INTRODUCTION TO FORECASTING & TIME SERIES STRUCTURE

Dr. Susan Simmons

Institute for Advanced Analytics

TIME SERIES DATA

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

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

Date	Y
January 2000	23
February 2000	18
March 2000	20
April 2000	25
May 2000	21

Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

Date	Y
January 2000	23
February 2000	18
March 2000	20
April 2000	25
May 2000	21

Y_1

Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

Date	Y
January 2000	23
February 2000	18
March 2000	20
April 2000	25
May 2000	21

Y_2

Time Series Data

- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

Date	Y
January 2000	23
February 2000	18
March 2000	20
April 2000	25
May 2000	21

Y_3

Time Series Data

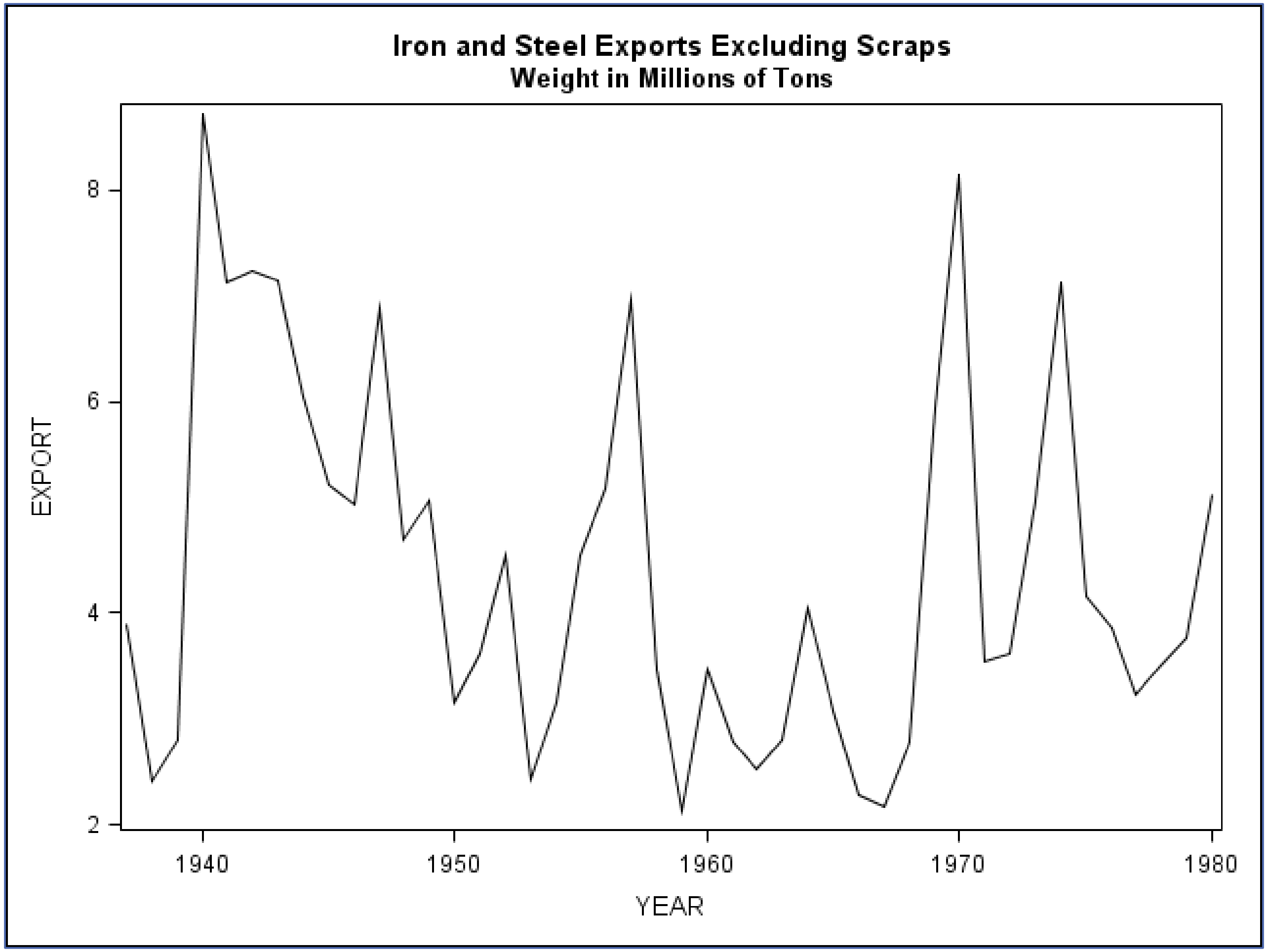
- We will begin our time series discussions with univariate time series (only one time series…one variable, we will call it Y).
- Multivariate time series will be in Fall 2.

