CORRELATION FUNCTIONS AND WHITE NOISE

Dr. Susan Simmons
Institute for Advanced Analytics

CORRELATION FUNCTIONS

Dependencies

- A time series is *typically* analyzed with an assumption that observations have a potential relationship across time.
  - Ex: Weight

- Same approach can be taken with space as well as time.
  - Ex: Temperature

Autocorrelation Function

- *Autocorrelation* is the correlation between two sets of observations, from the same series, that are separated by *k* points in time.
- The autocorrelation function (ACF) is the function of all autocorrelations (between two **sets of observations** Y_t and Y_{t-k}) across time (for all values of *k*).

$$\rho_k = \mathrm{Corr}(Y_t, Y_{t-k})$$

Autocorrelation Function

t	Y_t	Y_{t-1}	Y_{t-2}
1	20	.	.
2	2	20	.
3	16	2	20
4	-3	16	2
5	-14	-3	16
6	-28	-14	-3
…	…	…	
999	0	29	17
1000	-19	0	29

Autocorrelation Function

t	Y_t	Y_{t-1}	Y_{t-2}
1	20	.	.
2	2	20	.
3	16	2	20
4	-3	16	2
5	-14	-3	16
6	-28	-14	-3
…	…	…	
999	0	29	17
1000	-19	0	29

$$\hat{\rho}_1 = 0.46$$

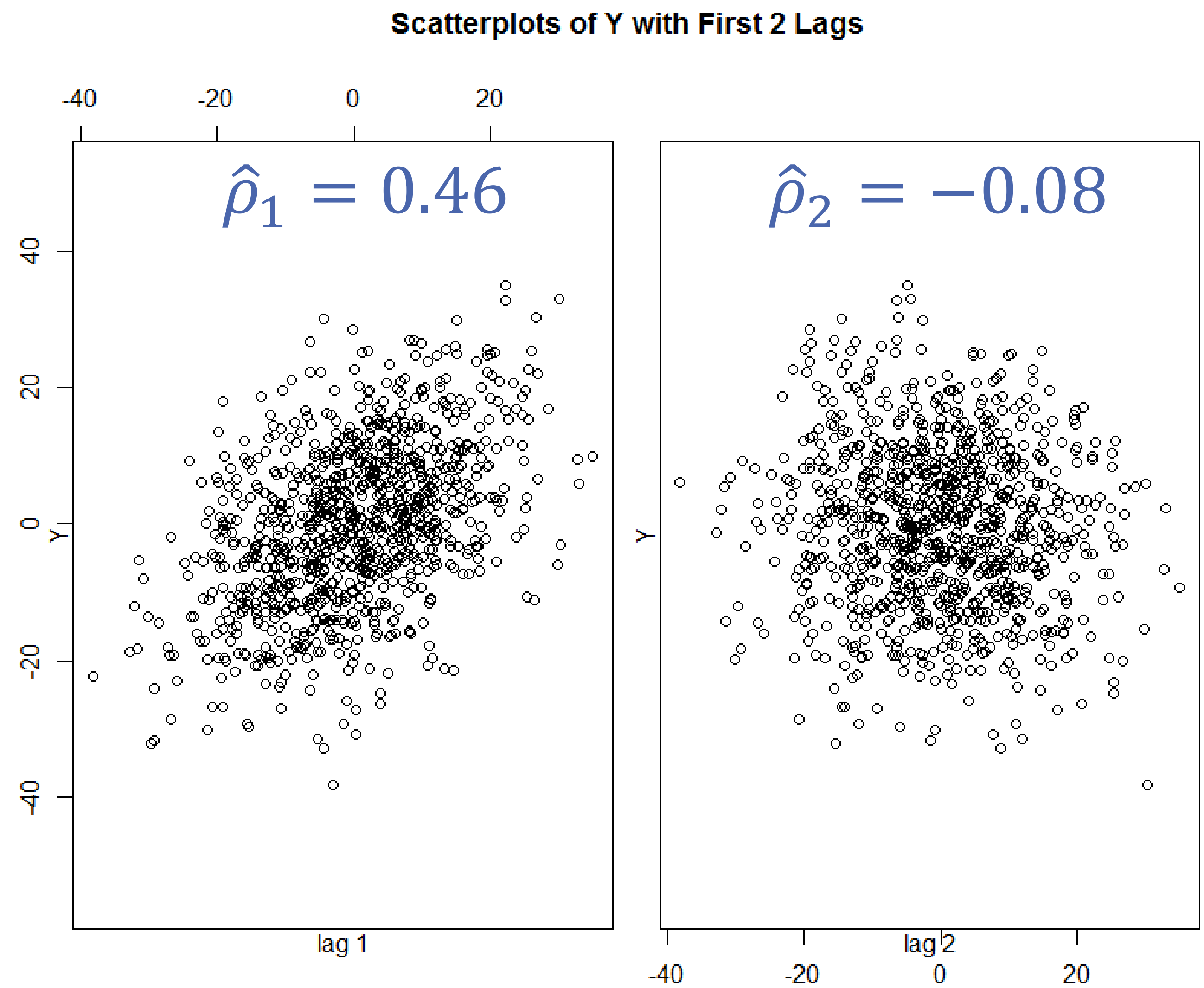
Autocorrelation Function

t	Y_t	Y_{t-1}	Y_{t-2}
1	20	.	.
2	2	20	.
3	16	2	20
4	-3	16	2
5	-14	-3	16
6	-28	-14	-3
…	…	…	
999	0	29	17
1000	-19	0	29

