INTRODUCTION TO FORECASTING & TIME SERIES STRUCTURE

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

TIME SERIES DATA

Objectives

- Introduce concepts behind modeling time series.
- Visually assess time series structures.
- Describe time series with respect to signal (explained variation) and noise (unexplained variation).

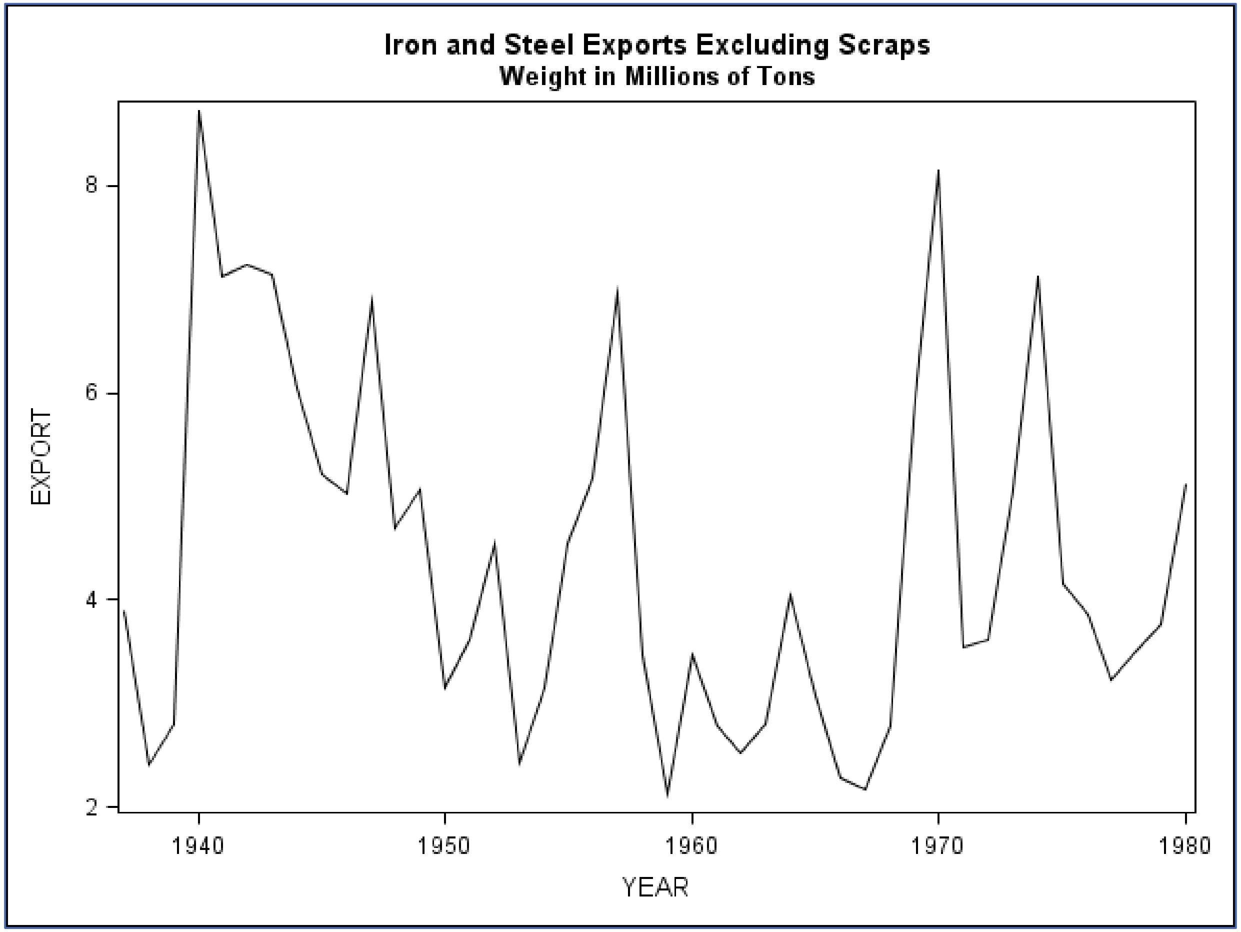
Time Series Data

- A time series is an ordered sequence of observations.
  - Ordering is typically through equally spaced time intervals.
  - Possibly through space as well.
- Used in a variety of fields:
  - Agriculture: Crop Production
  - Economics: Stock Prices
  - Engineering: Electric Signals
  - Meteorology: Wind Speeds
  - Social Sciences: Crime Rates

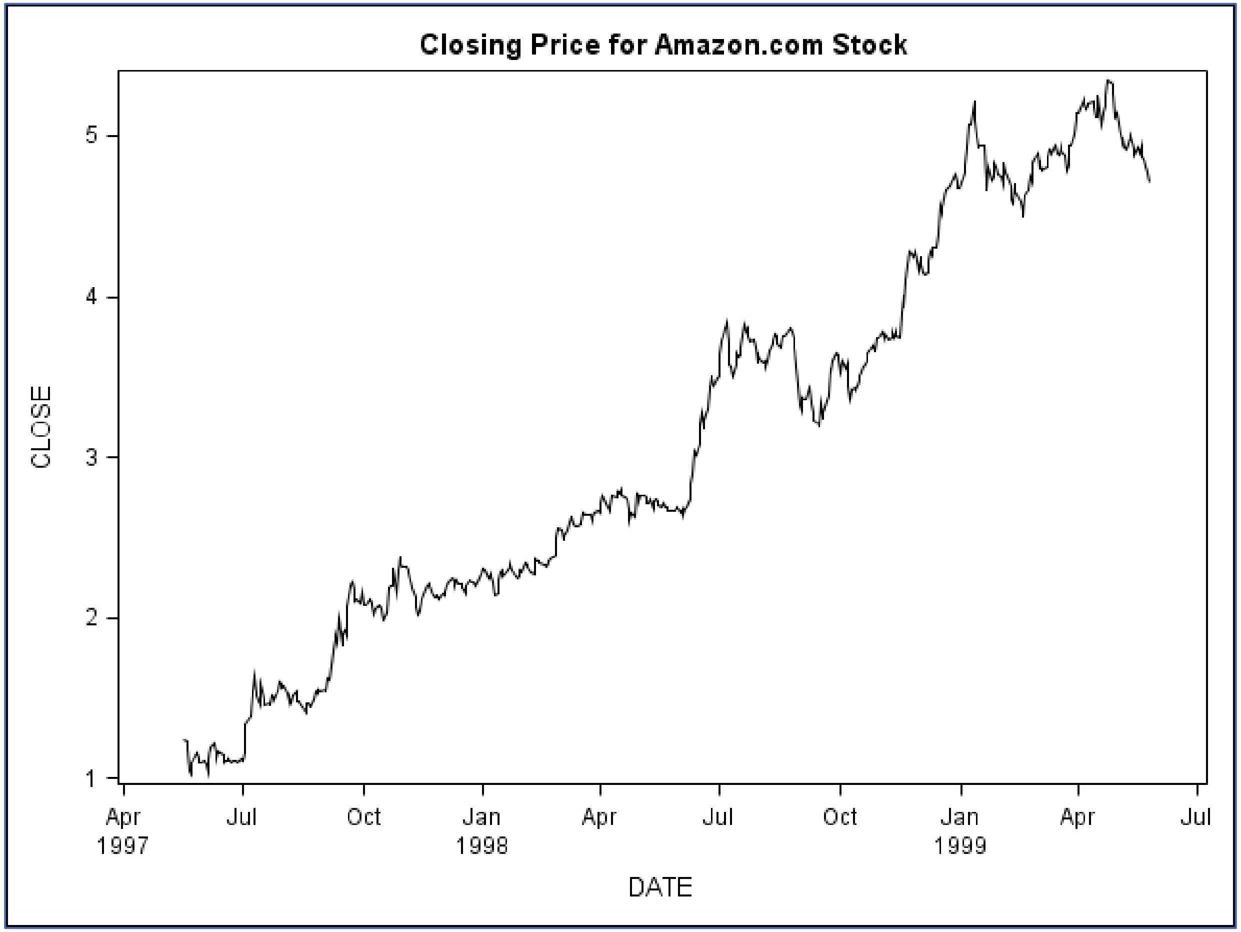
Time Series Data

- A time series might exhibit variation that can be explained with one of the following:
  - Trend
  - Seasonal variation
  - Cyclical patterns