Date	Y
January 2000	23
February 2000	18
March 2000	20
April 2000	25
May 2000	21

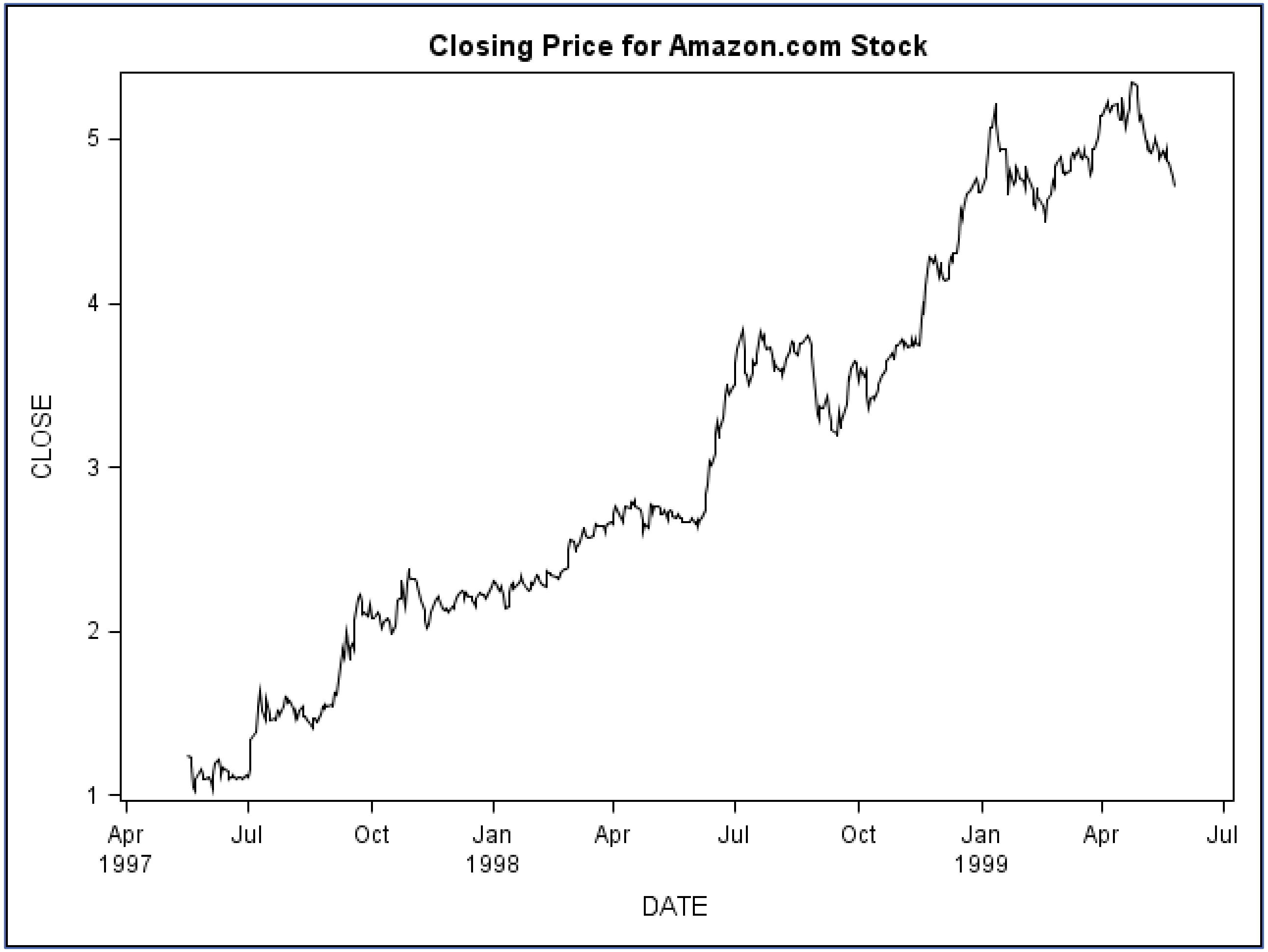
Y_3

Y_t

Example 1: Iron and Steel Exports

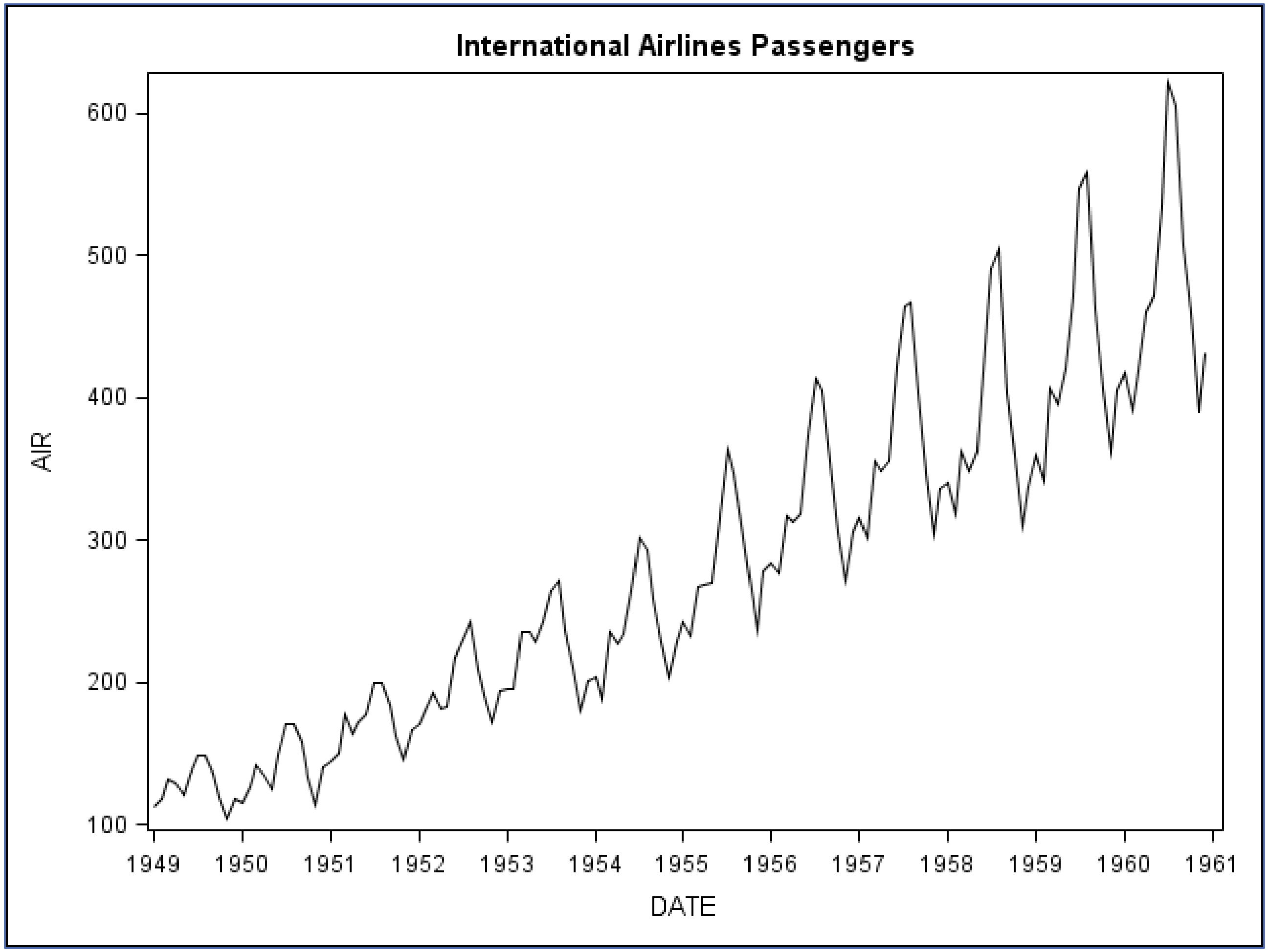


Example 2: Amazon.com Stock

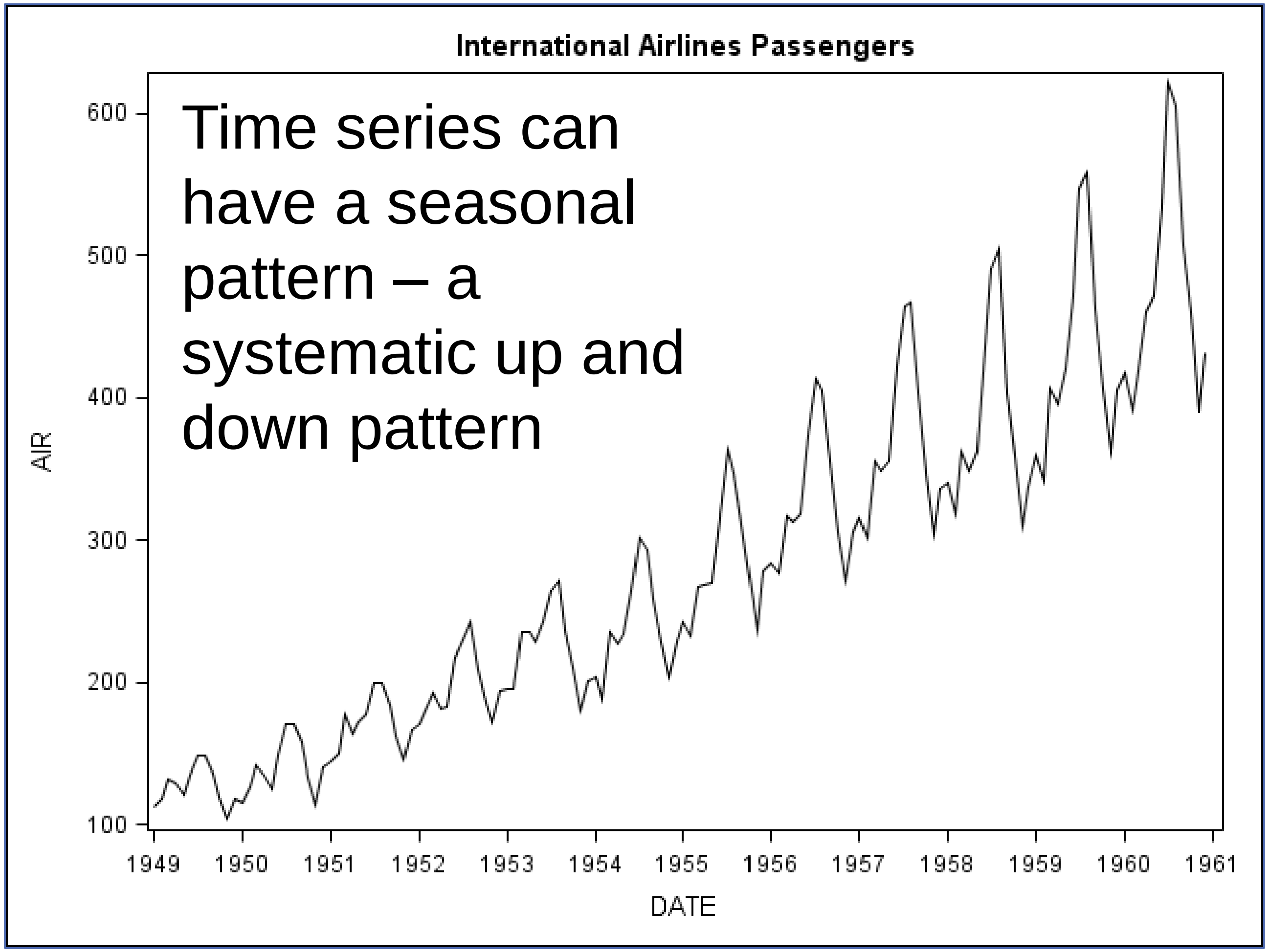


Example 2: Amazon.com Stock

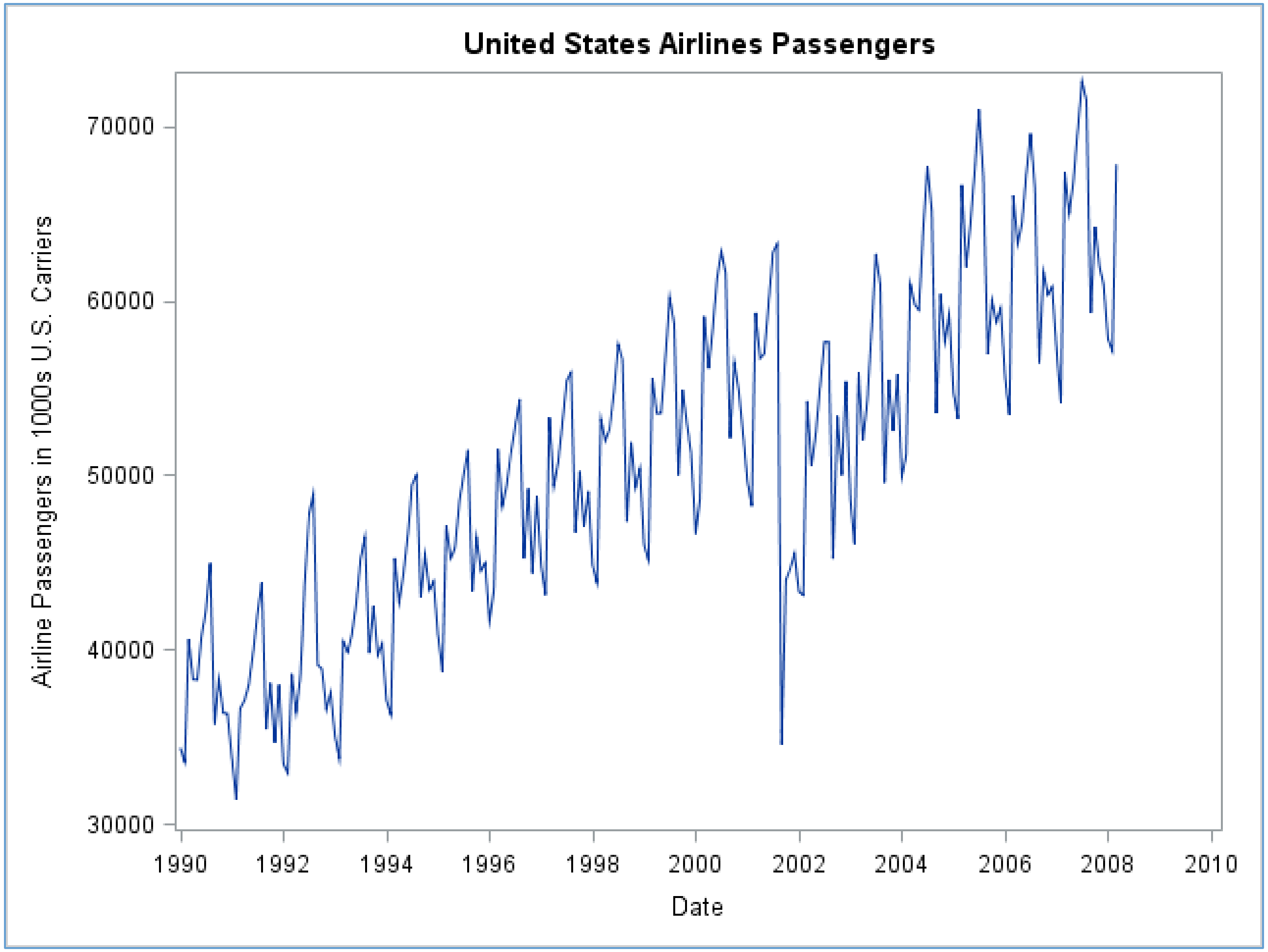
Example 3: Airlines Passengers



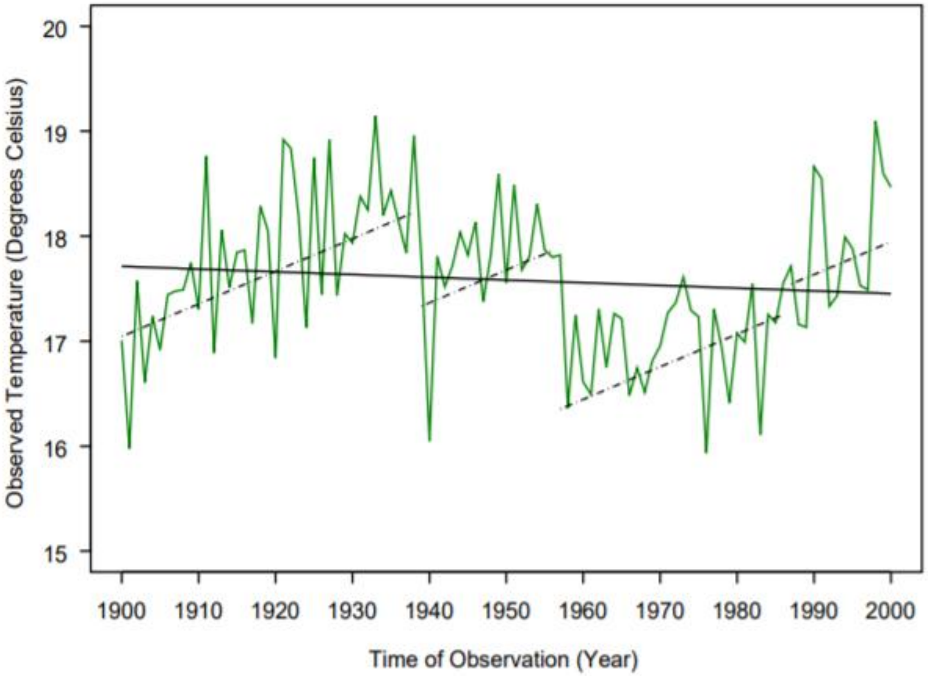