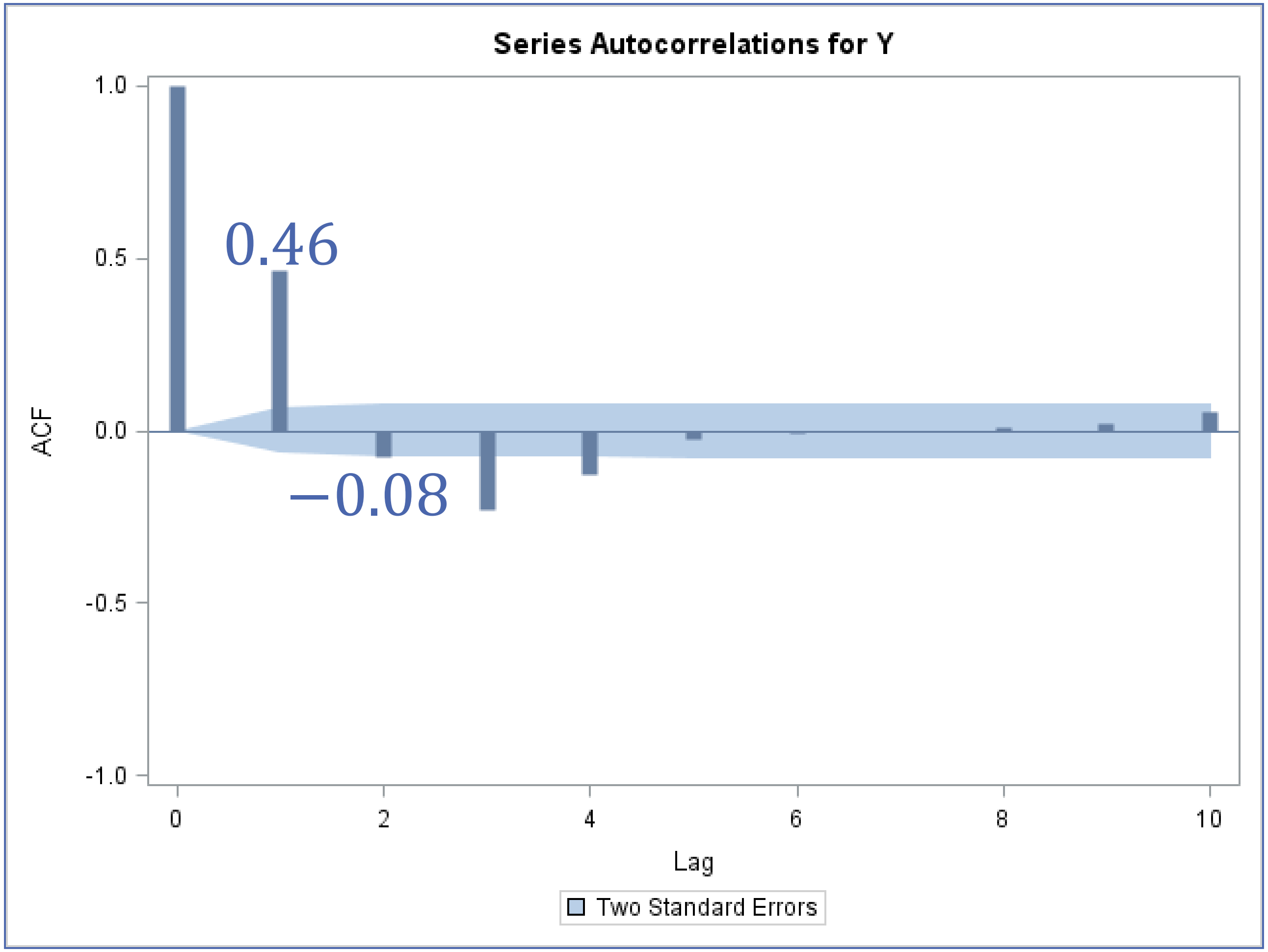
$$\hat{\rho}_1 = 0.46$$

$$\hat{\rho}_2 = -0.08$$

Autocorrelation Function



Autocorrelation Function



Autocorrelation Function

- Suppose that the first autocorrelation value (ACF(1)) is significant.
- This implies that two consecutive time points are related to each other.
  - March is related to April, April is related to May, etc.
  - Monday is related to Tuesday, Tuesday is related to Wednesday, etc.

Autocorrelation Function

- This relationship can be both in a positive and negative direction:
  - Positive – High Mondays imply high Tuesdays
  - Negative – High Mondays imply low Tuesdays

- This same relationship goes for all lags of the autocorrelation function.

