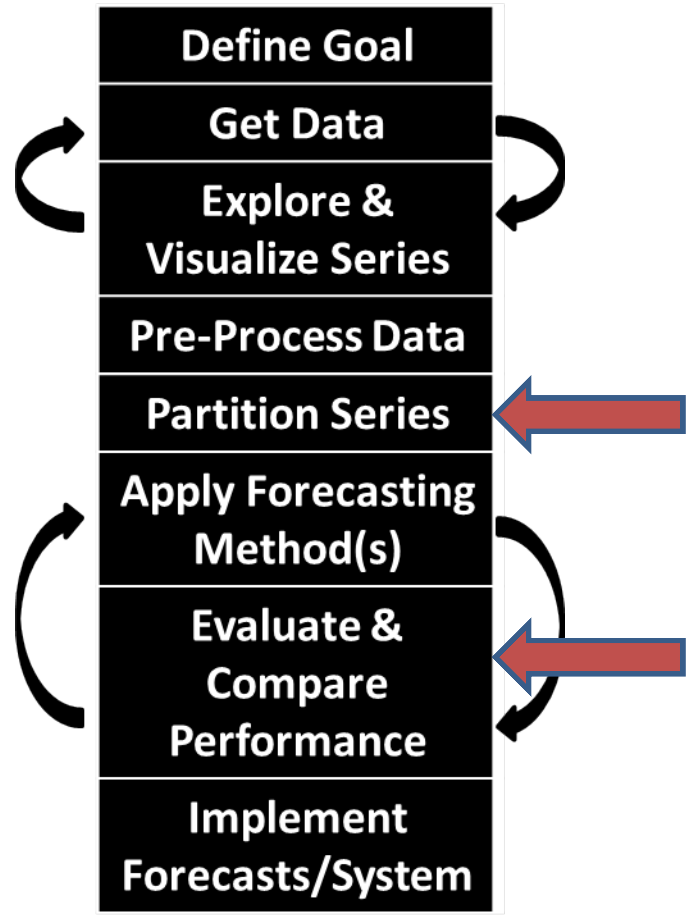
PERFORMANCE EVALUATION

Forecasting Analytics

Recall the Forecasting Process

Today: Partitioning, Naïve Forecasts, and Evaluation



How well will the forecasting model work in practice?

Forecast accuracy
Practical implications

Is the best fitting model the best forecaster of future values?

How well will the forecasting model work in practice?

Forecast accuracy
Practical implications

Is the best fitting model the best forecaster of future values?

A model which fits the data well does not necessarily forecast well.

A perfect fit can always be obtained by using a model with enough parameters.