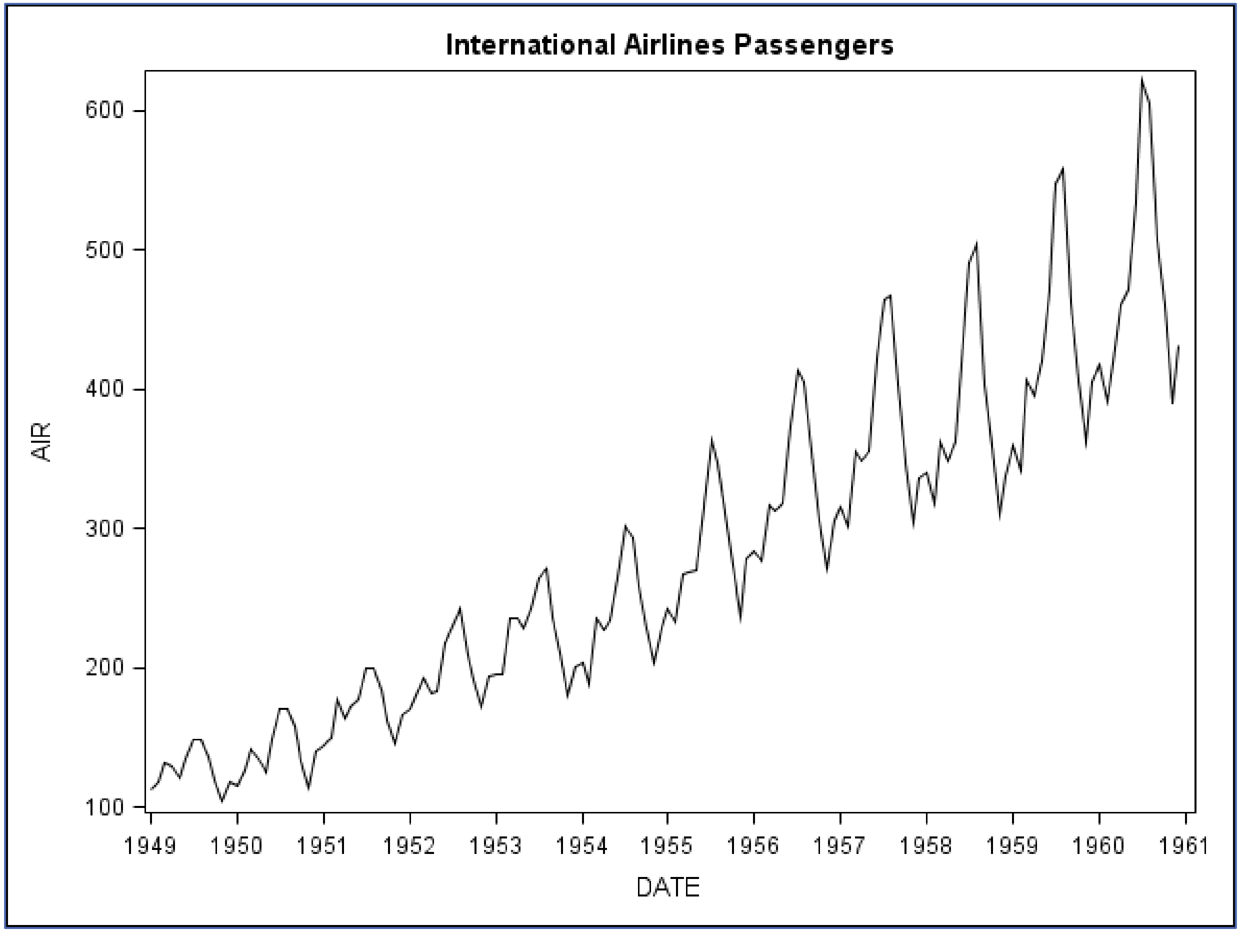
Example 1: Iron and Steel Exports



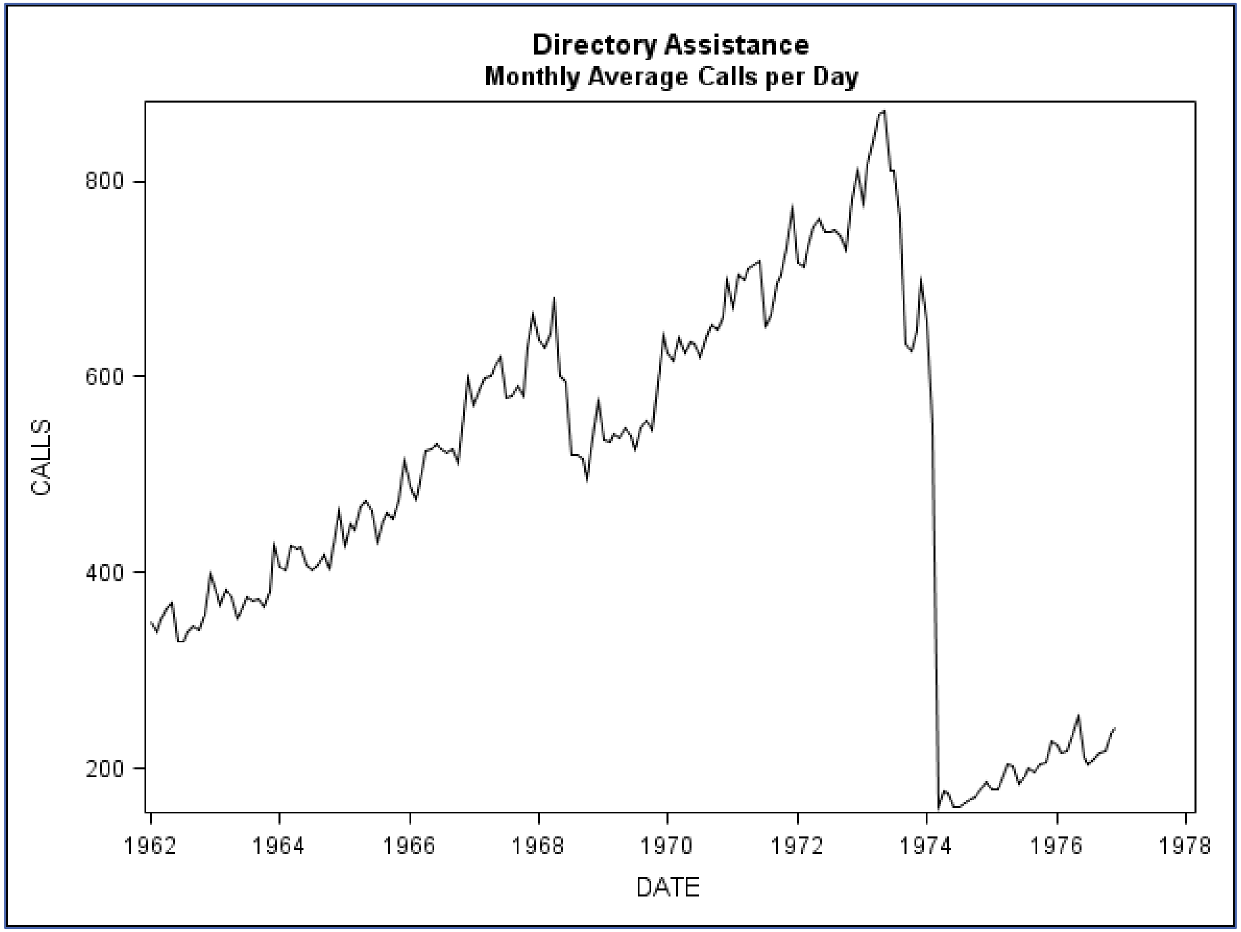
Example 2: Amazon.com Stock



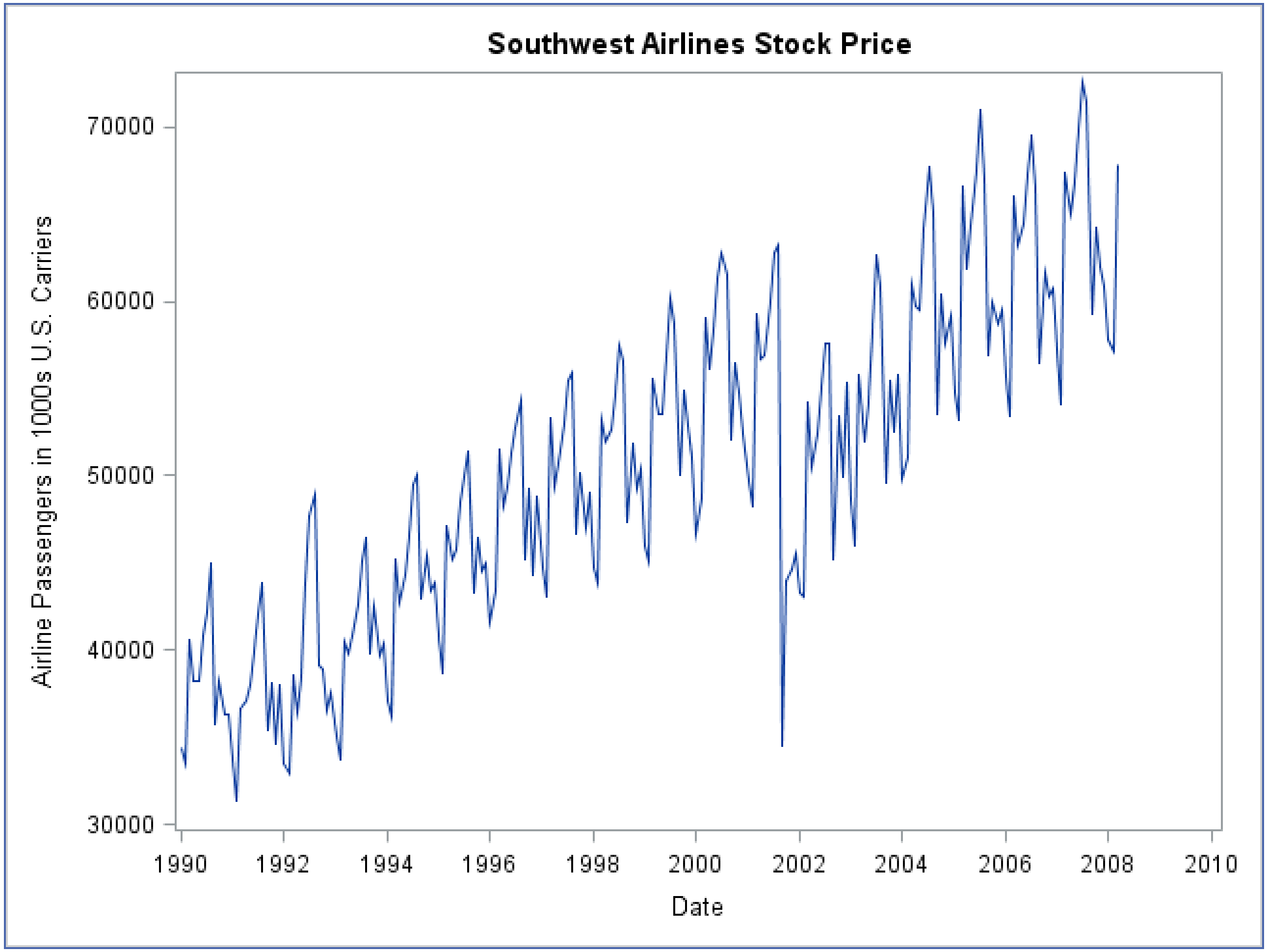
Example 3: Airlines Passengers



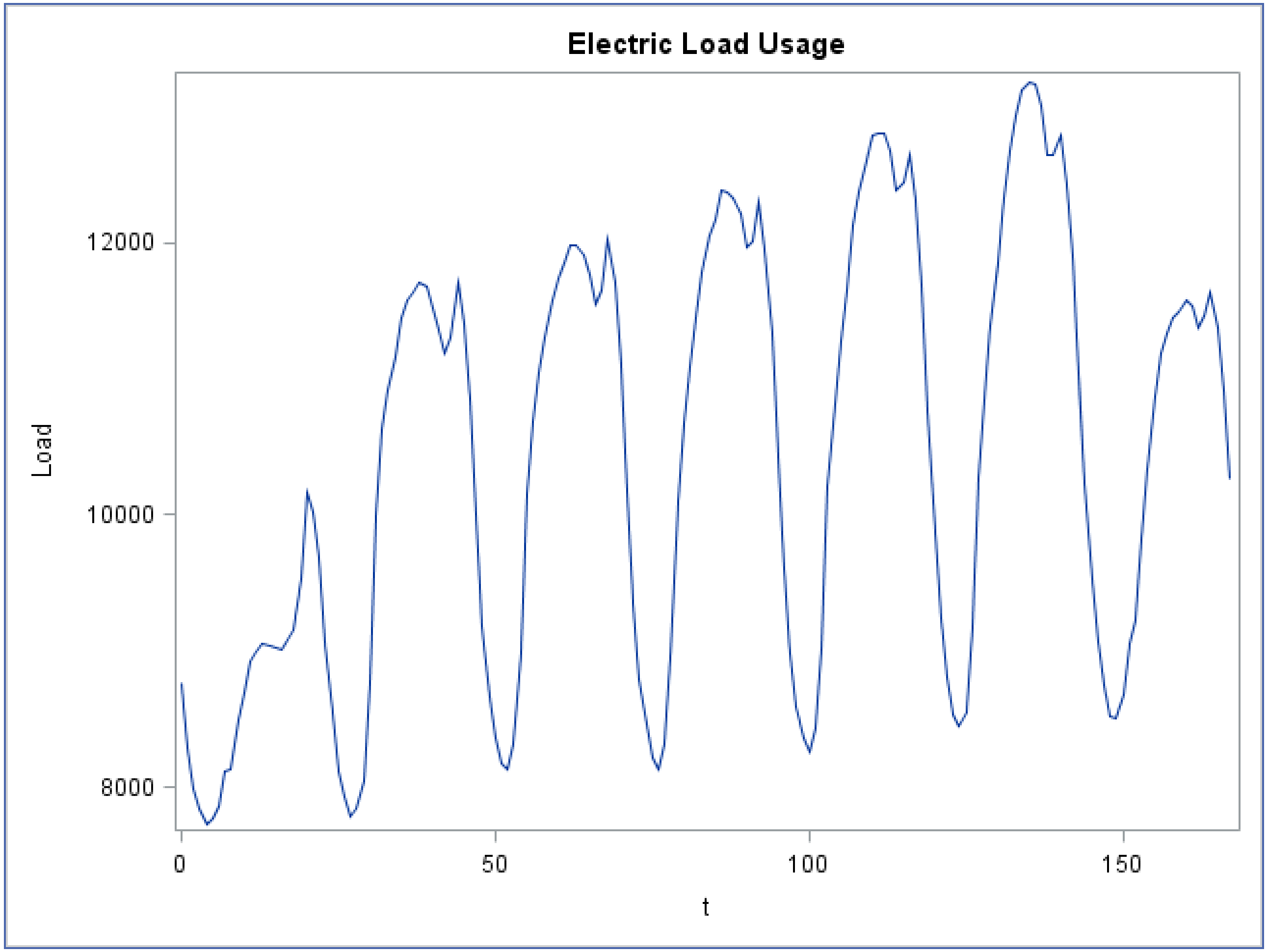
Example 4: Directory Assistance Calls



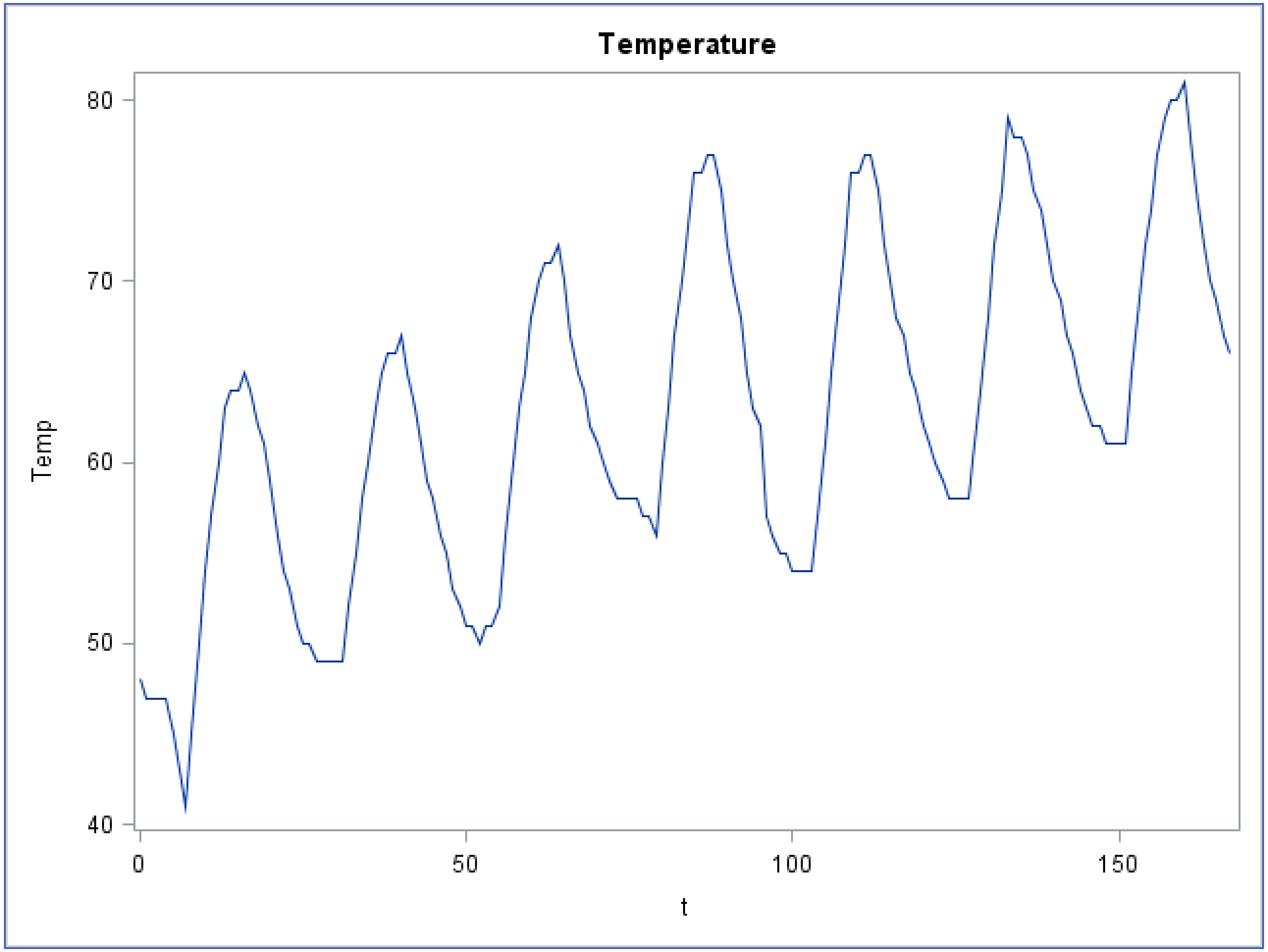
Example 5: Airline Passengers Again



Example 6: Electric Load