Example 3: Airlines Passengers



Example 5: Airline Passengers Again



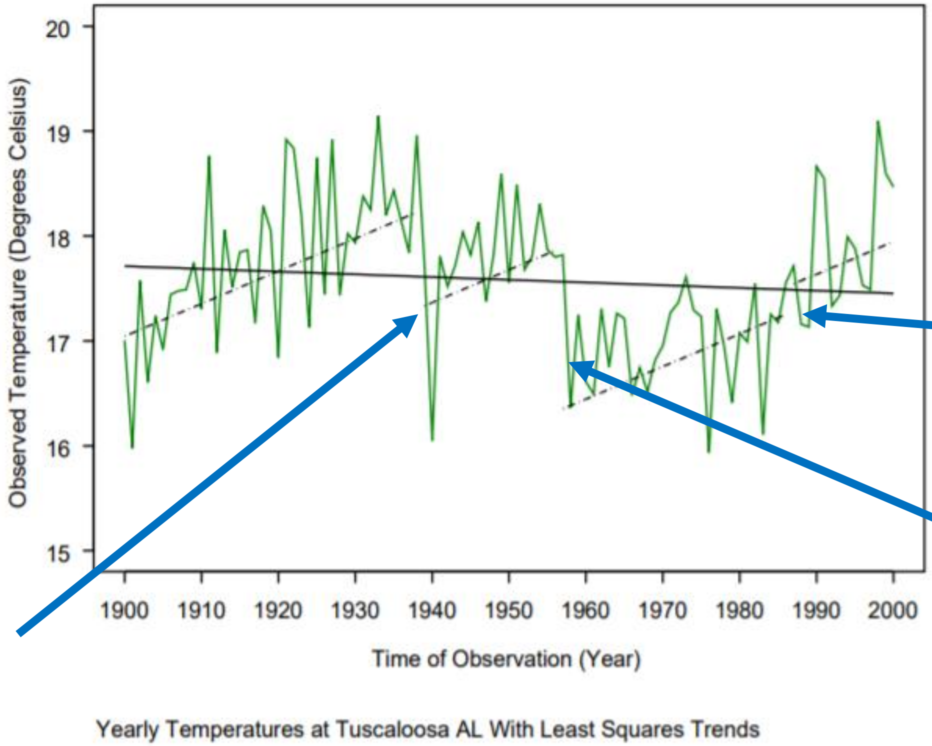
Temperature over the past century for Tuscaloosa, Alabama



Yearly Temperatures at Tuscaloosa AL With Least Squares Trends

Source: Dr. Robert Lund

Temperature over the past century for Tuscaloosa, Alabama



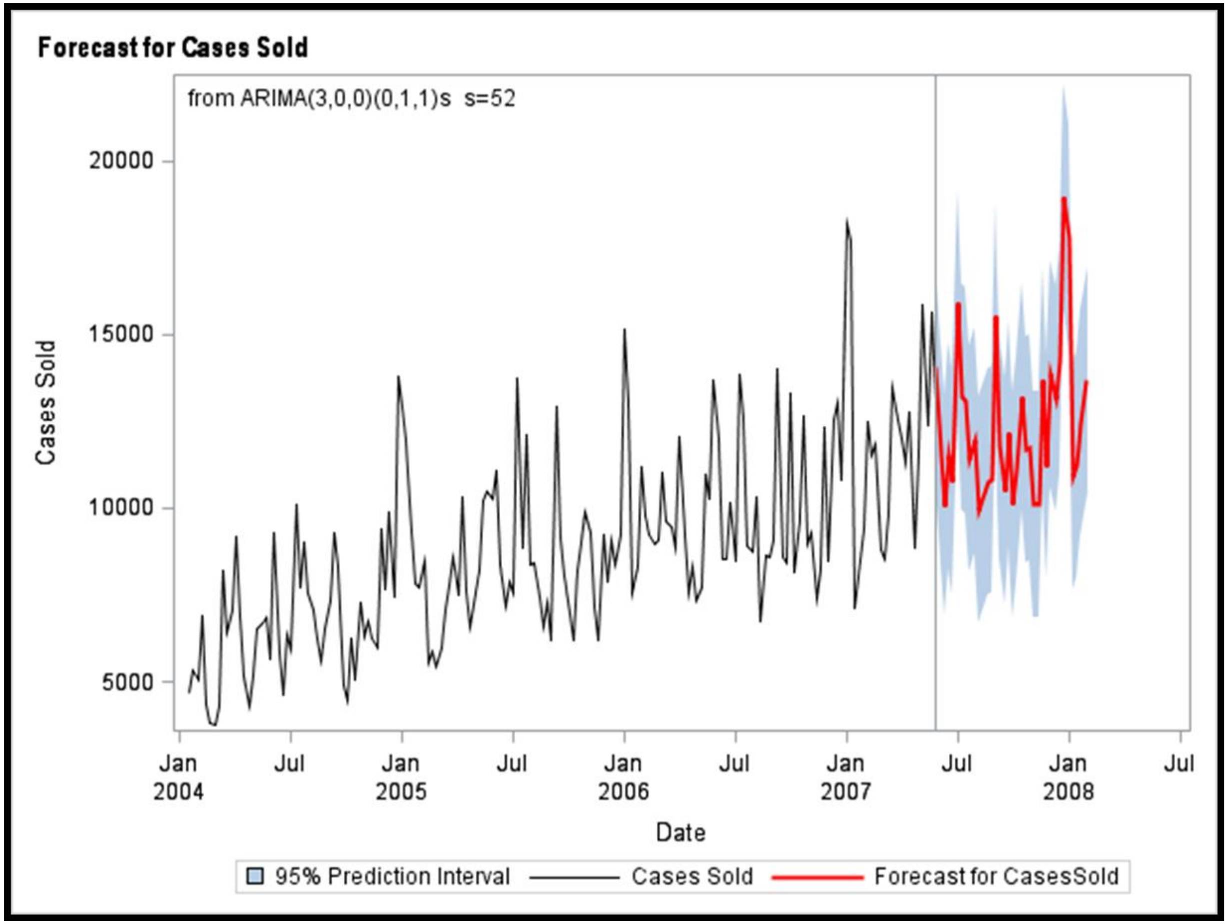
Station relocated and instrumentation was changed