Partial Autocorrelation Function

- *Partial autocorrelation* is the correlation between two sets of observations, from the same series, that are separated by *k* points in time, **after adjusting for all previous (1, 2, …, *k*-1) autocorrelations**.
- Partial autocorrelations are conditional correlations.
- The partial autocorrelation function (PACF) is the function of all partial autocorrelations (between two **sets of observations** Y_t and Y_{t-k}) across time (for all values of *k*).

$$\phi_k = \mathrm{Corr}(Y_t, Y_{t-k} \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k-1})$$

Partial Autocorrelation Function

t	Y_t	Y_{t-1}	Y_{t-2}
1	20	.	.
2	2	20	.
3	16	2	20
4	-3	16	2
5	-14	-3	16
6	-28	-14	-3
…	…	…	
999	0	29	17
1000	-19	0	29

Partial Autocorrelation Function

t	Y_t	Y_{t-1}	Y_{t-2}
1	20	.	.
2	2	20	.
3	16	2	20
4	-3	16	2
5	-14	-3	16
6	-28	-14	-3
…	…	…	
999	0	29	17
1000	-19	0	29

$\hat{\rho}_1 = 0.46$

$\hat{\phi}_1 = 0.46$

No time points in between to influence results!

Partial Autocorrelation Function

t	Y_t	Y_{t-1}	Y_{t-2}
1	20	.	.
2	2	20	.
3	16	2	20
4	-3	16	2
5	-14	-3	16
6	-28	-14	-3
…	…	…	
999	0	29	17
1000	-19	0	29

$\hat{\rho}_1 = 0.46$

$\hat{\phi}_1 = 0.46$

$\hat{\rho}_2 = -0.08$

$\hat{\phi}_2 = -0.37$

Must remove influence of time point in between!

Partial Autocorrelation Function

- The partial autocorrelation for the ***k*** ***k^{th}* lag** is calculated from the following regression:

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_k Y_{t-k} + e_t$$

Partial Autocorrelation Function

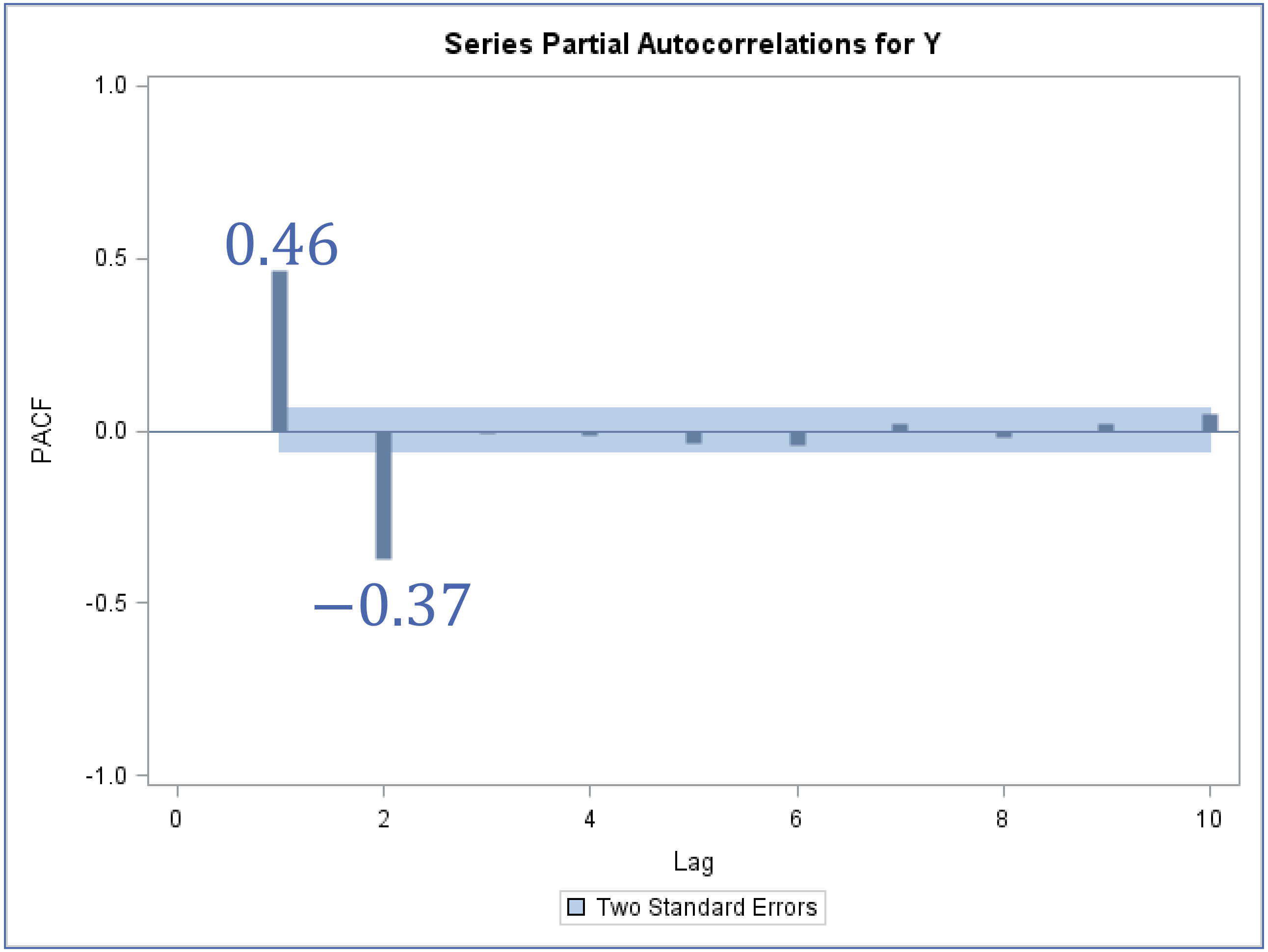
- The partial autocorrelation for the ***k**^th^* **lag** is calculated from the following regression:

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_k Y_{t-k} + e_t$$

- For example, the 2nd partial autocorrelation (ϕ_2) is estimated from:

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2}$$

Partial Autocorrelation Function



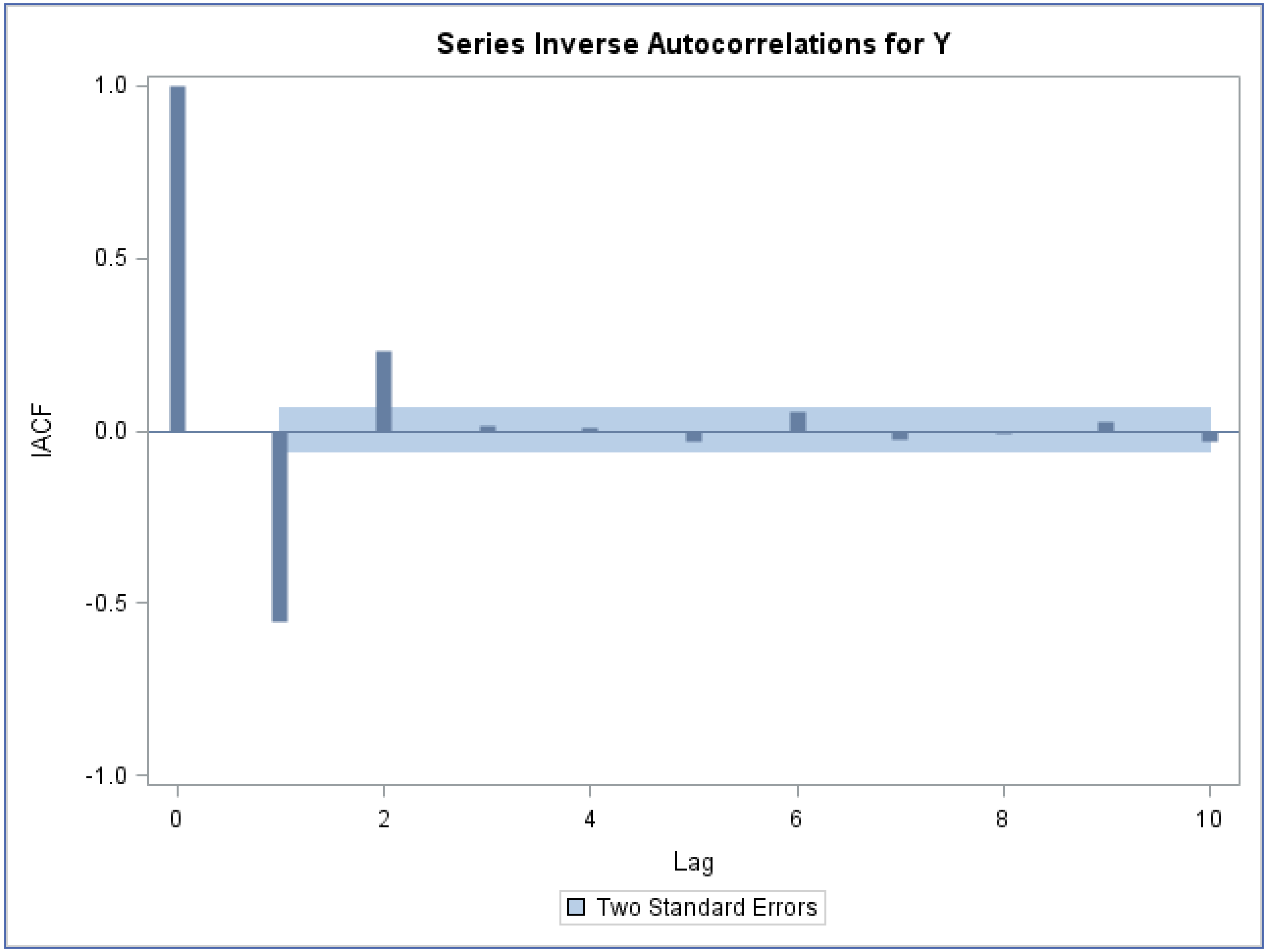
Partial Autocorrelation Function

- The partial autocorrelation functions tries to measure the direct relationship between two sets of observations, without the influence of other sets of time in between.

