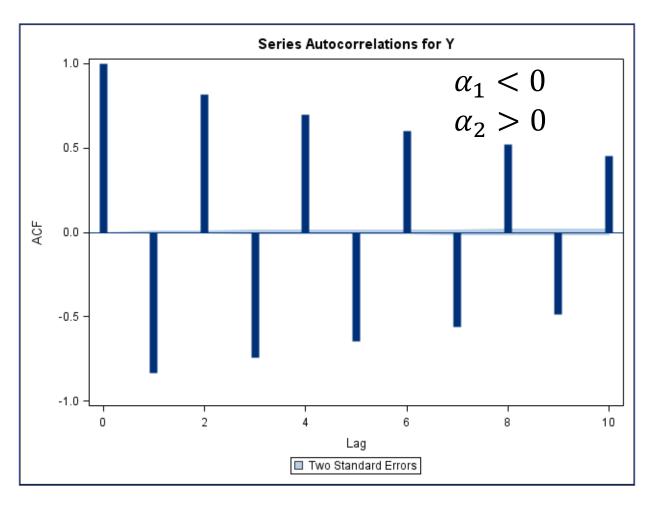
AR(2) – Autocorrelation Function



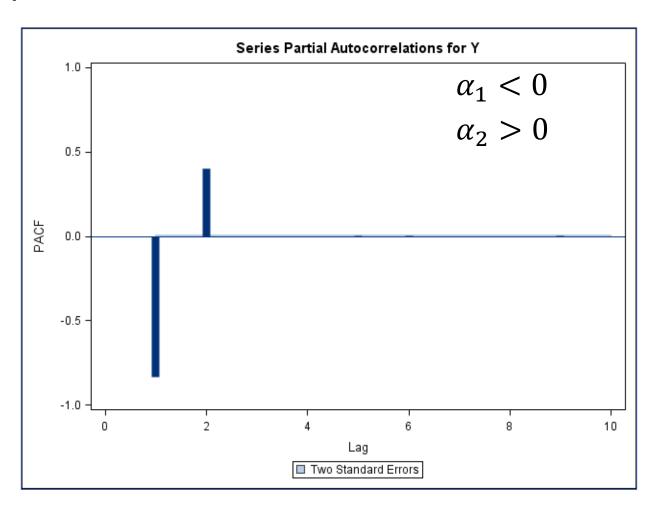
AR(2) – Partial Autocorrelation Function