Thermometer was changed

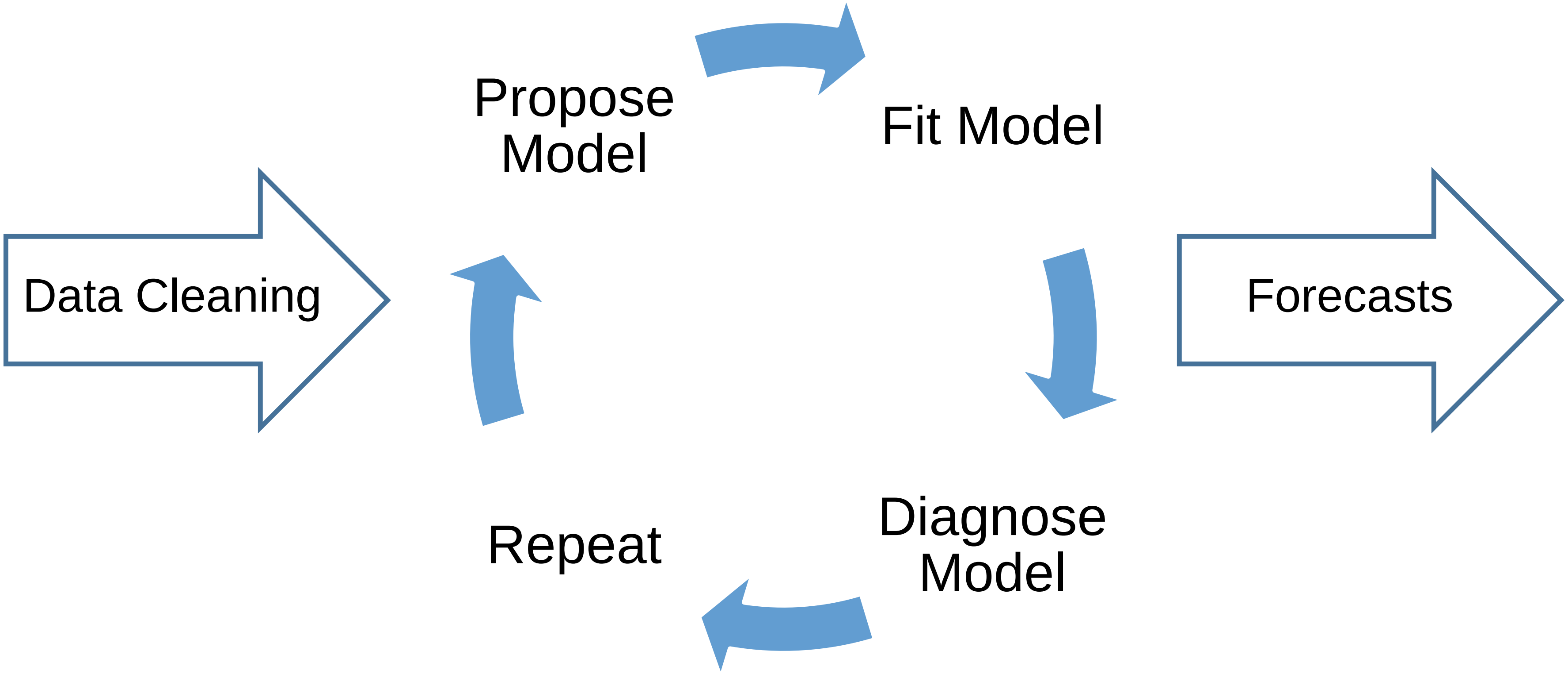
Station was relocated

Source: Dr. Robert Lund

Time Series to Forecast

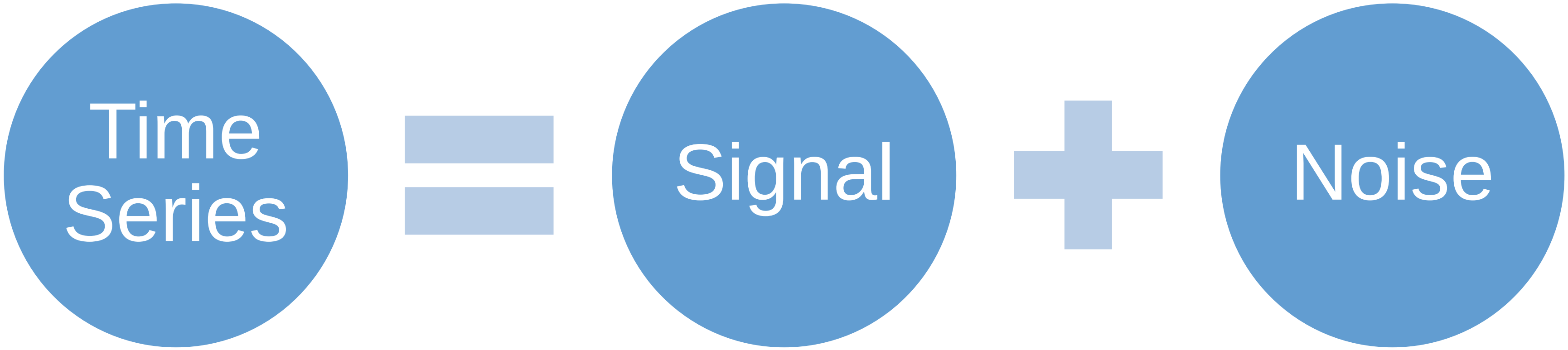


Forecasting Process

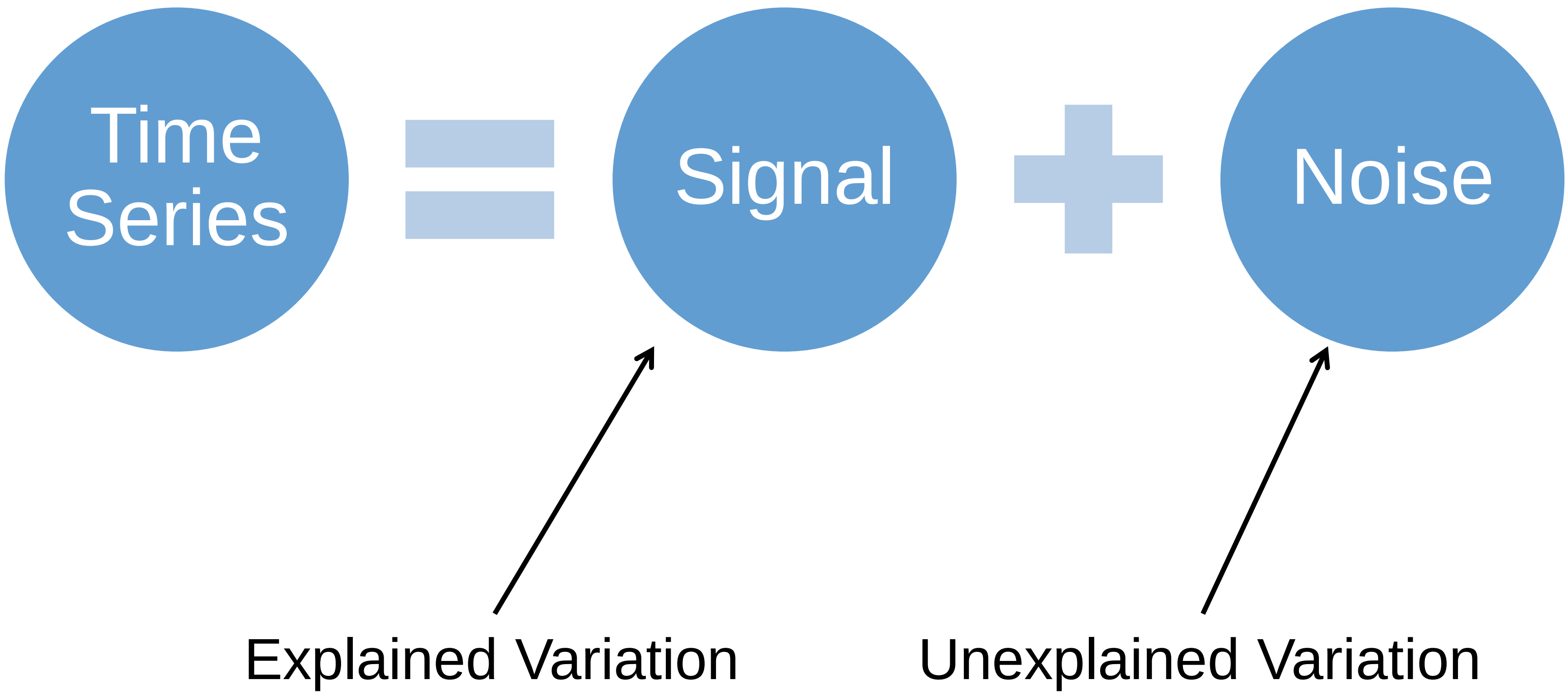


SIGNAL AND NOISE

Statistical Forecasting



Statistical Forecasting



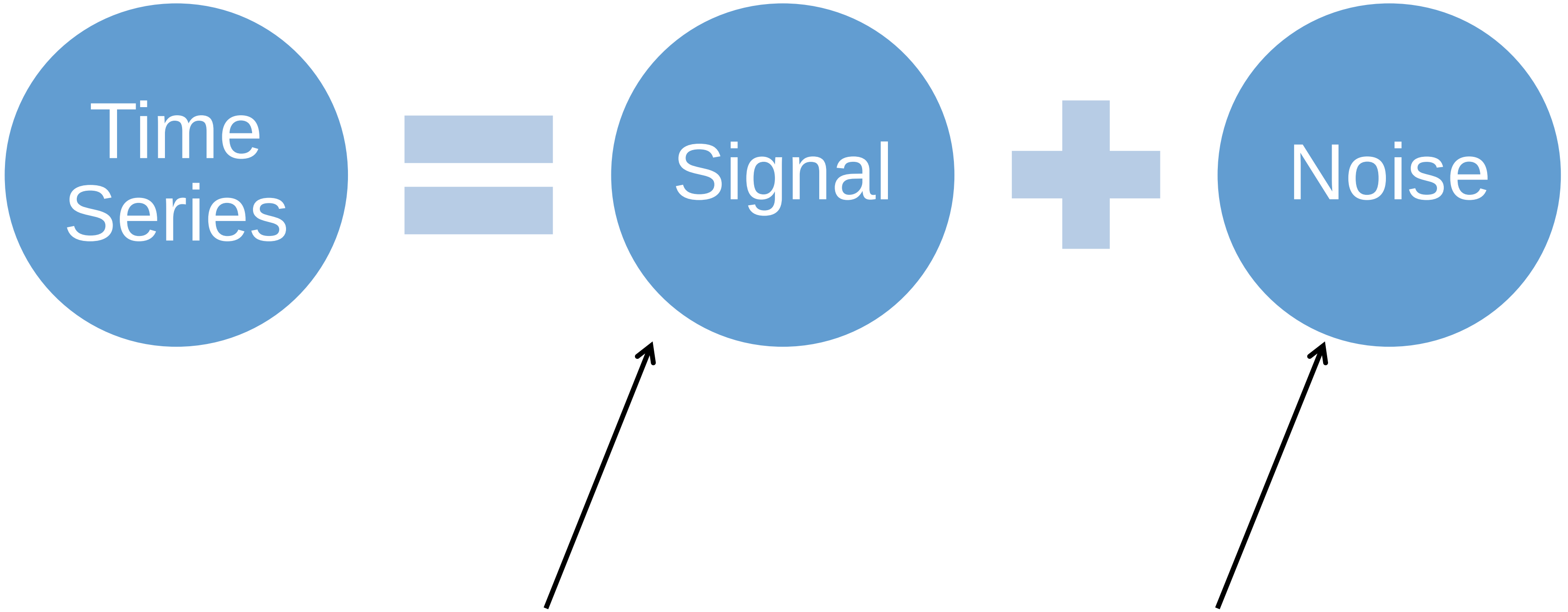
Statistical Forecasting

Time Series = Signal + Noise

Explained variation:
Trend/Cycle
Seasonality
Correlation

Error

Statistical Forecasting



Forecasts extrapolate signal portion of model.

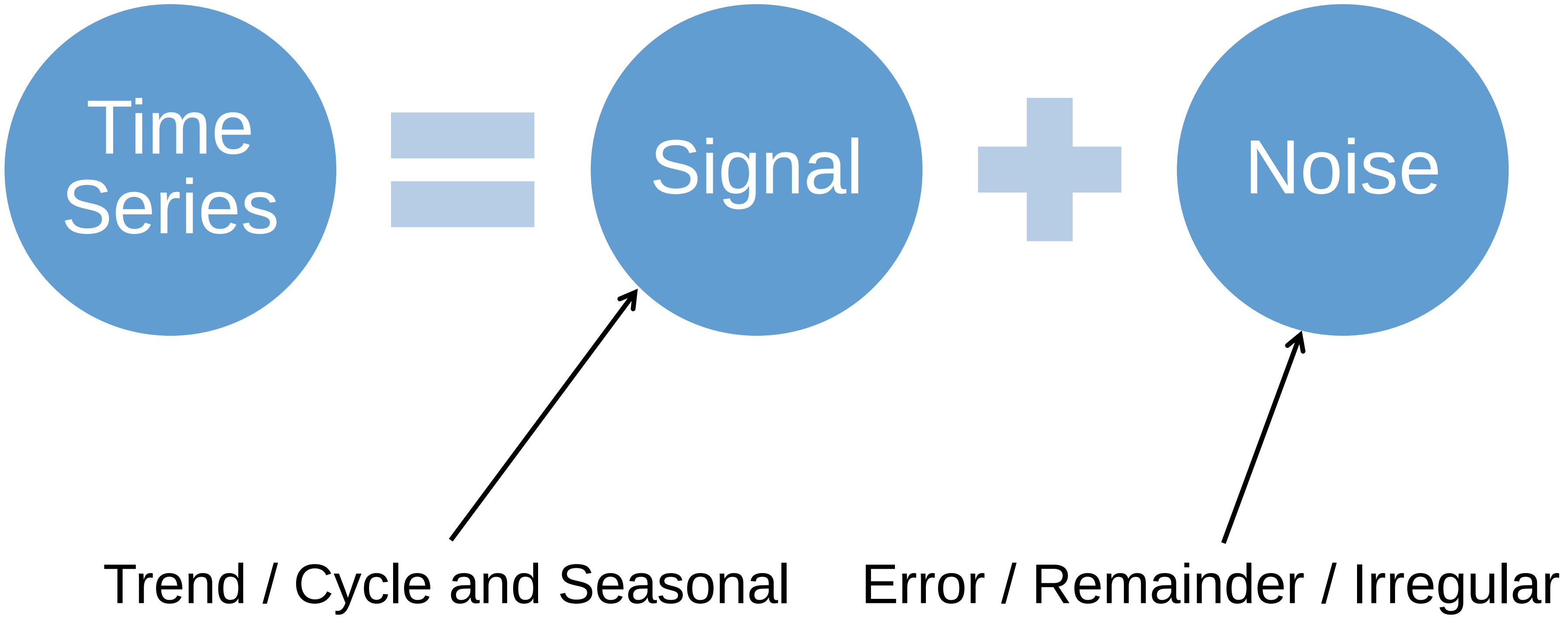
Confidence intervals account for uncertainty.