Inverse Autocorrelation Function

- *Inverse autocorrelation* is the correlation between two sets of observations, from the same series, that are separated by k points in time, **after adjusting for all previous (1, 2, …, *k*-1) autocorrelations**.
- Similar to the PACF, but without the same calculations.
- IACF typically has opposite signs as the PACF.

Inverse Autocorrelation Function



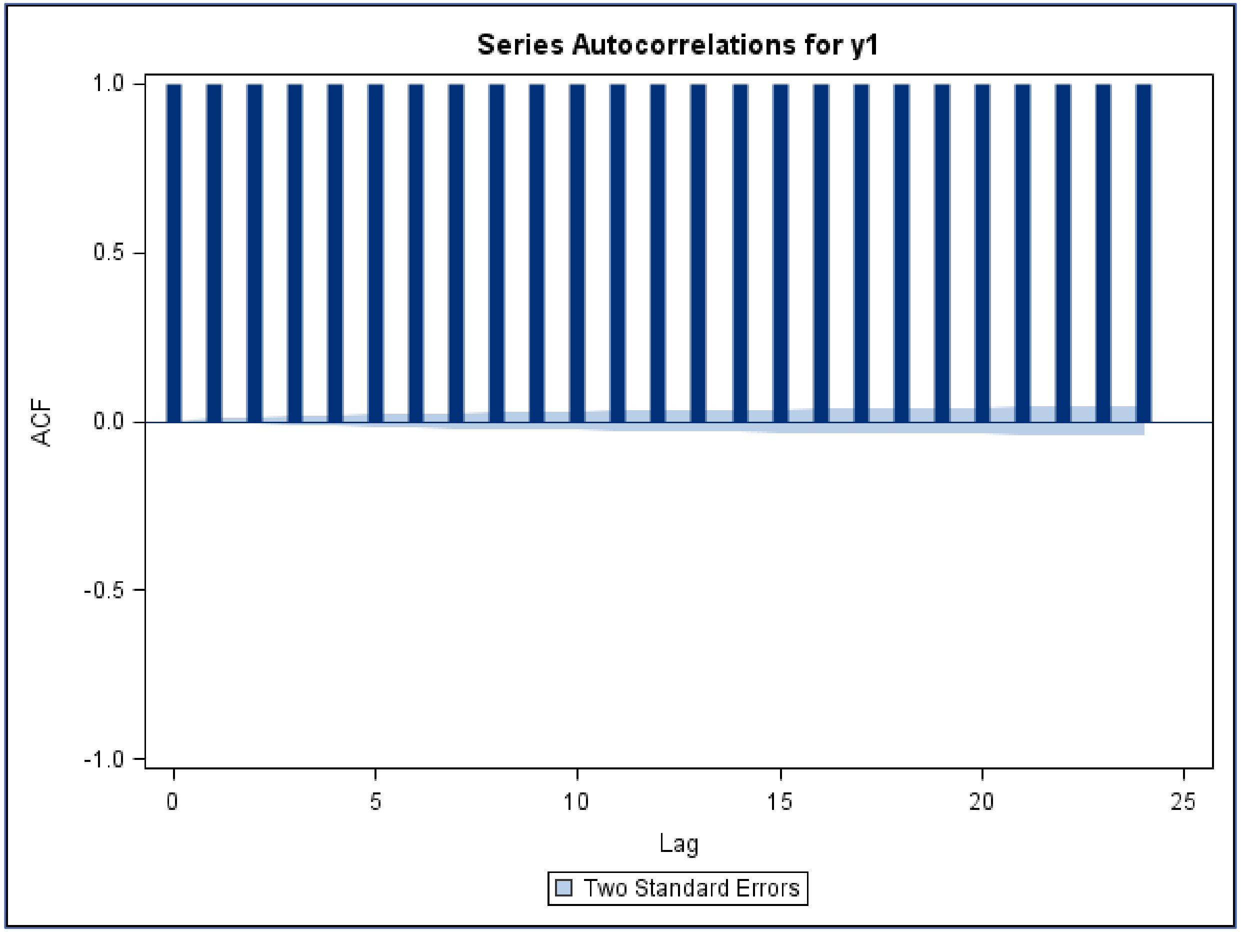
Correlation Functions – SAS

```
proc arima data=Time.AR2 plot(unpack)=all;
     identify var=y nlag=10;
run;
quit;
```