Example 6: Electric Load

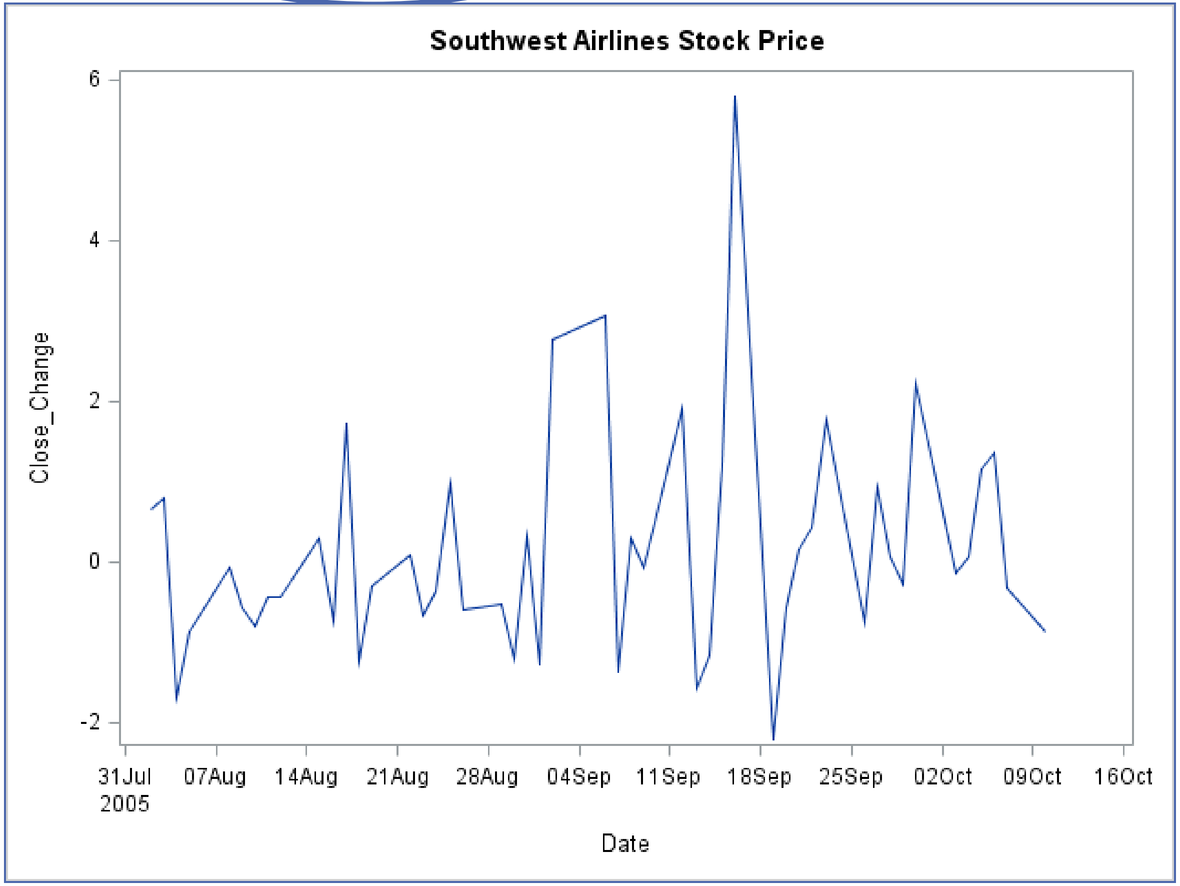


Explanation vs. Forecasting

- Similar to linear regression analysis, time series analysis is used for either explanation or forecasting.

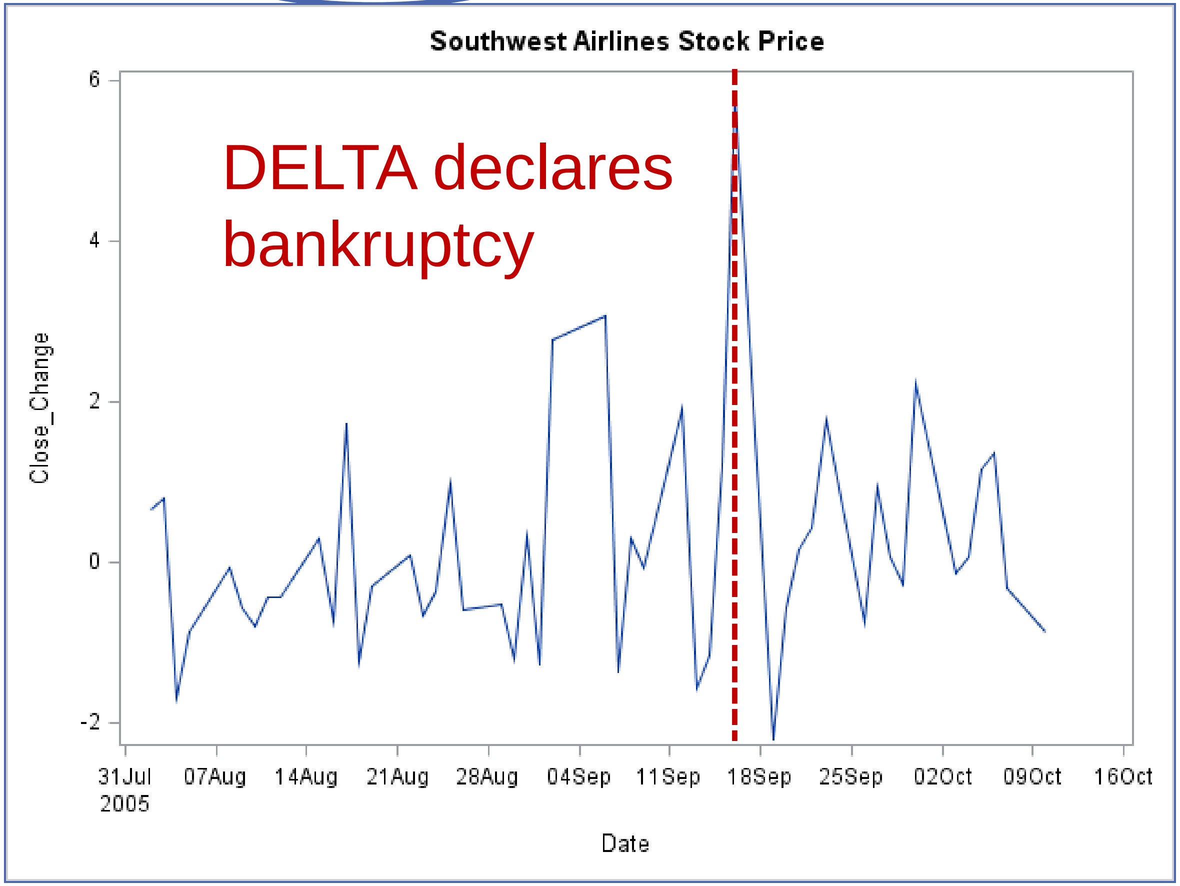
Explanation vs. Forecasting

- Similar to linear regression analysis, time series analysis is used for either explanation or forecasting.