Time Series Decomposition

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

Time Series Decomposition

- The signal part of the time series can typically be broken down into two components:



Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

Trend / Cycle

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

Seasonal

  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

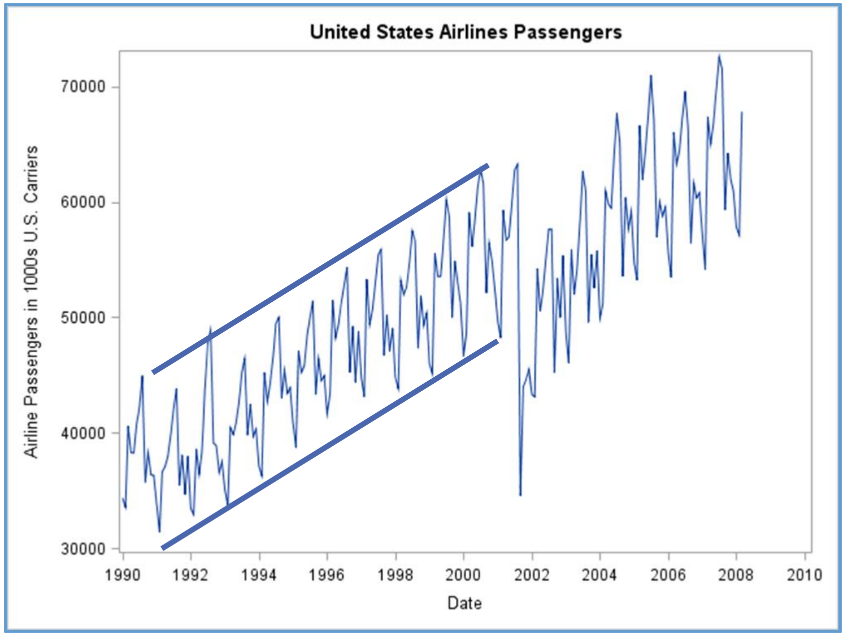
Error

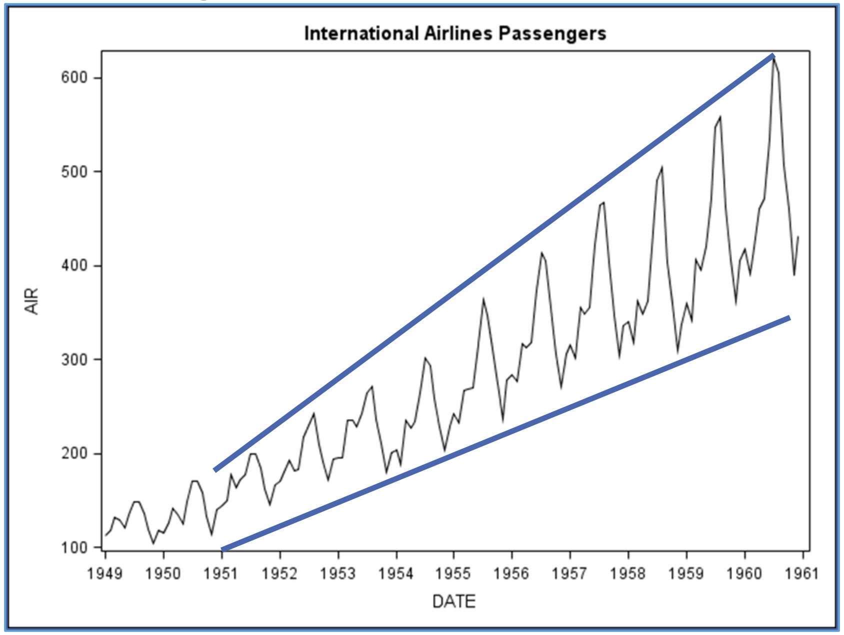
  - Multiplicative:

$$Y_t = T_t \times S_t \times E_t$$

Additive vs. Multiplicative

- Additive – magnitude of variation around trend / cycle remains constant.
- Multiplicative – magnitude of the variation around trend / cycle proportionally changes.





Time Series Decomposition

- The whole time series can now be thought of like the equations below.
  - Additive:

$$Y_t = T_t + S_t + E_t$$

  - Multiplicative:

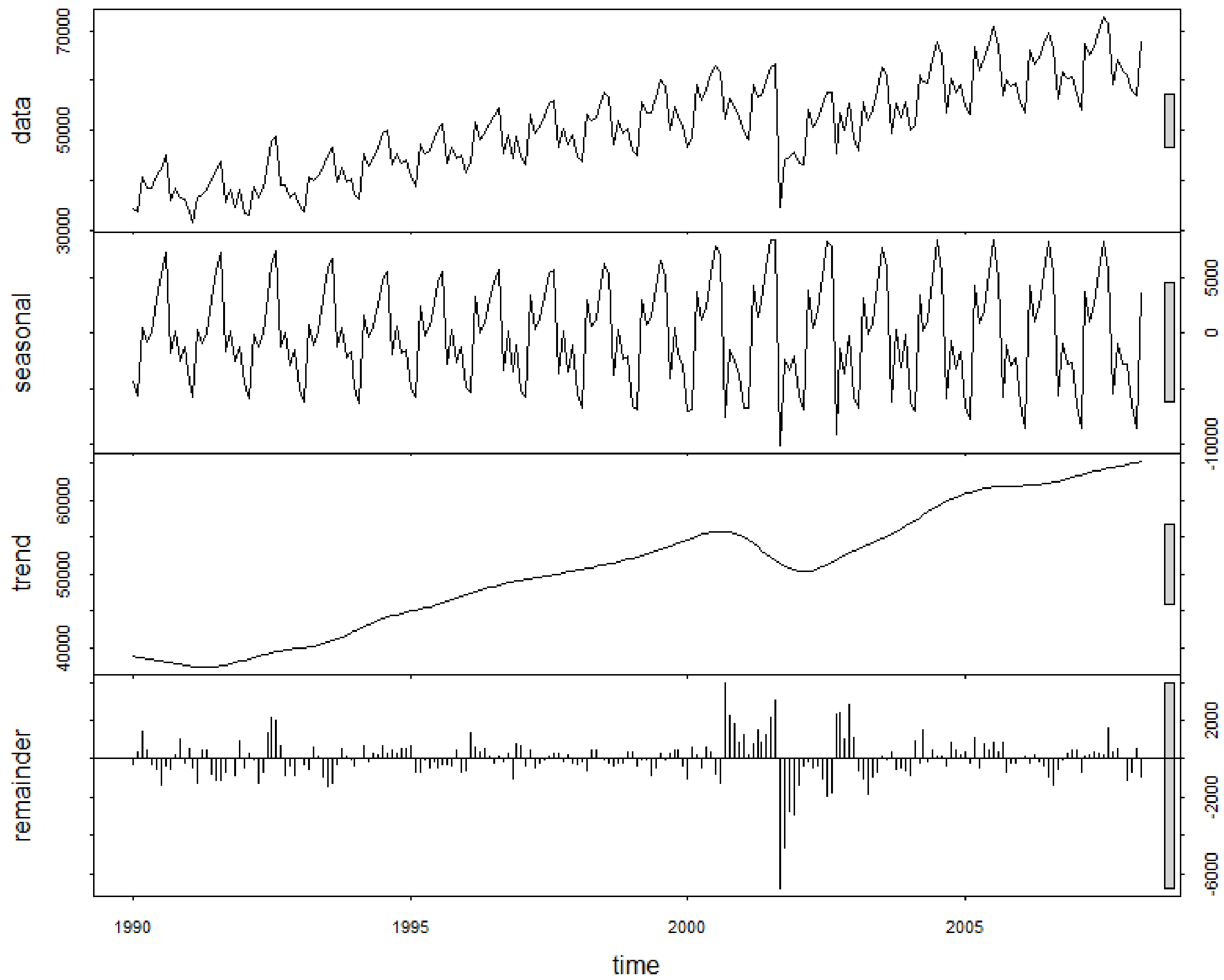
$$Y_t = T_t \times S_t \times E_t$$

OR

$$\log(Y_t) = \log(T_t) + \log(S_t) + \log(E_t)$$

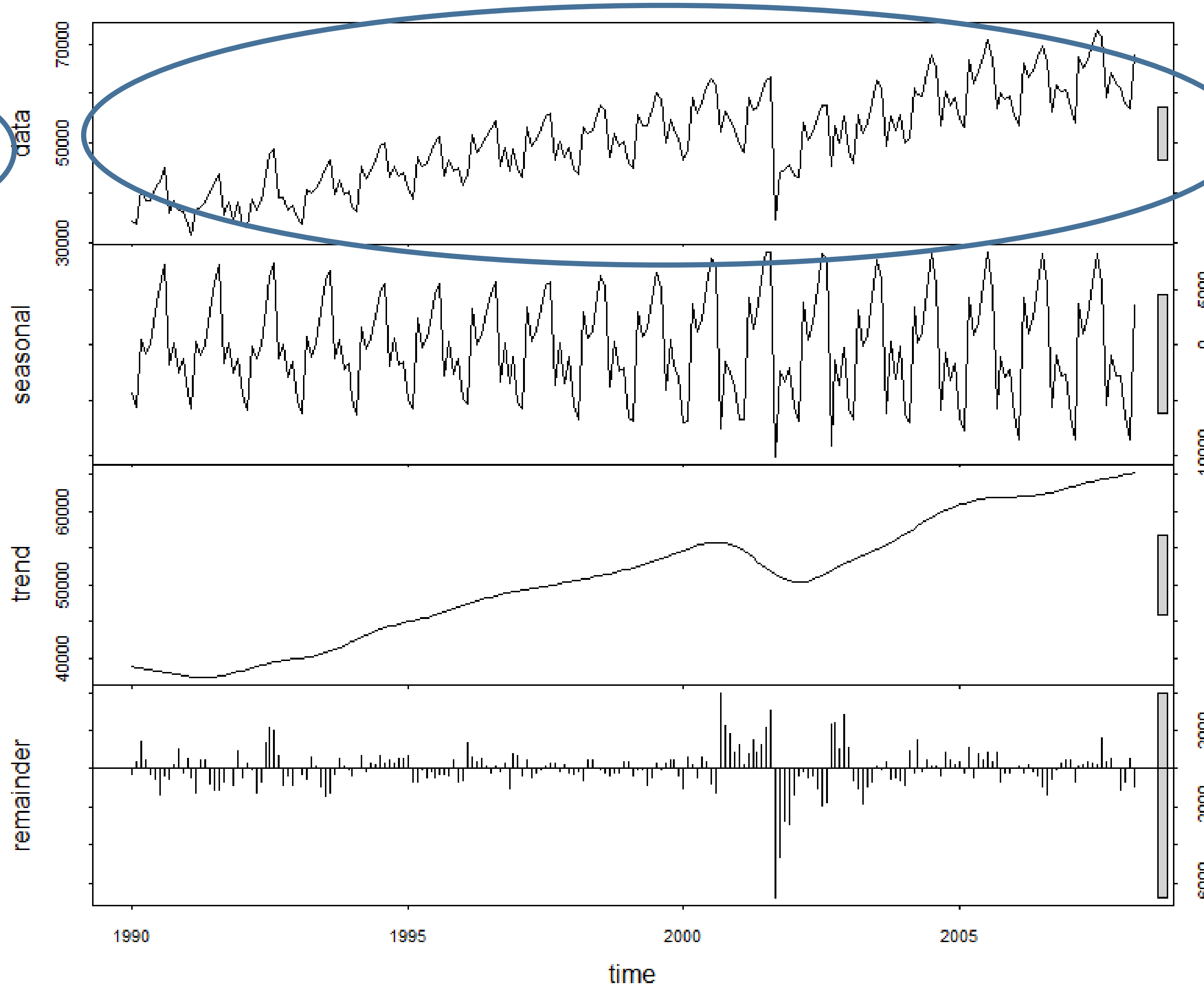
Time Series Decomposition

$$Y_t = S_t + T_t + E_t$$



Time Series Decomposition

$$Y_t = S_t + T_t + E_t$$



Original data

Time Series Decomposition

Y_t

$=$

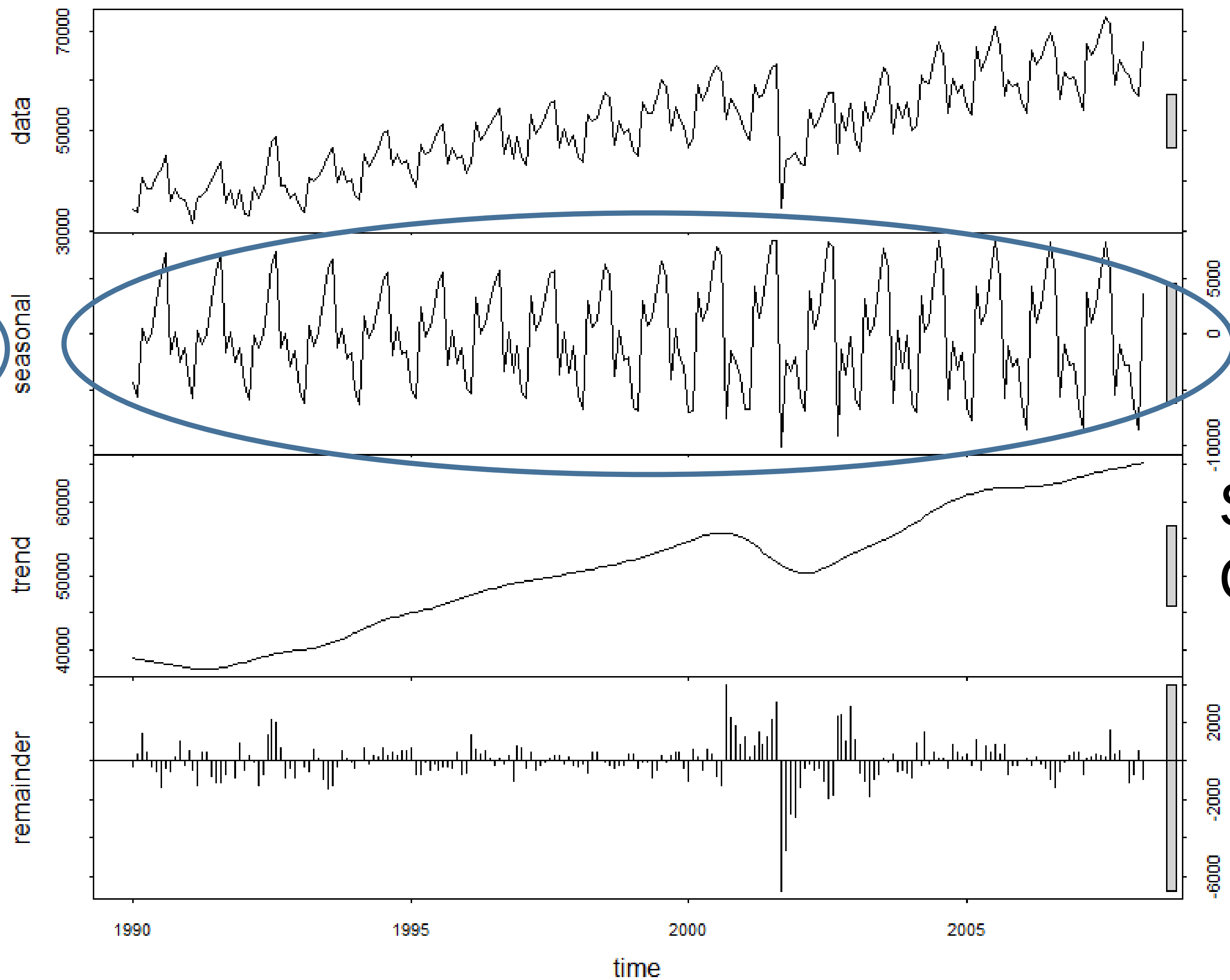
S_t

$+$

T_t

$+$

E_t



Seasonal Component

Time Series Decomposition

Y_t

$=$

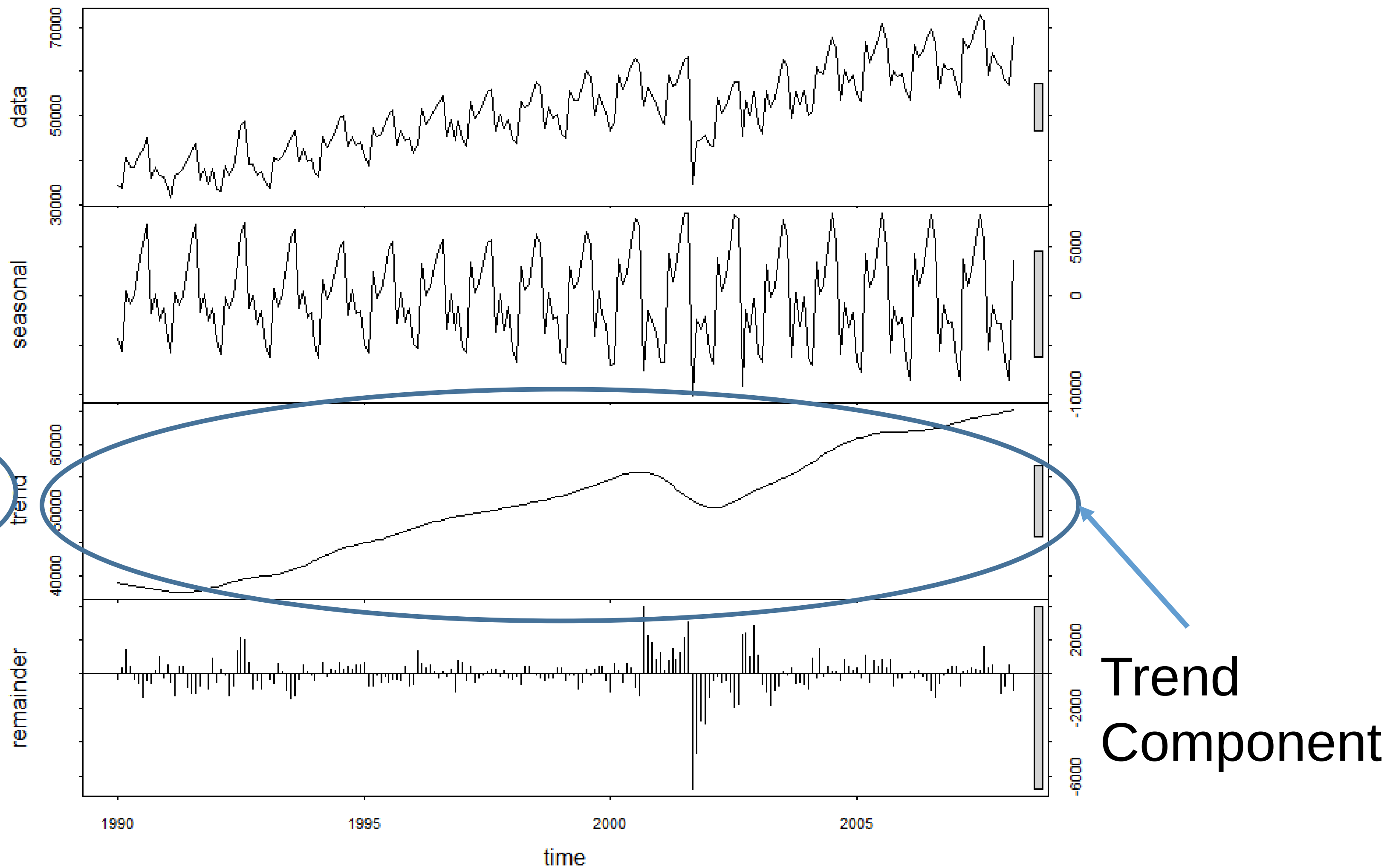
S_t

$+$

T_t

$+$

E_t



Time Series Decomposition

Y_t

$=$

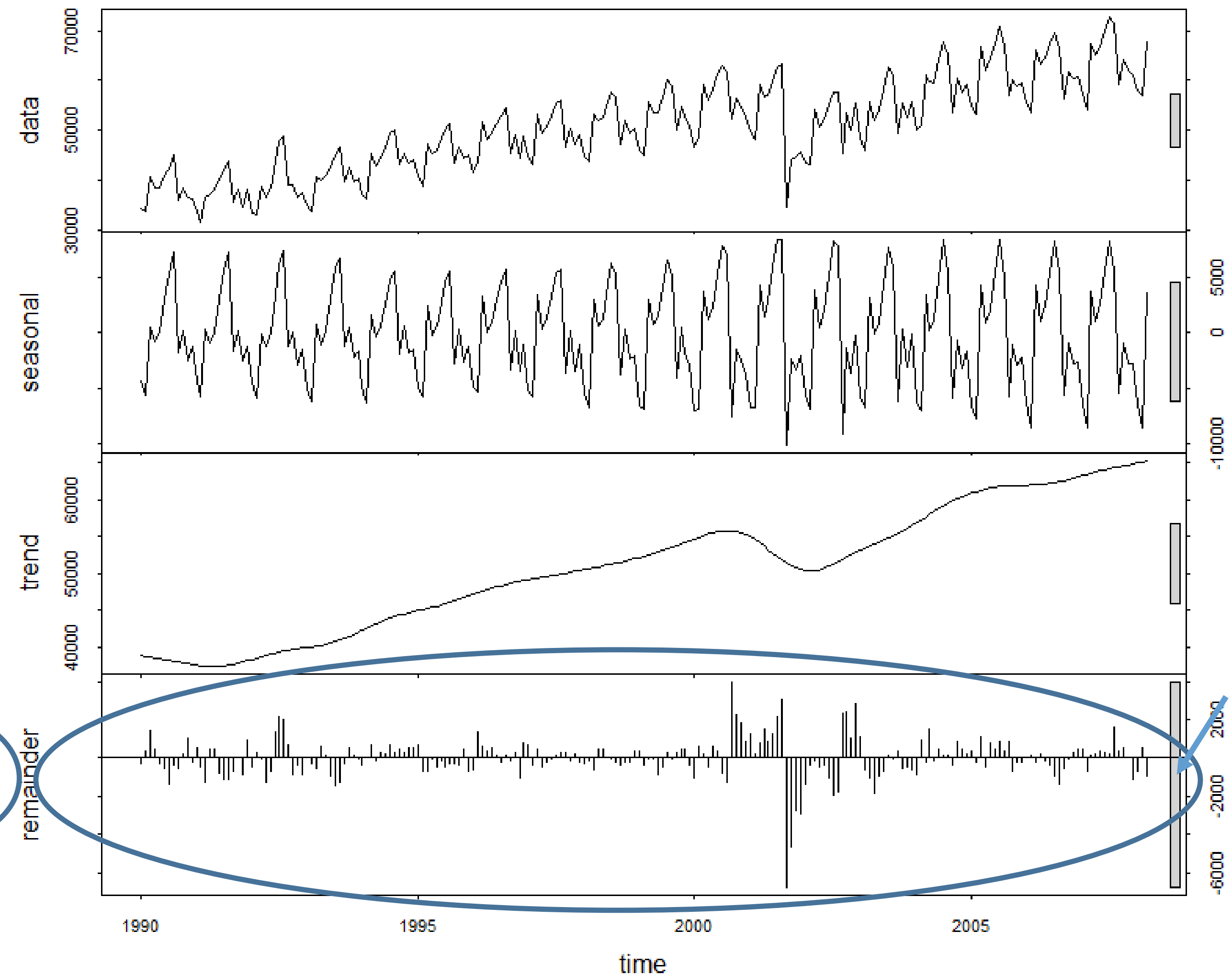
S_t

$+$

T_t

$+$

E_t



Remainder (What's left over)

Needed libraries in R

```
install.packages(‘forecast’,dependencies=T)
install.packages('tseries')
install.packages(c('expsmooth','lmtest','zoo'))

library(haven)
library(fma)
library(tseries)
library(expsmooth)
library(lmtest)
library(zoo)
```

Importing SAS data sets

```
file.dir <- “filename/"