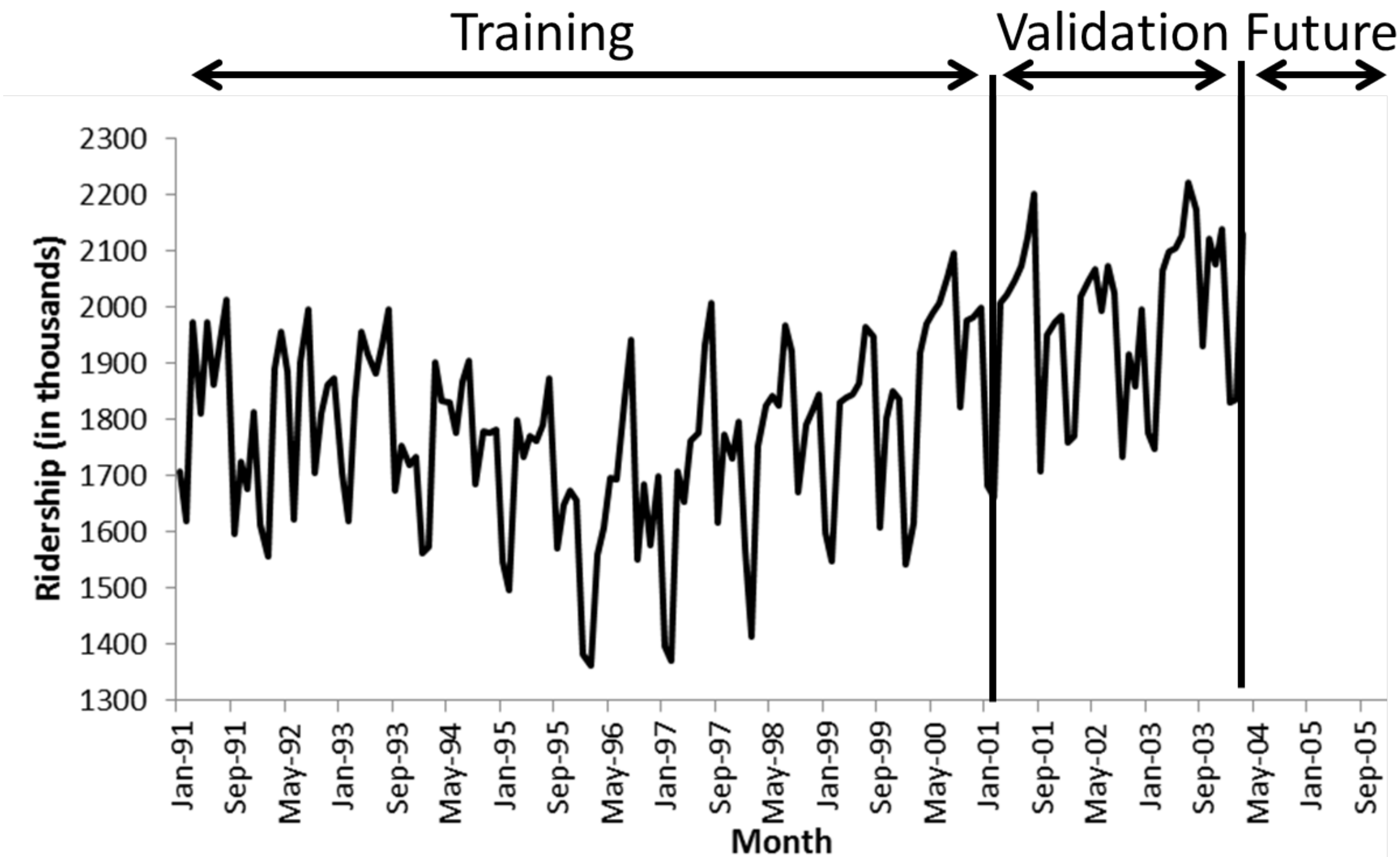
With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.

John von Neumann (physicist on Manhattan Project and principal contributor to Monte Carlo methods, game theory, etc.)

Over-fitting a model to data is as bad as failing to identify the systematic pattern in the data.

Forecasting: principles and practice
An online textbook by Hyndman & Athanasopoulos

Data partitioning



How/when to use the partitions?

Fit the model only to **training** period.

Assess performance on **validation** period.