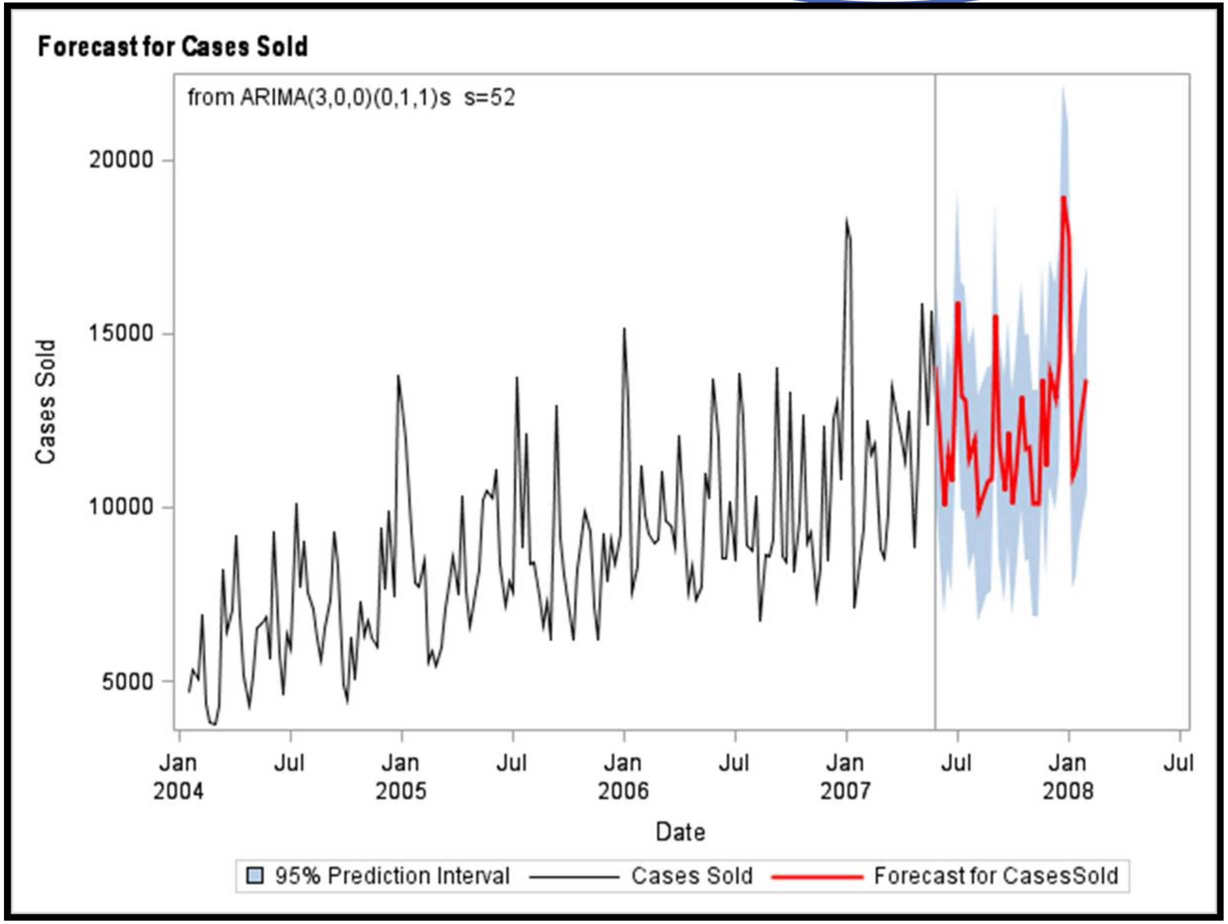
Explanation vs. Forecasting

- Similar to linear regression analysis, time series analysis is used for either explanation or forecasting.

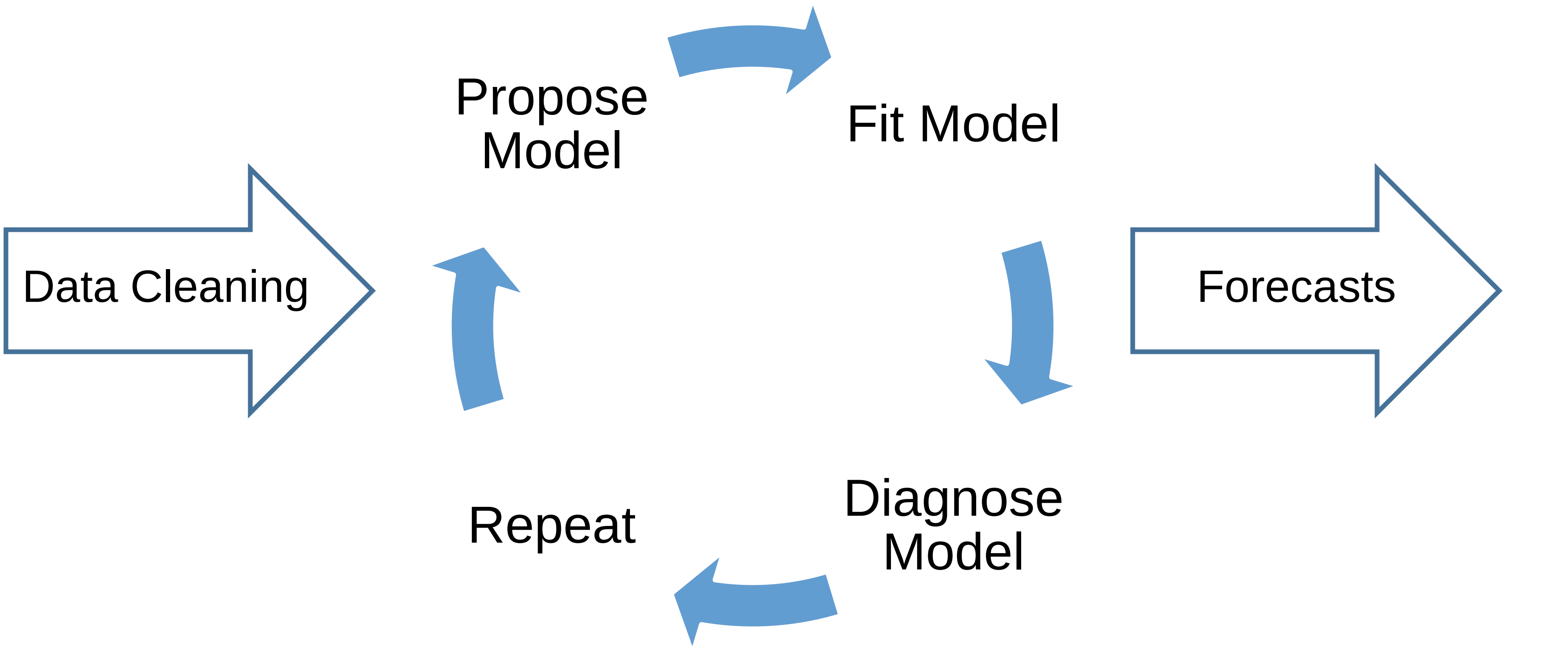


Explanation vs. Forecasting

- Similar to linear regression analysis, time series analysis is used for either explanation or forecasting.