input.file1 <- "usair.sas7bdat"
input.file2 <- "ar2.sas7bdat"

USAirlines <- read_sas(paste(file.dir, input.file1,sep = ""))
AR2 <- read_sas(paste(file.dir, input.file2, sep = ""))
```

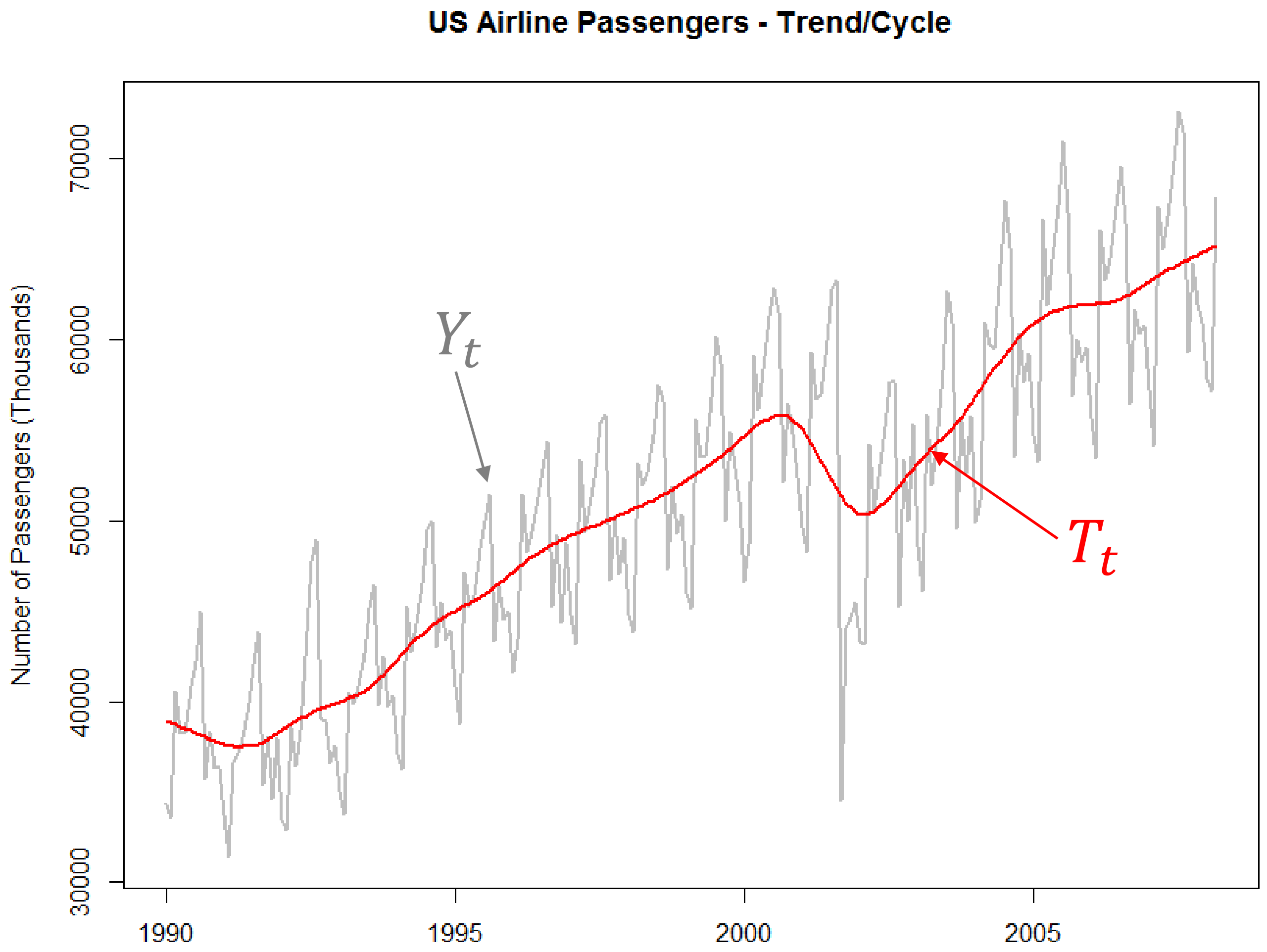
Time Series Decomposition – R

```
Passenger <- ts(USAirlines$Passengers, start = 1990,
      frequency = 12)
model <- stl(Passenger, s.window=7)
plot(model)
```

Time Series Decomposition – R

```
plot(Passenger, col="grey",
     main="US Airline Passengers - Trend/Cycle",
     xlab="", ylab="Number of Passengers (Thousands)",
     lwd=2)
lines(model$time.series[,2],col="red", lwd=2)
```

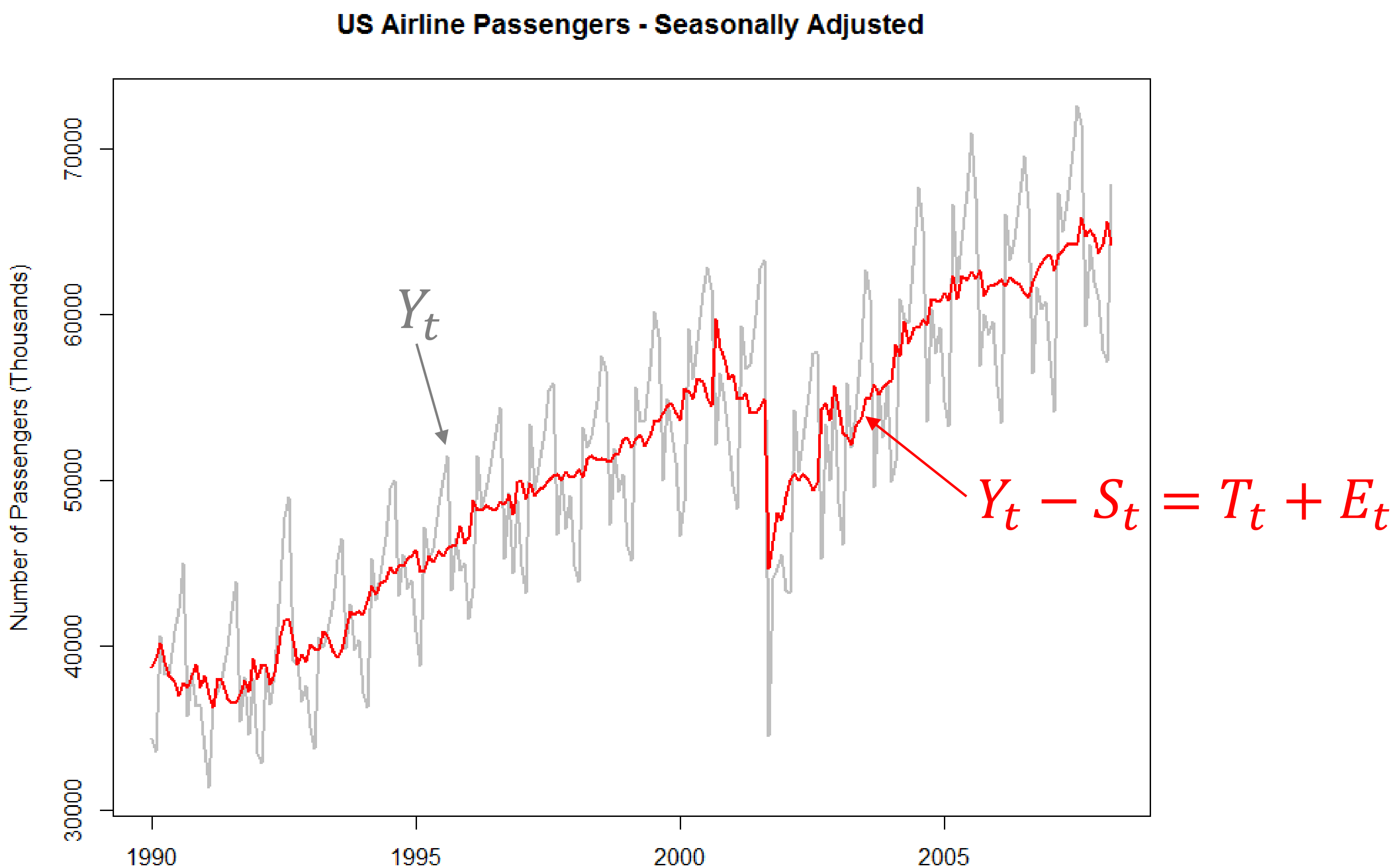
Time Series Decomposition



Time Series Decomposition – R

```
seas=Passenger-model$time.series[,1]
plot(Passenger, col = "grey", main =
   "US Airline Passengers - Seasonally Adjusted",
    xlab = "", ylab = "Number of Passengers
   (Thousands)", lwd = 2)
lines(seas, col = "red", lwd = 2)
```