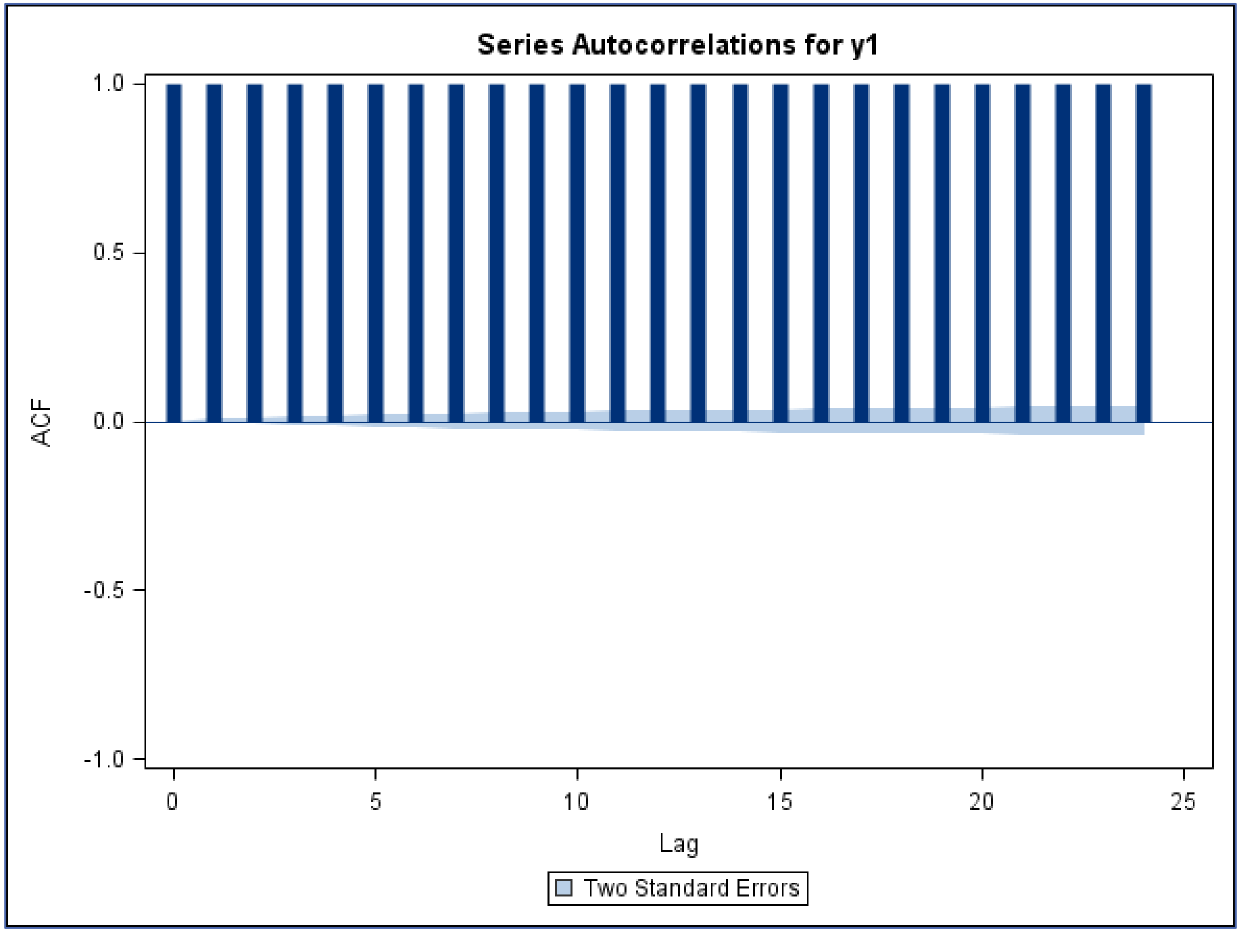
Correlation Functions – R

```
Acf(Y, lag=10)$acf
Pacf(Y, lag=10)$acf
```