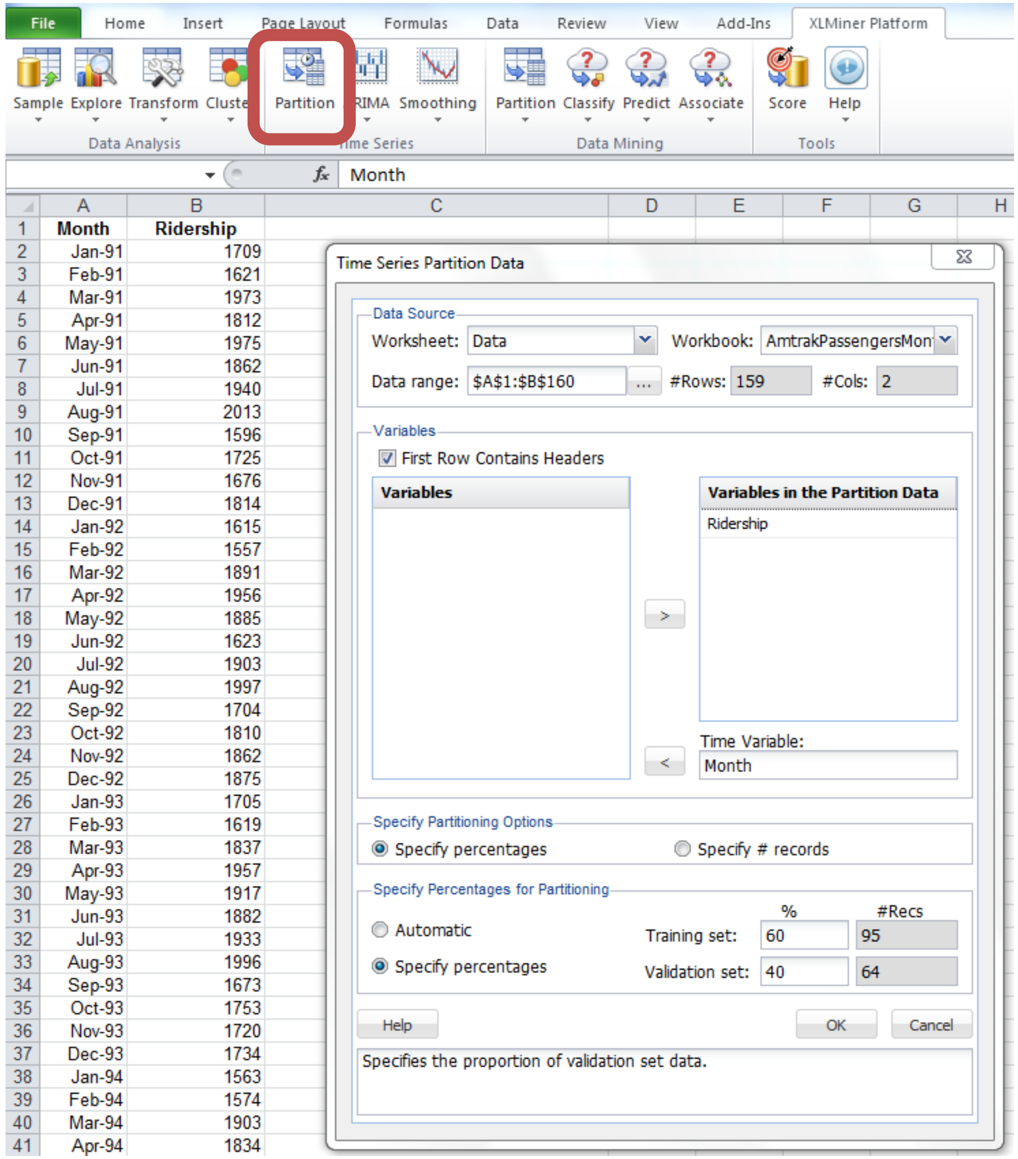
Deploy model by joining **training+validation**; rerun the chosen model from validation on all the data to forecast the future.

How to choose a validation period?

Depends on:

- Forecast horizon
- Seasonality
- Length of series
- Underlying conditions affecting series