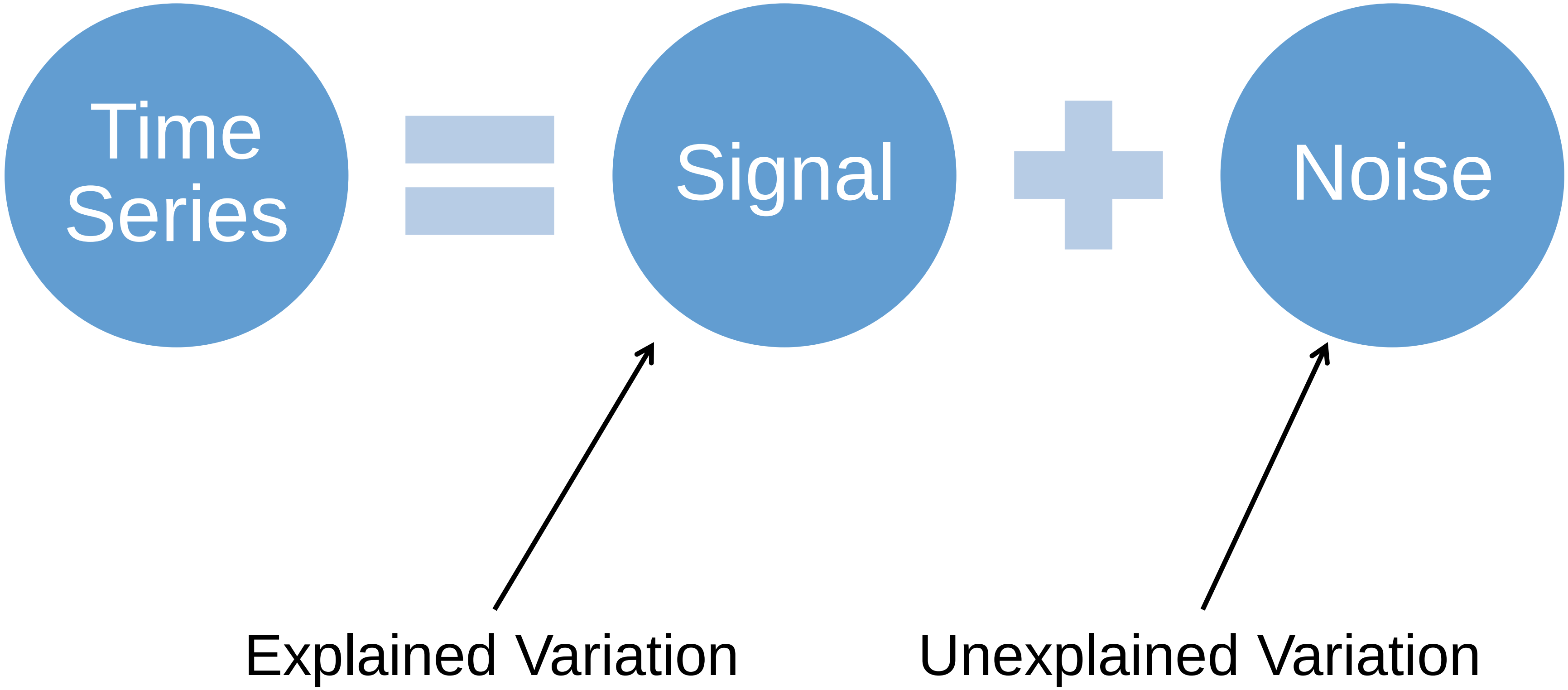
Forecasting Process



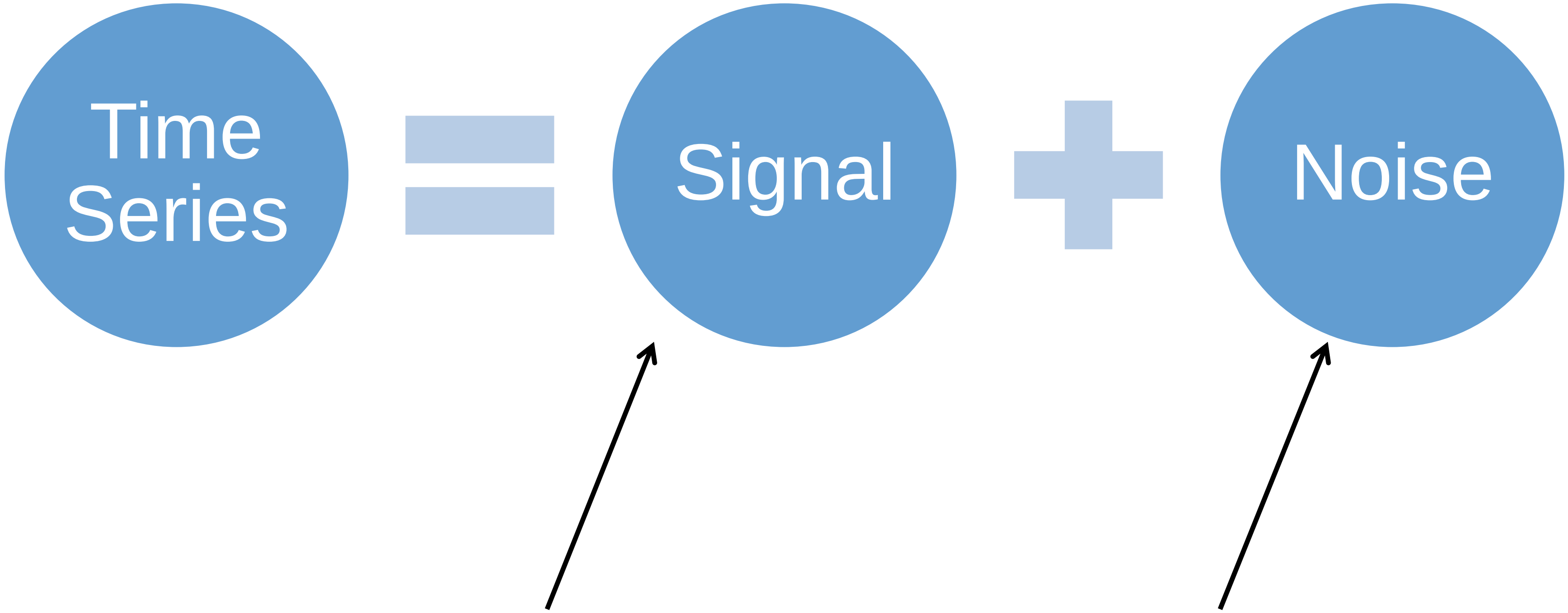
Statistical Forecasting

Time Series = Signal + Noise

Statistical Forecasting



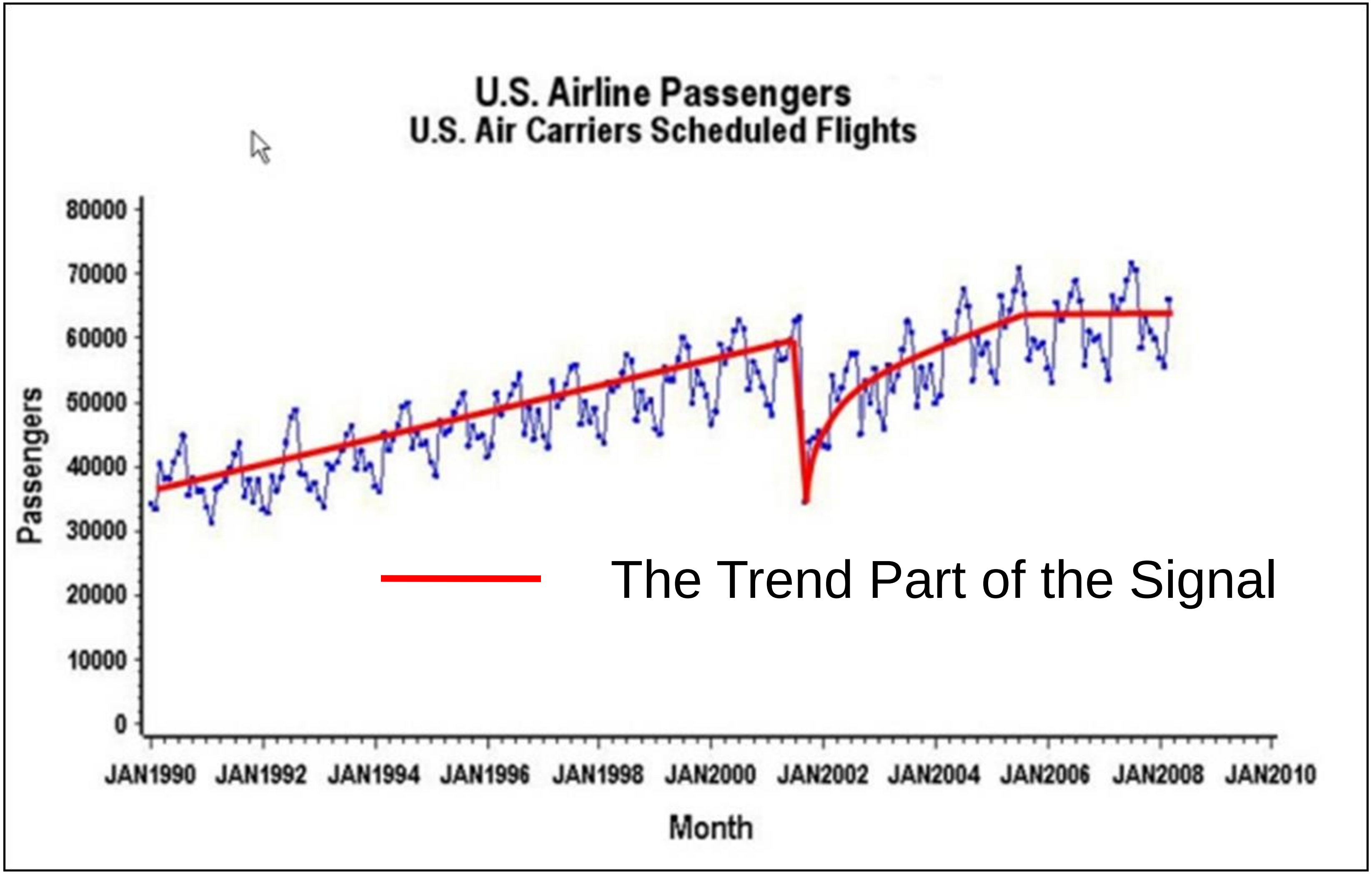
Statistical Forecasting



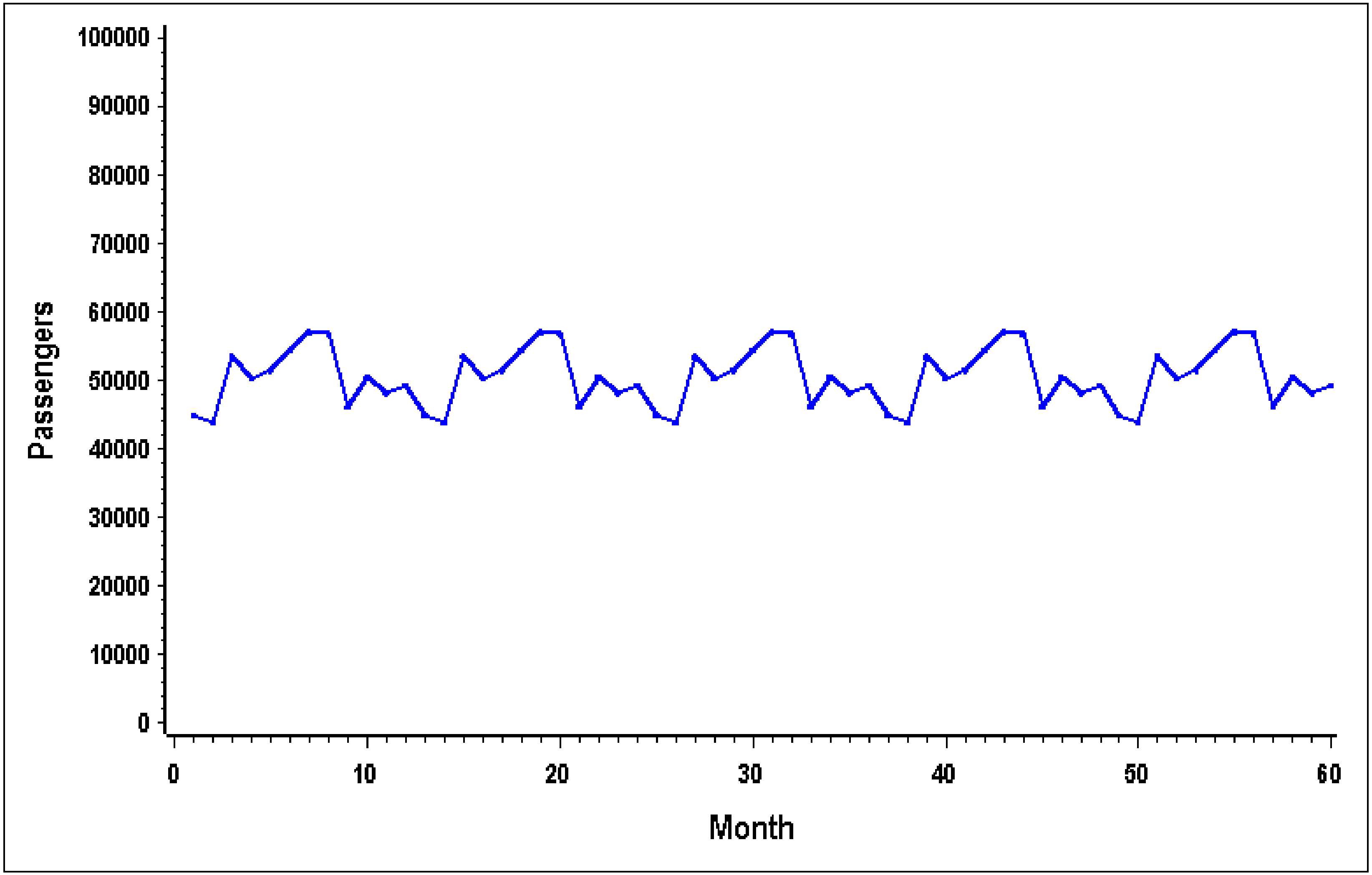
Forecasts extrapolate signal portion of model.