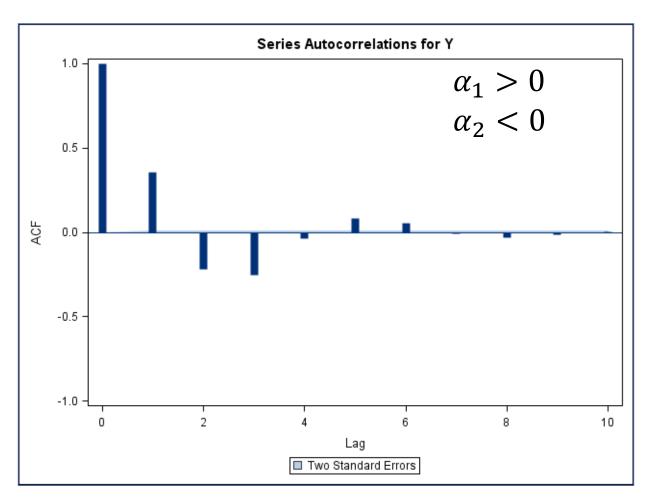
AR(2) – Autocorrelation Function



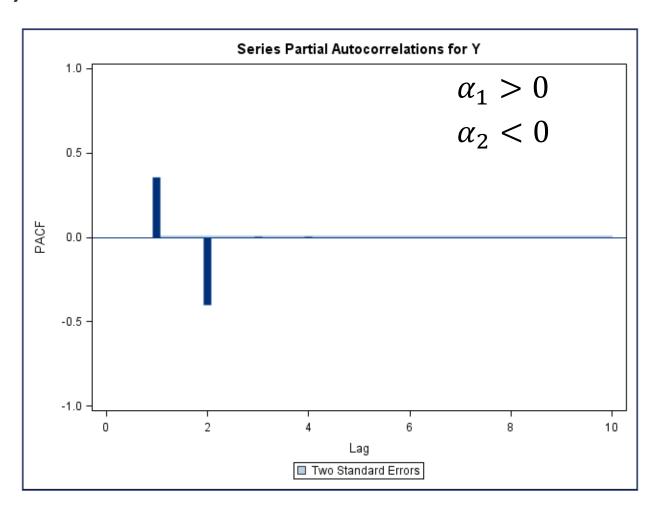
AR(2) – Partial Autocorrelation Function



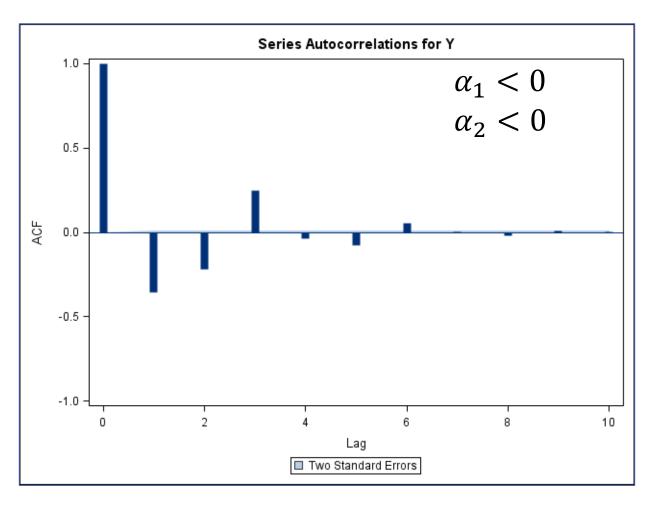
AR(2) – Autocorrelation Function



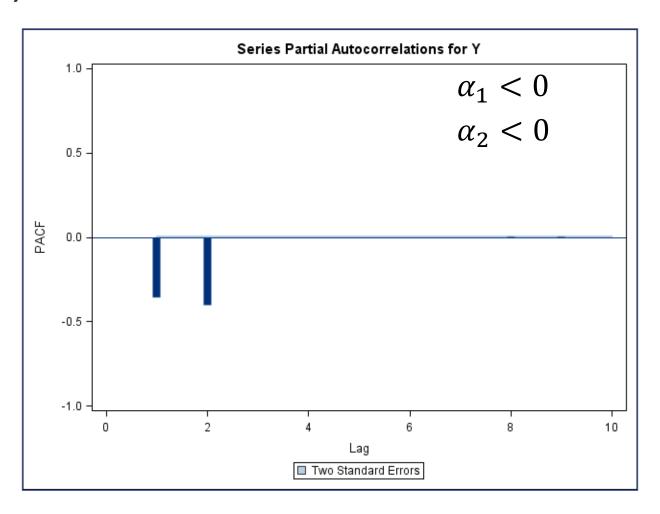
AR(2) – Partial Autocorrelation Function