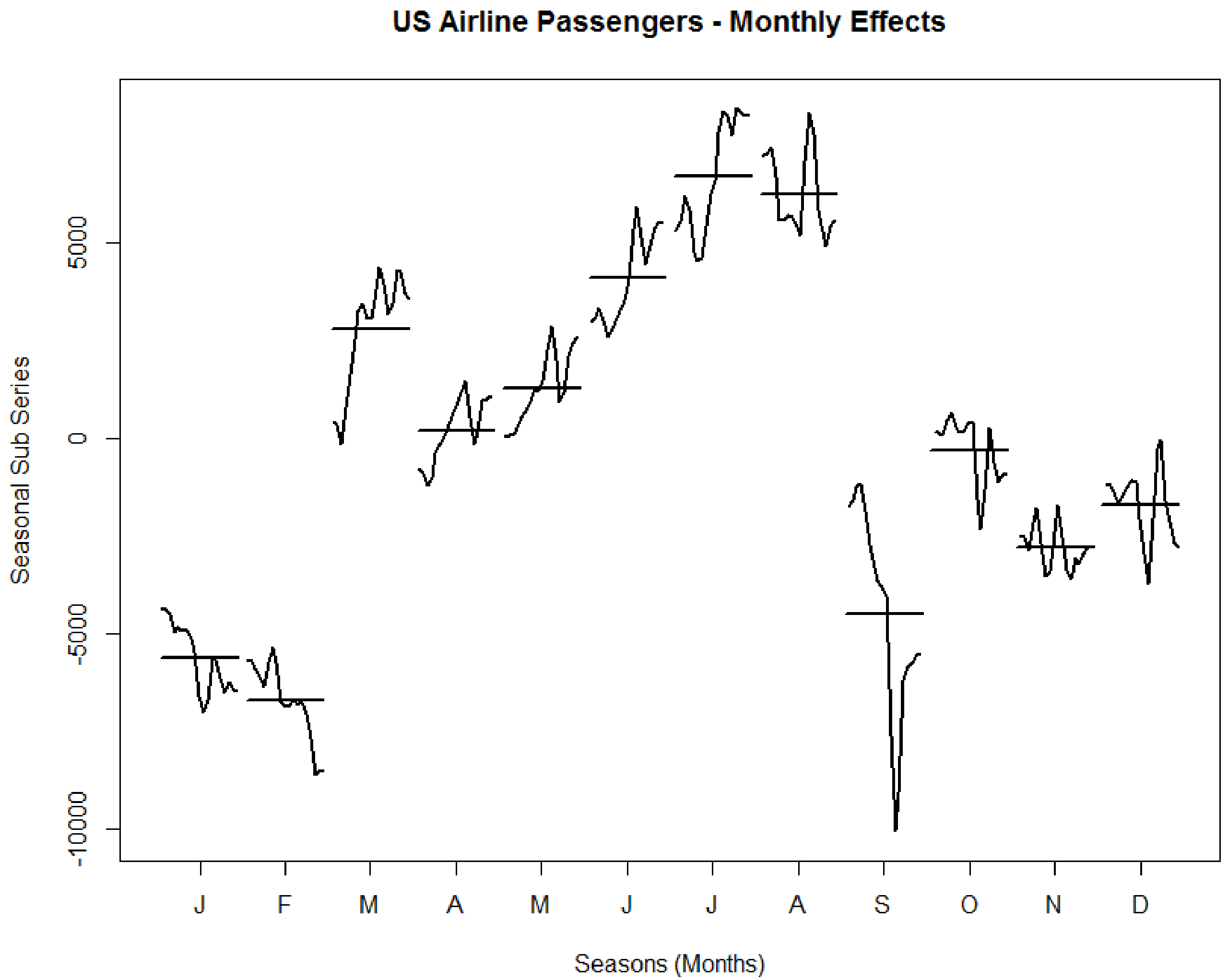
Time Series Decomposition



Time Series Decomposition – R

```
monthplot(model$time.series[,"seasonal"],
          main="US Airline Passengers -
                Monthly Effects",
          ylab="Seasonal Sub Series",
          xlab="Seasons (Months)", lwd=2)
```

Time Series Decomposition

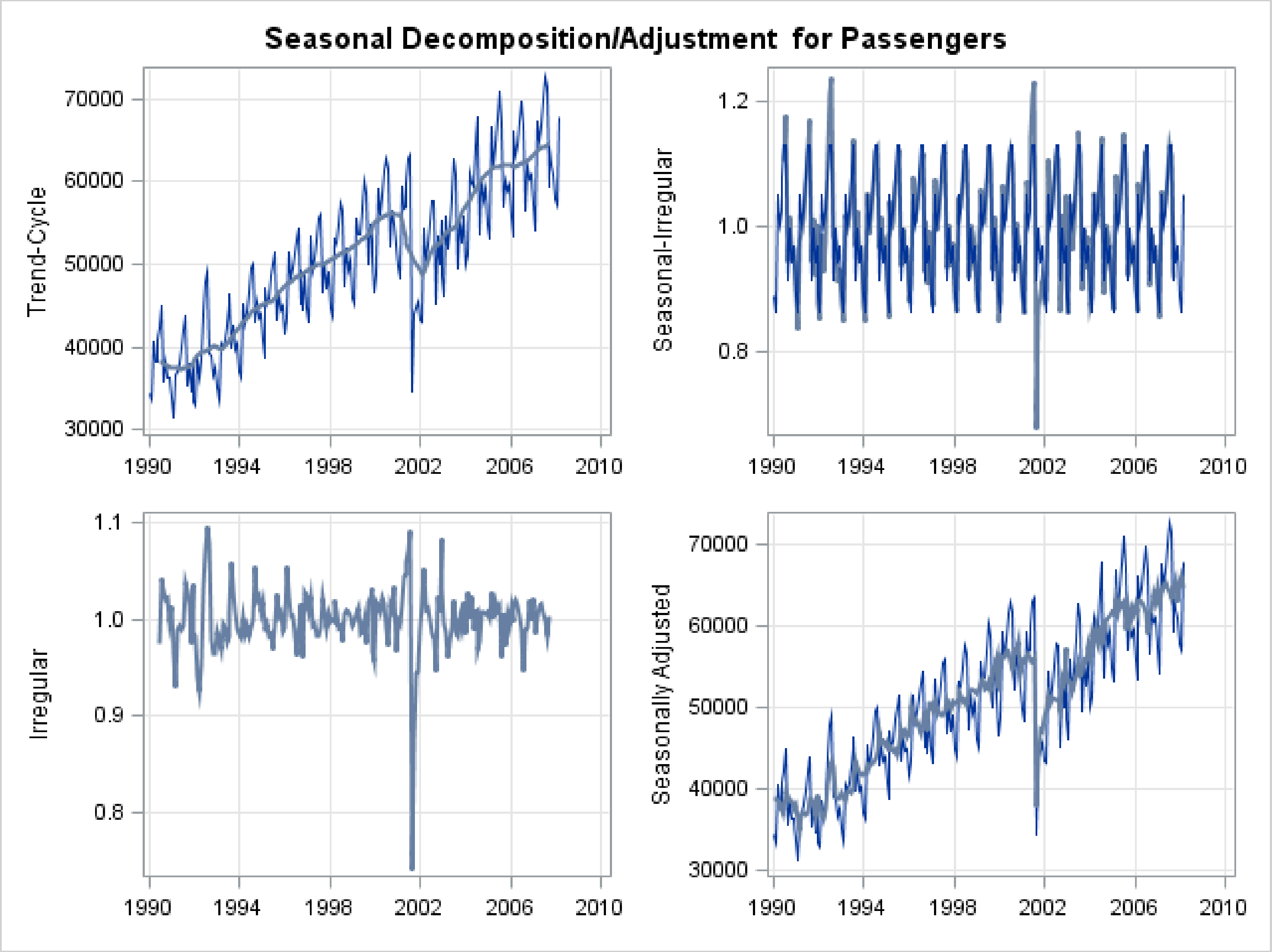


Time Series Decomposition – SAS

```
proc timeseries data=Time.USAirlines plots=(series decomp sc);
          id date interval=month;
          var Passengers;
run;
```

```
proc timeseries data=Time.USAirlines plots=(series decomp sc)
seasonality=12;
          var Passengers;
run;
```

Time Series Decomposition (decomp)



Decomposition Techniques

- There are many different ways to calculate the trend/cycle and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition

Decomposition Techniques

- There are many different ways to calculate the trend/cycle and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition
     a. Default in SAS (Can be done in R)
     b. Trend – Uses Moving / Rolling Average Smoothing
     c. Seasonal – Average De-trended Values Across Seasons

Decomposition Techniques

- There are many different ways to calculate the trend/cycle and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition
  2. X-12 ARIMA Decomposition (now at X-13…self study)
     a. Trend – Uses Moving / Rolling Average Smoothing
     b. Seasonal – Uses Moving / Rolling Average Smoothing
     c. Iteratively Repeats Above Methods and ARIMA Modeling

Decomposition Techniques

- There are many different ways to calculate the trend/cycle, and seasonal effects inside time series data.
- Here are 3 common techniques:
  1. Classical Decomposition
  2. X-12 ARIMA Decomposition
  3. STL (Seasonal and Trend using LOESS estimation) Decomposition
     a. Default of stl Function in R (Not available in SAS)
     b. Uses **LO**cal regr**ESS**ion Techniques to Estimate Trend and Seasonality
     c. Allows Changing Effects for Trend and Season
     d. Adapted to Handle Outliers

Comparison of seasonal component in SAS versus R