RW – Autocorrelation Function

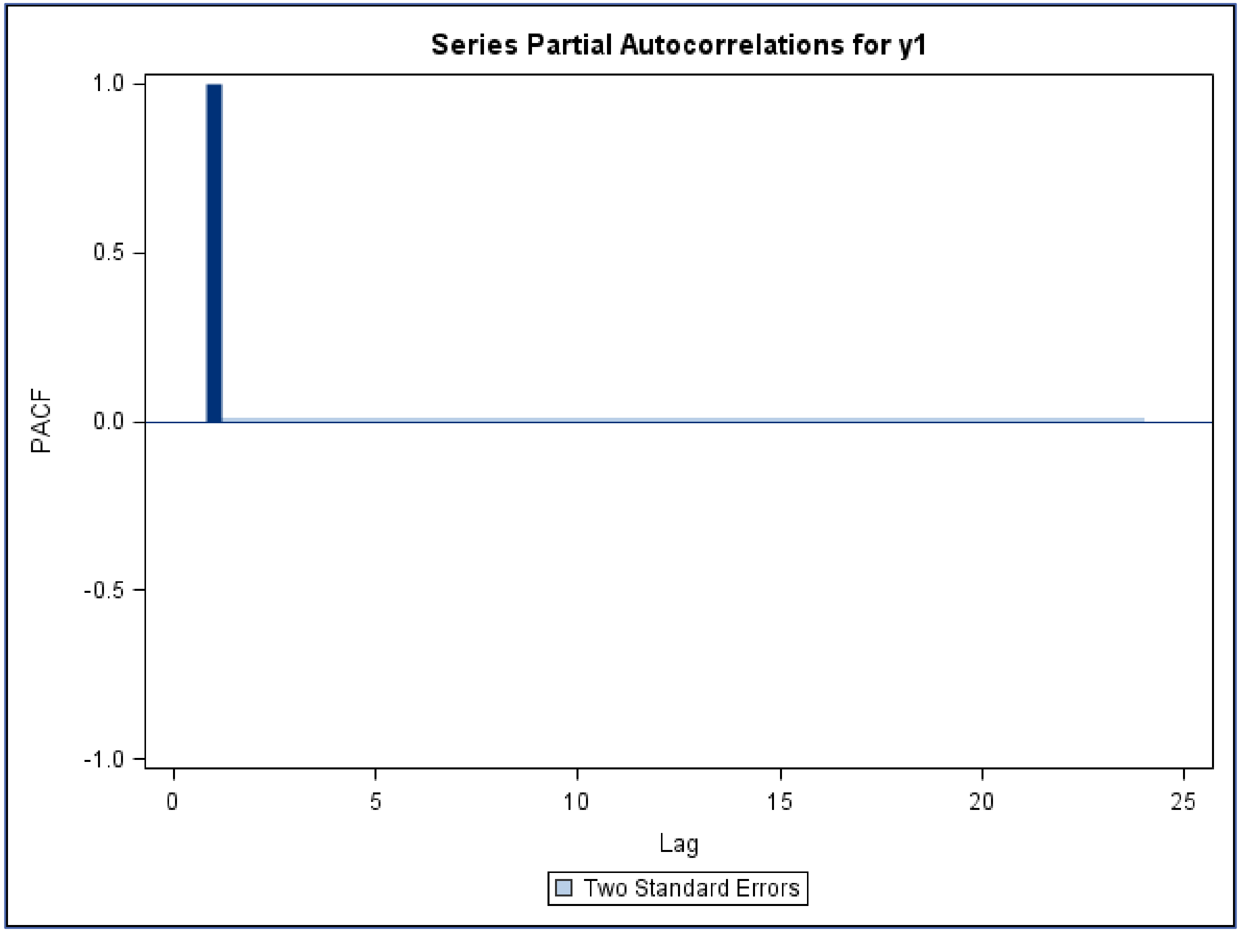


RW – Autocorrelation Function



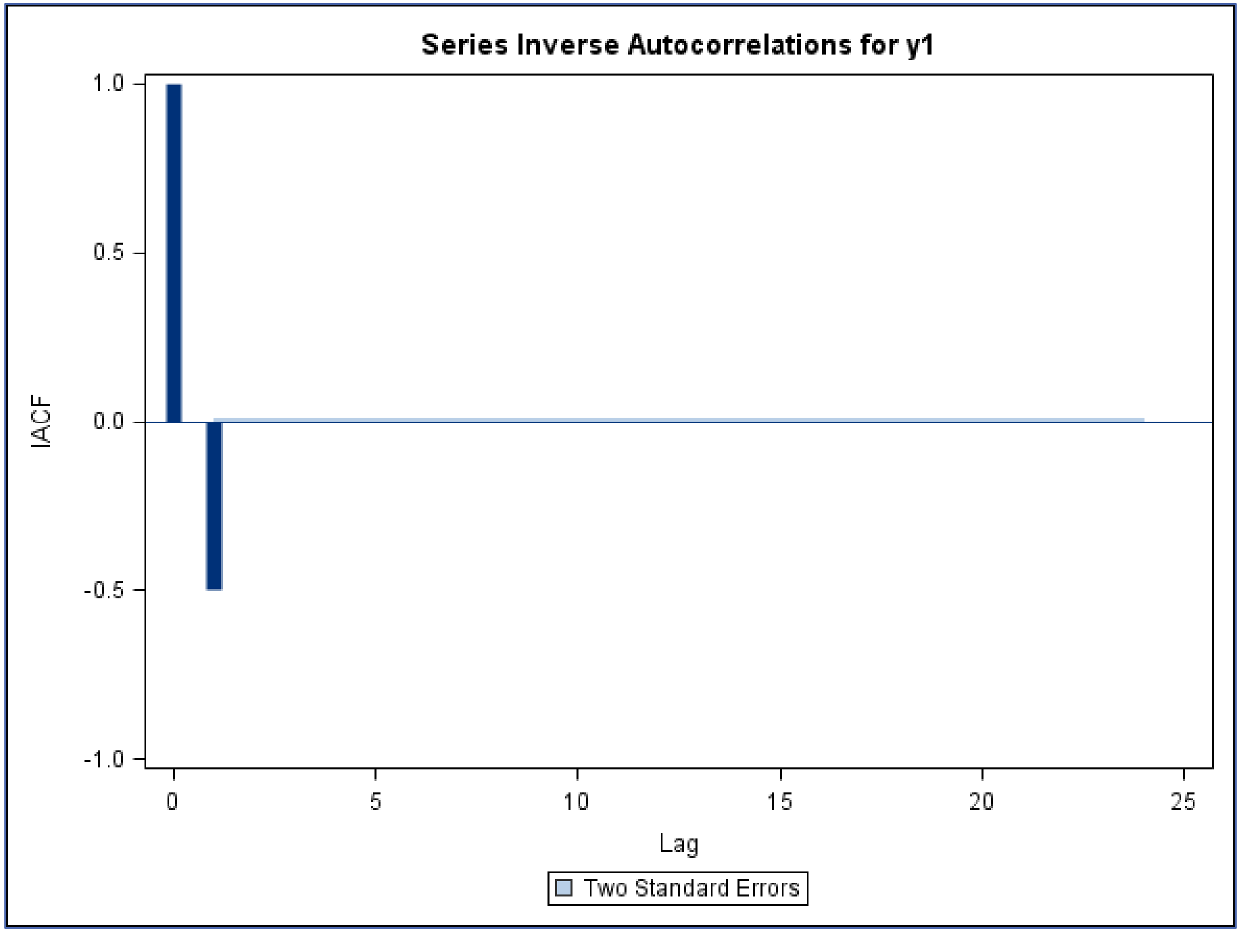
Notice that RW affect the correlation plots