Confidence intervals account for uncertainty.

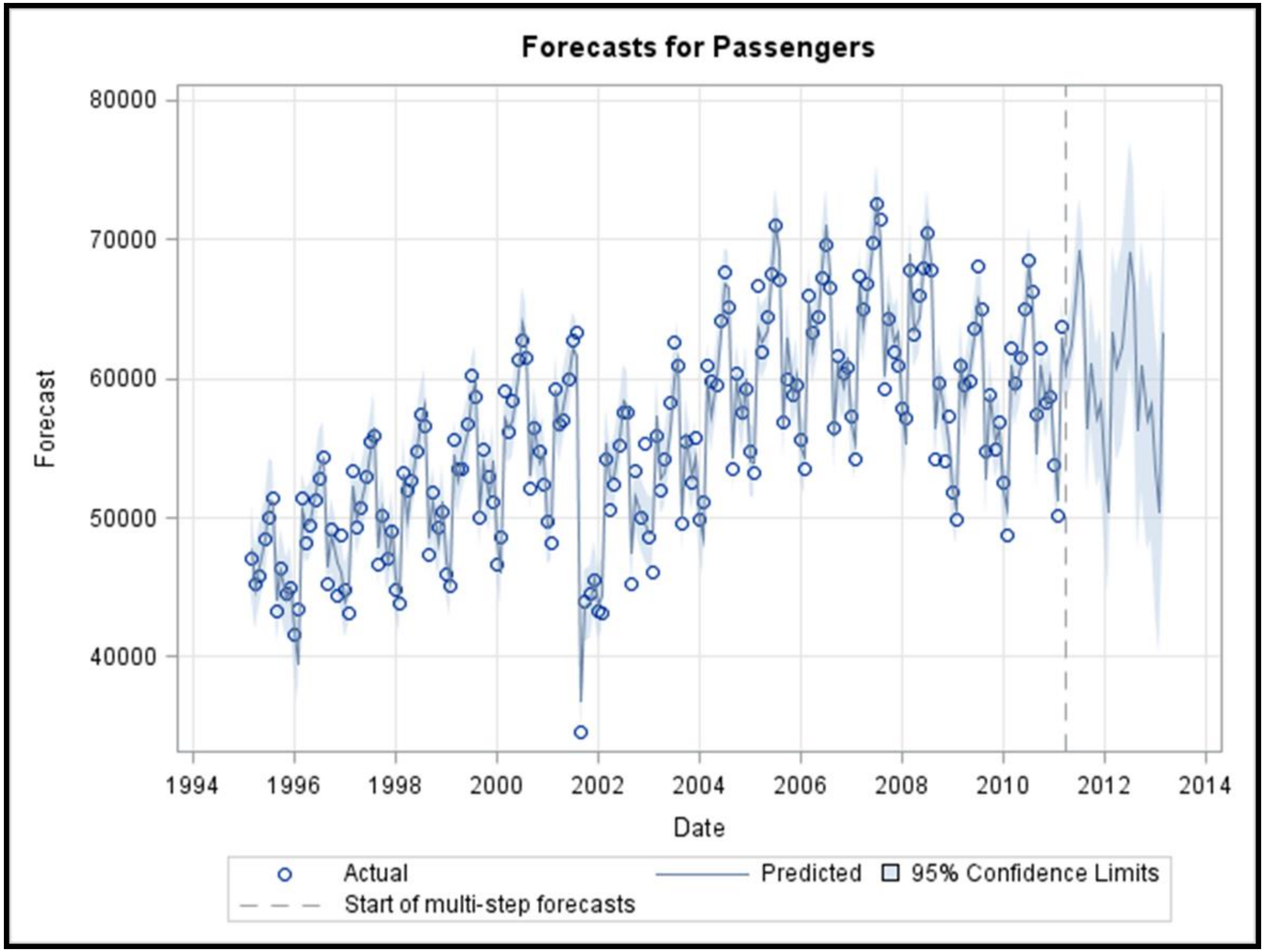
Time Series = Signal + Noise



Time Series = Signal + Noise



Time Series = Signal + Noise



CORRELATION FUNCTIONS

Objectives

- Examine the correlation structure in times series data.
- Define the autocorrelation, partial autocorrelation, and inverse autocorrelation function.
- Visually relate these functions to sets of data.

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

Characteristics of Time Series

- There are three common characteristics of a time series:
  1. Trend
  2. Seasonality
  3. Autocorrelation

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

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

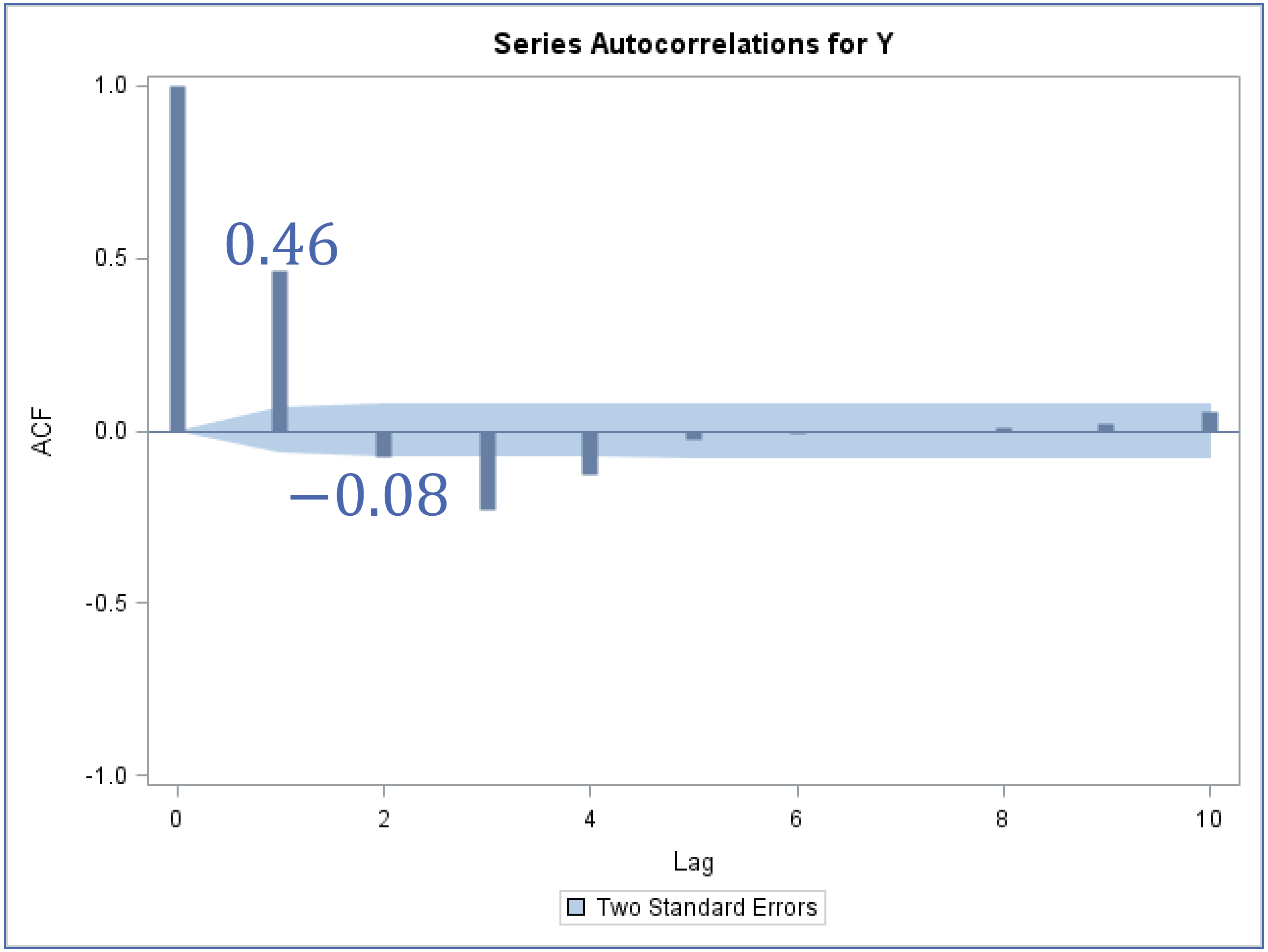
Autocorrelation Function

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

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

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

$\hat{\rho}_1 = 0.46$

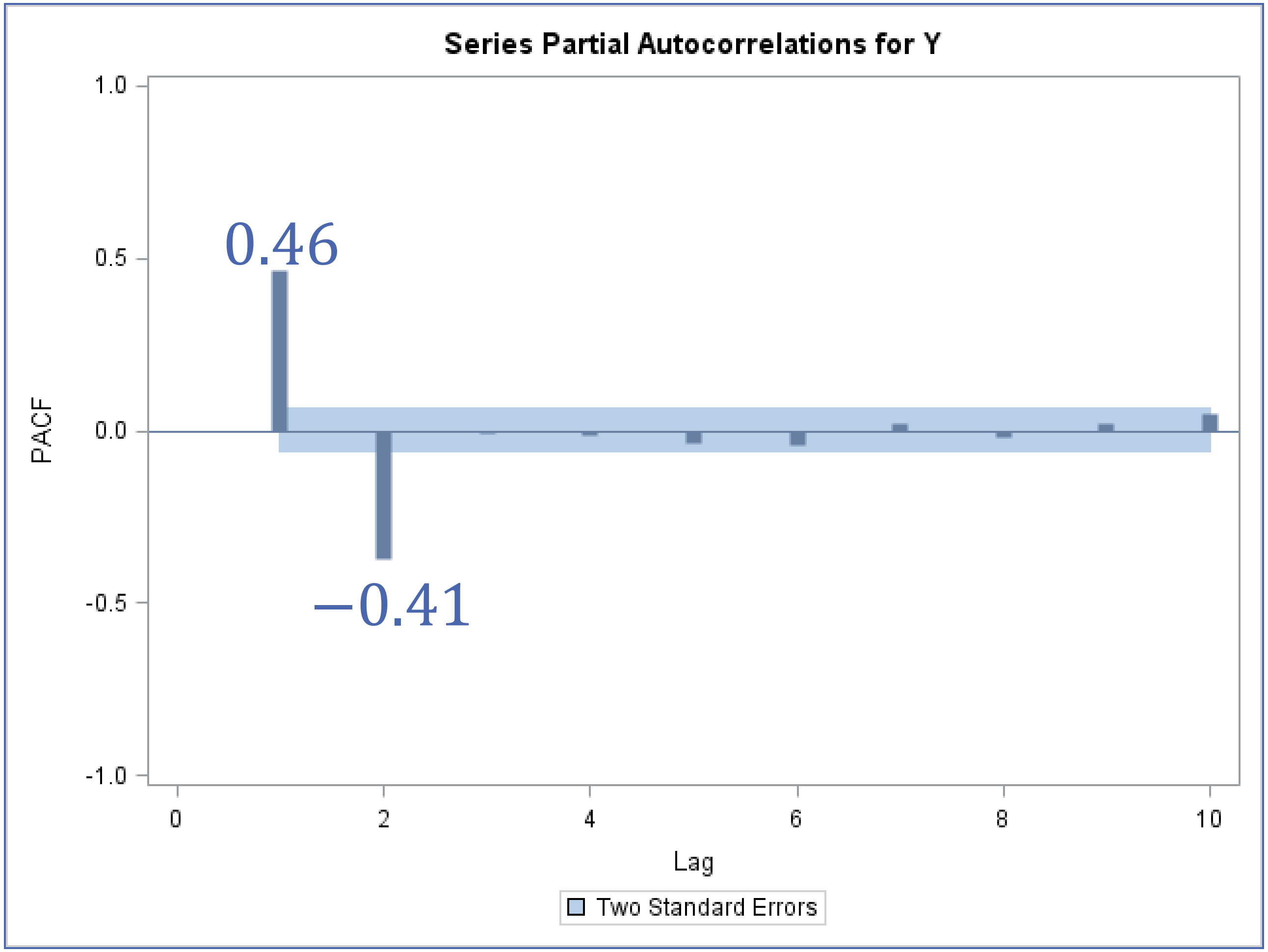
$\hat{\phi}_1 = 0.46$

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

$\hat{\phi}_2 = -0.41$

Must remove influence of time point in between!