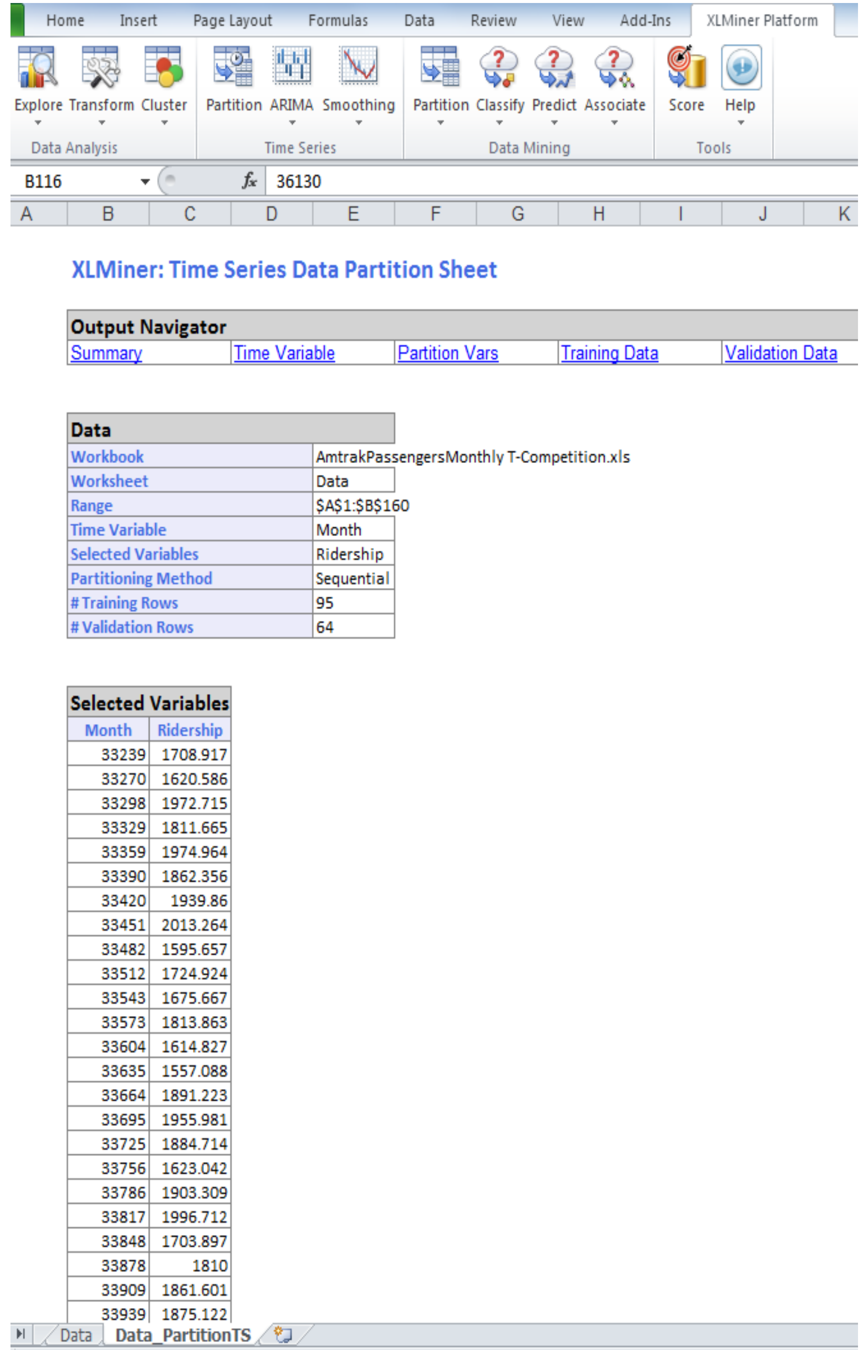
Partitioning time series in XLMiner



New partitioned worksheet: Data_ParttionTS1

Running any forecasting method on the partitioned sheet will **fit the model to the training period** only

We get performance on the training and validation periods **separately**



XLMiner: Time Series Data Partition Sheet

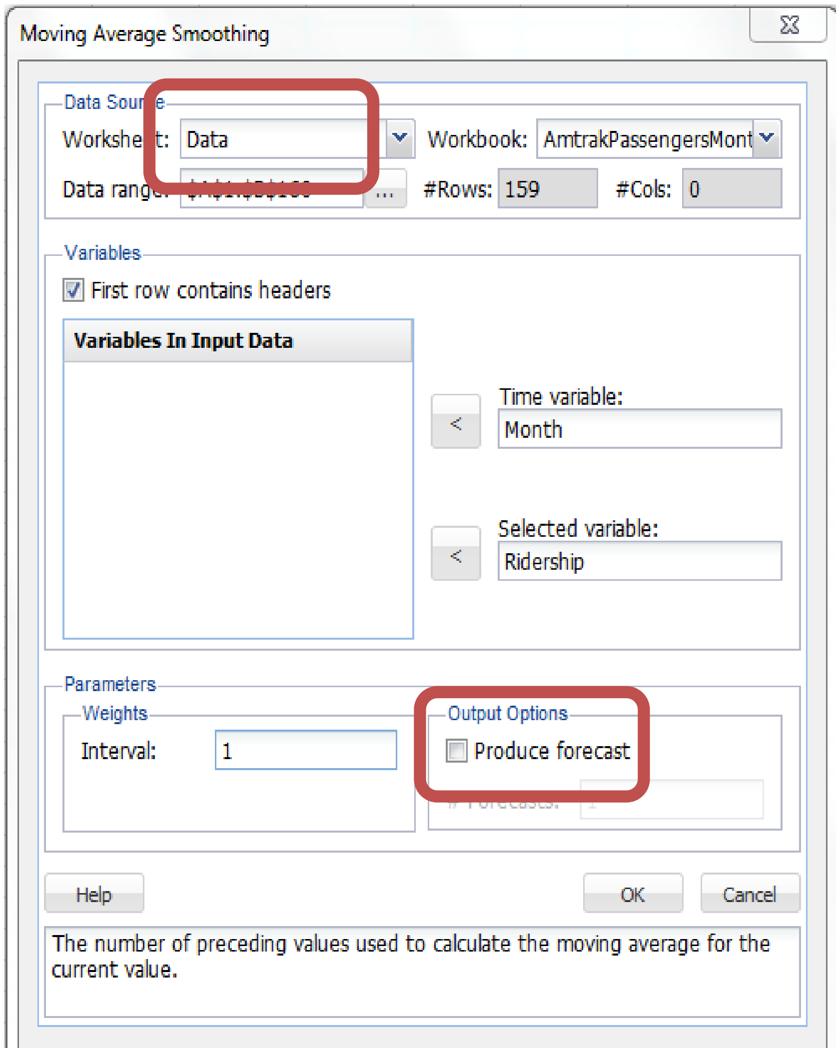
Output Navigator				
Summary	Time Variable	Partition Vars	Training Data	Validation Data