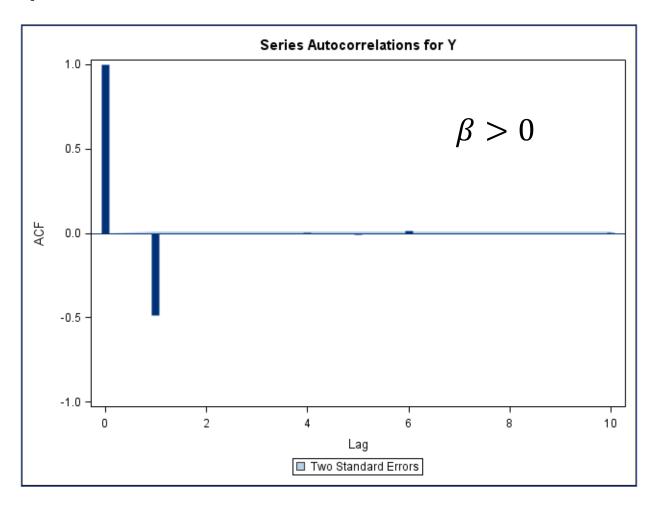
AR(2) – Autocorrelation Function



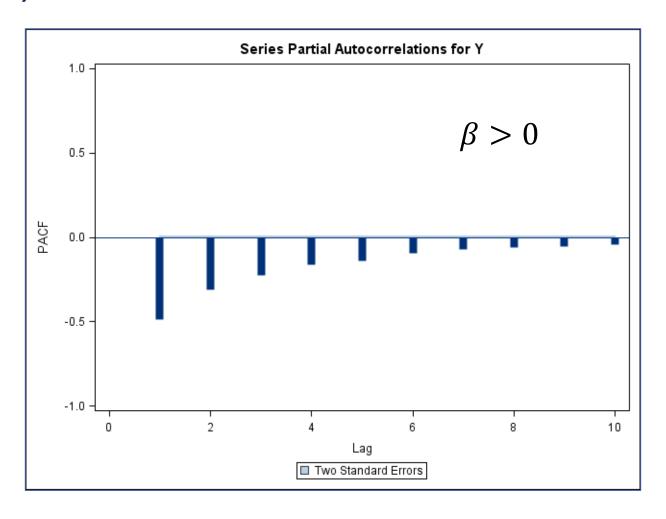
AR(2) – Partial Autocorrelation Function



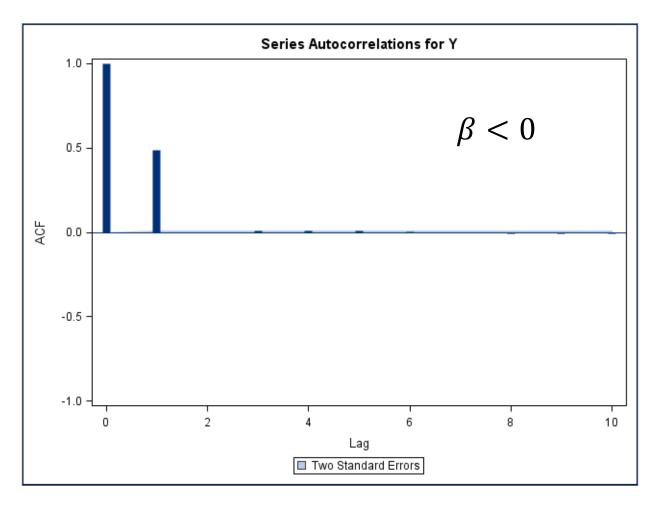
MA(1) – Autocorrelation Function



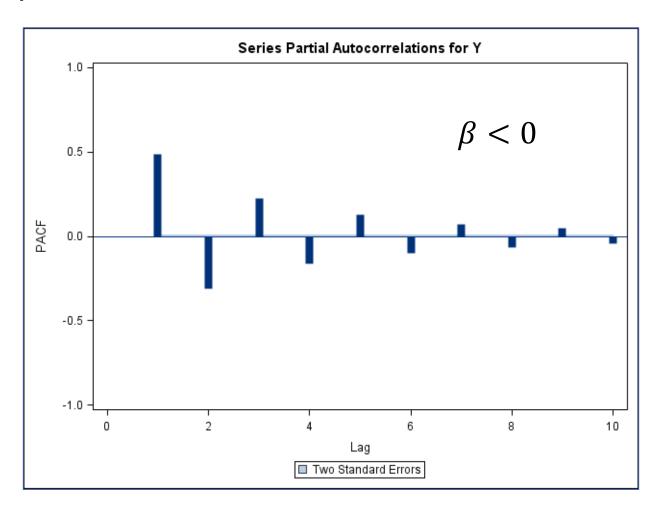
MA(1) – Partial Autocorrelation Function



MA(1) – Autocorrelation Function



MA(1) – Partial Autocorrelation Function



MA(2) – Autocorrelation Function