RW– Partial Autocorrelation Function



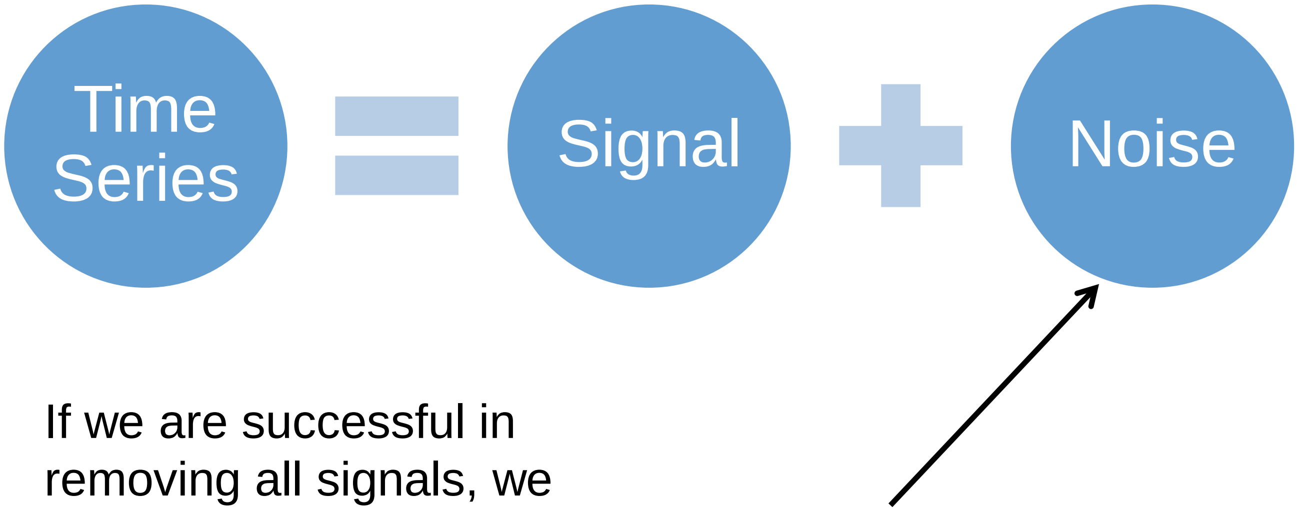
Only dependent on previous observation. Perfect correlation

RW – Inverse Autocorrelation Function



WHITE NOISE

Statistical Forecasting



If we are successful in removing all signals, we are left with independent errors.

White Noise

- A white noise time series follows a Normal distribution (or bell-shaped) with mean zero and positive, *constant* variance in which all observations are independent of each other.
- Autocorrelation and partial autocorrelation functions have a value close to zero at every time point (except for lag of 0).

White Noise

- The goal of modeling time series is to be left with white noise time series in the residuals.
- If the residuals still have a correlation structure, then more modeling can typically be done.
- How do we know when we are left with white noise at the end of the model?

Ljung-Box χ^2 Test for White Noise

- The Ljung-Box test may be applied to the original data or to the residuals after fitting a model.
- The null hypothesis is that the series has NO autocorrelation, and the alternative hypothesis is that one or more autocorrelations up to lag *m* are not zero.

$$\chi_m^2 = n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k}$$

Testing for White Noise – SAS

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

Testing for White Noise – R

```
White.LB <- rep(NA, 10)
for(i in 1:10){
  White.LB[i] <- Box.test(Y.Model$residuals, lag=i,
                          type="Lj", fitdf=2)$p.value
}

barplot(White.LB, main="Ljung-Box Test P-values",
        ylab="Probabilities", xlab="Lags")
abline(h=0.01, lty="dashed", col="black")
abline(h=0.05, lty="dashed", col="black")
```