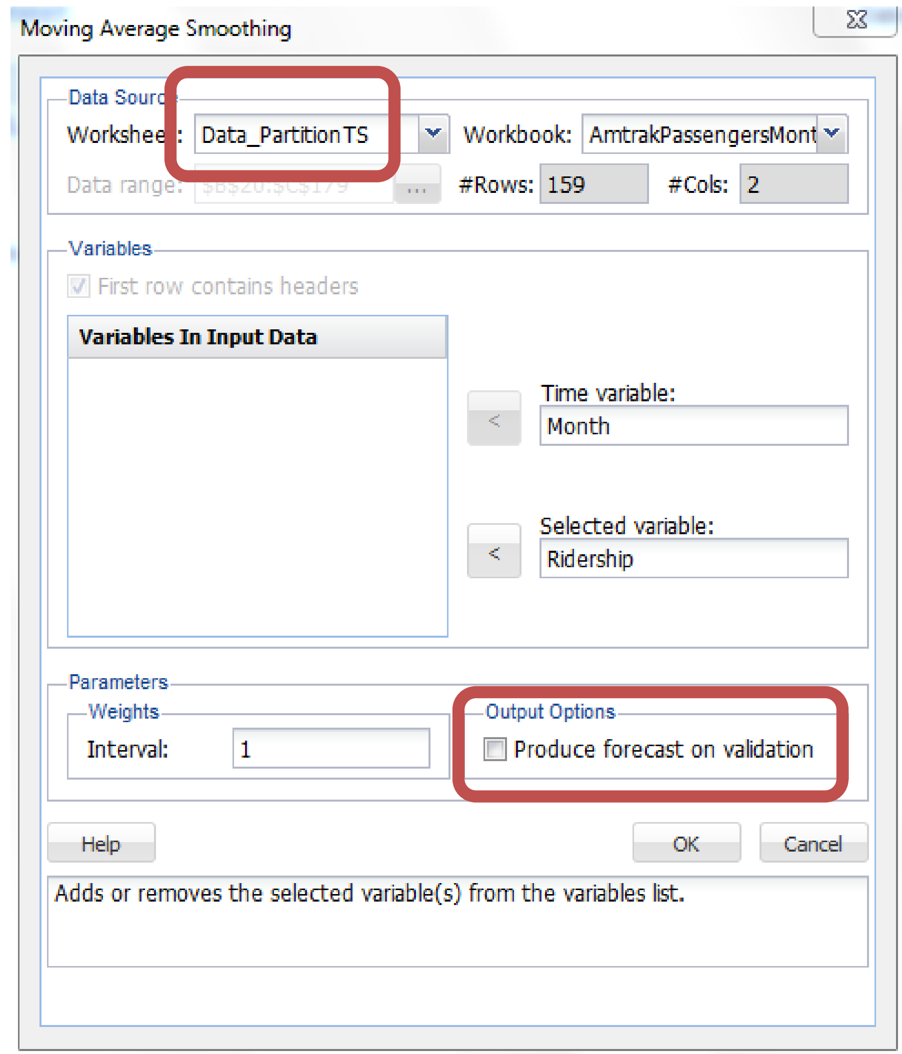
Data	
Workbook	AmtrakPassengersMonthly T-Competition.xls
Worksheet	Data
Range	$A$1:$B$160
Time Variable	Month
Selected Variables	Ridership
Partitioning Method	Sequential
# Training Rows	95
# Validation Rows	64