Partial Autocorrelation Function



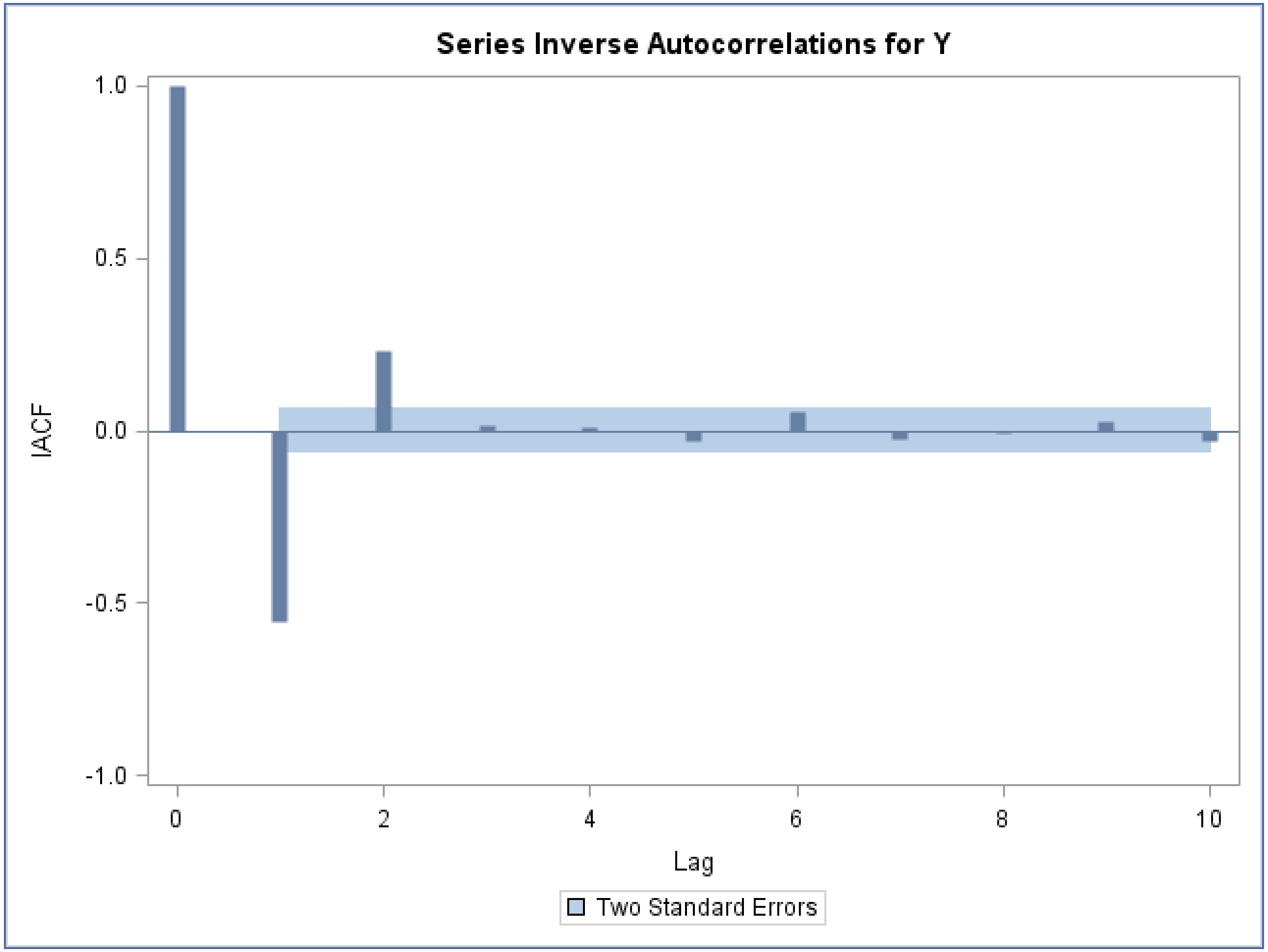
Partial Autocorrelation Function

- 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

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

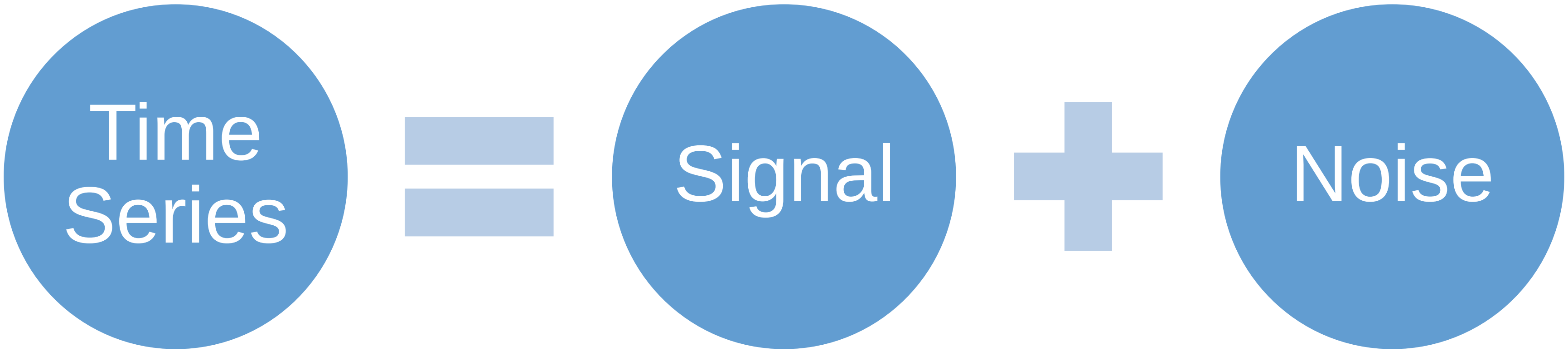
?

WHITE NOISE

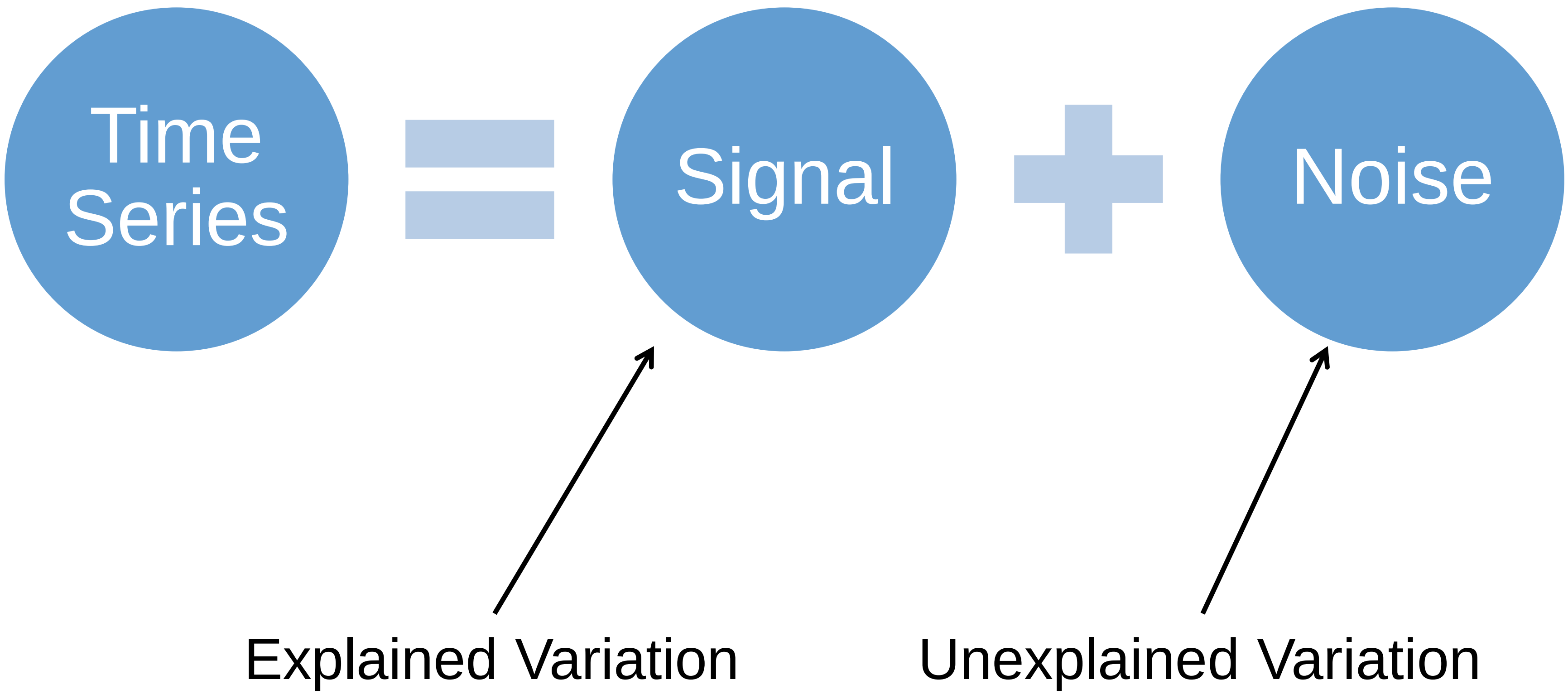
Time Series Dependencies

- A Time series is analyzed with an assumption that observations have a potential relationship that exists across time periods.
- We try to explain this dependency using previous signals across time.

Statistical Forecasting



Statistical Forecasting



Statistical Forecasting

Time Series = Signal + Noise

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

White Noise

White Noise

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

White Noise

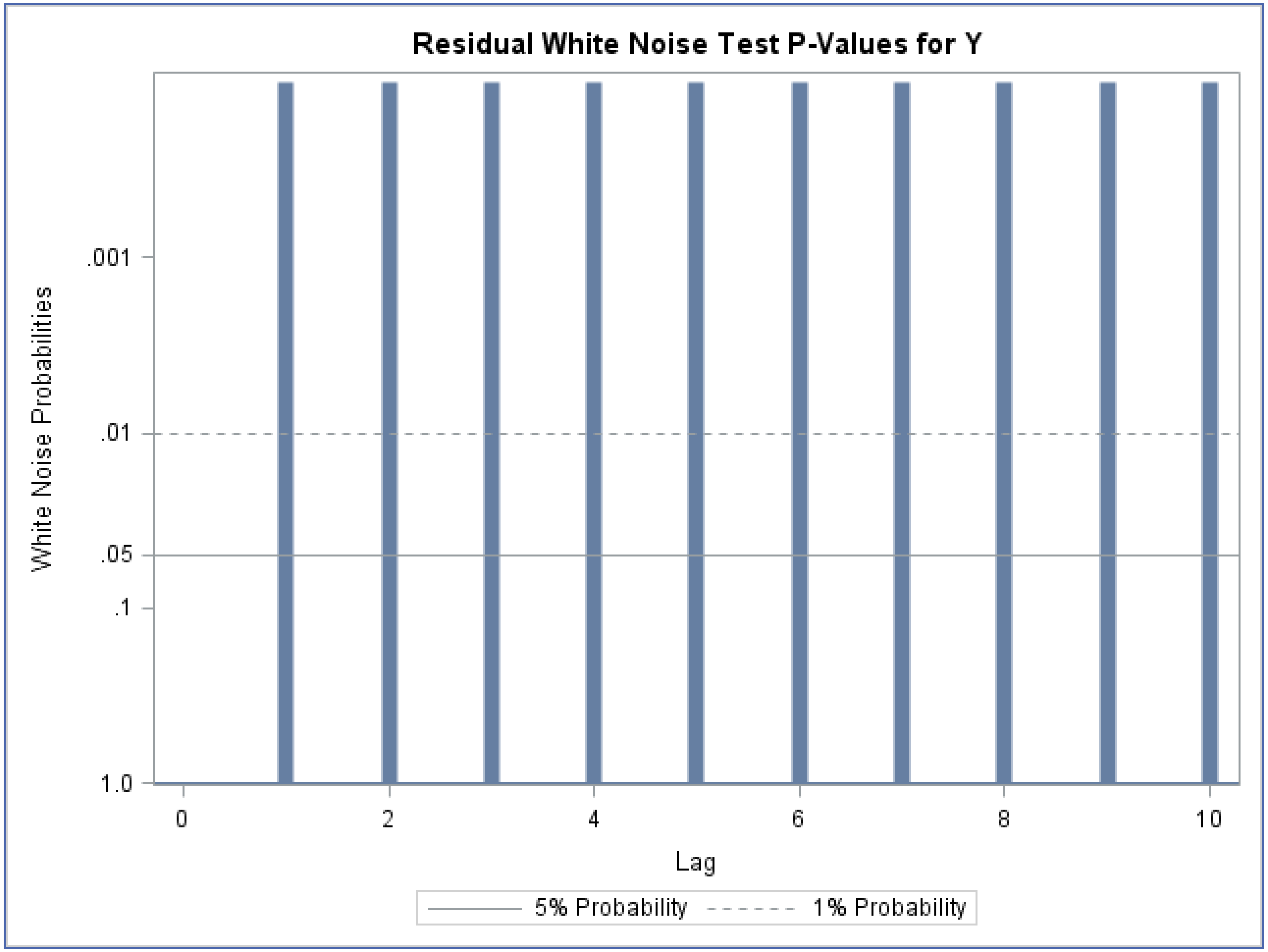
- The goal of modeling time series is to be left with white noise time series in the residuals.
- If the residuals still have a correlation structure, then more modeling can typically be done.
- How do we know when we are left with white noise at the end of the model?

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 is white noise, and the alternative hypothesis is that one or more autocorrelations up to lag *m* are not zero.

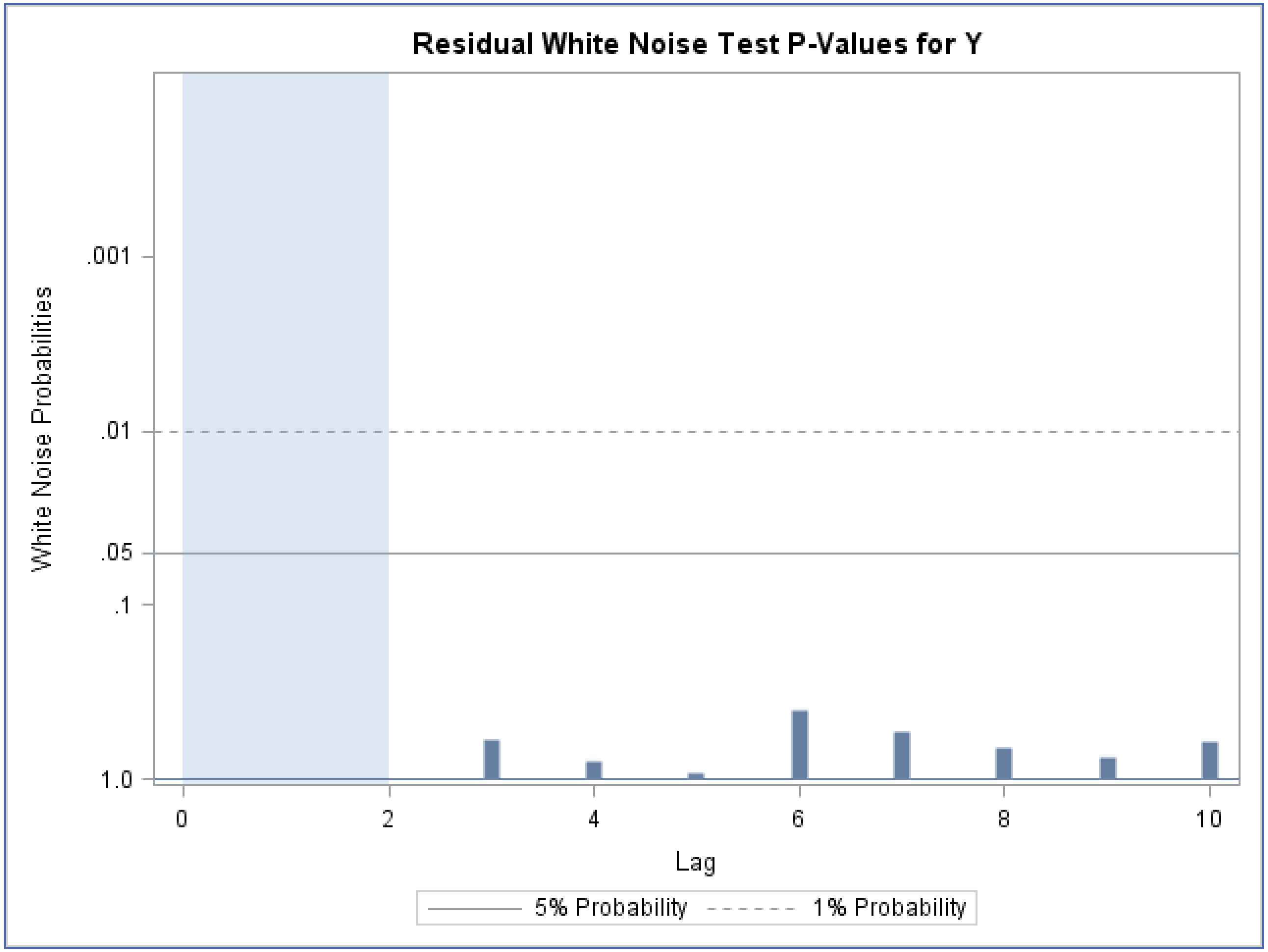
$$\chi_m^2 = n(n+2)\sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k}$$

Ljung-Box χ^2 Test for White Noise



Data itself

Ljung-Box χ^2 Test for White Noise



Residuals after modeling

Testing for White Noise

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

?

EVALUATING FORECASTS

Liability

- "Will you stake your reputation on the accuracy of these forecasts?"
- "**No, but I will stake my reputation on the methodology that was used to generate the forecasts.**"

- You have NO control over data accuracy, validity, or future events.

Forecasting Strategy

- Accuracy of forecasts depends on your definition of accuracy.
  - Different across different fields of industry.
- Good forecasts should have the following characteristics:
  - Be highly correlated with actual series values
  - Exhibit small forecast errors
  - Capture the important features of the original time series.

Judgment Forecasting

- Forecasting, just like most statistical methodologies, depends on the data used to derive the forecast.
- There are some aspects that data cannot account for that may lead to *judgment forecasting*.
  - Labor strike
  - New (unannounced) projects in region
  - Natural disasters
  - Lack of data availability

Accuracy vs. Goodness-of-Fit

- A diagnostic statistic calculated using the same sample that was used to build the model is a *goodness-of-fit* statistic.
- A diagnostic statistic calculated using a hold out sample that was not used in the building of the model is an *accuracy* statistic.

Hold-out Sample

- A hold out sample in time series analysis is different than cross-sectional analysis.
- The hold-out sample is always at the end of the time series, and doesn't typically go beyond 25% of the data.
- Ideally, an entire season should be captured in a hold-out sample.

Hold-out Sample

1. Divide the time series into two segments – training and validation (hold-out).
2. Derive a set of candidate models.
3. Calculate the chosen *accuracy* statistic by forecasting the validation data set.
4. Pick the model with the best accuracy statistic.

Model Diagnostic Statistics

1. Mean Absolute Percent Error:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{Y_t}$$

2. Mean Absolute Error:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|$$

Model Diagnostic Statistics

3. Symmetric Mean Absolute Percent Error:

$$\text{SMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|Y_t - \hat{Y}_t\right|}{\frac{1}{2}\left(Y_t + \hat{Y}_t\right)}$$

4. Mean Absolute Error Divided by Mean:

$$\text{MAE/Mean} = \frac{\frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|}{\bar{Y}}$$

Model Diagnostic Statistics

5. Akaike's Information Criterion

$$\mathrm{AIC} = -2\log(L) + 2k$$

$$\mathrm{AIC} = n\log\left(\frac{SSE}{n}\right) + 2k$$

Likelihood Based

6. Schwarz's Bayesian Information Criterion:

$$\mathrm{SBC} = -2\log(L) + k\log(n)$$

$$\mathrm{SBC} = n\log\left(\frac{SSE}{n}\right) + k\log(n)$$

Model Diagnostic Statistics

5. Akaike's Information Criterion

$$\mathrm{AIC} = -2\log(L) + 2k$$

$$\mathrm{AIC} = n\log\left(\frac{SSE}{n}\right) + 2k$$

Error Based

6. Schwarz's Bayesian Information Criterion:

$$\mathrm{SBC} = -2\log(L) + k\log(n)$$

$$\mathrm{SBC} = n\log\left(\frac{SSE}{n}\right) + k\log(n)$$