Selected Variables

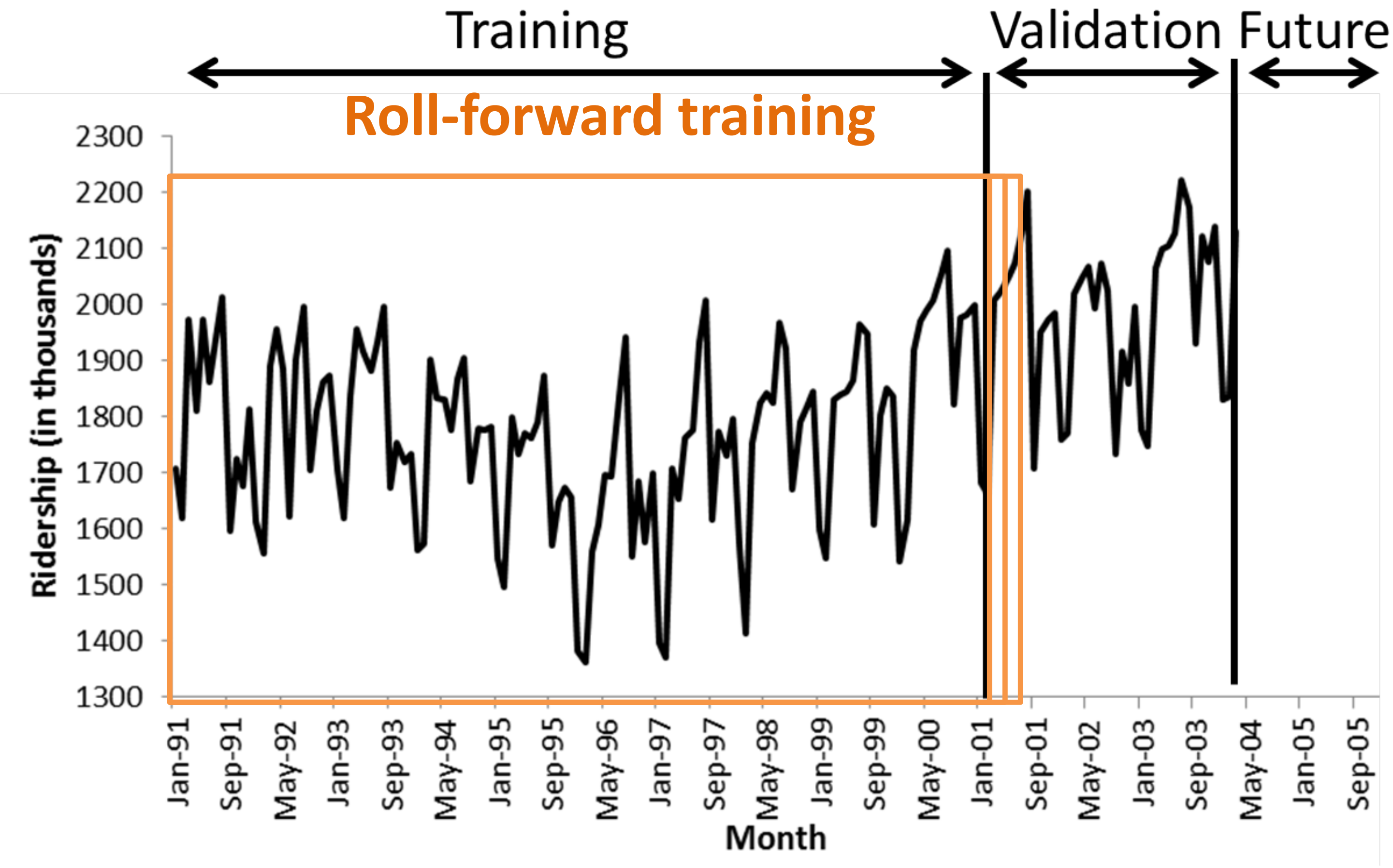
Month	Ridership
33239	1708.917
33270	1620.586
33298	1972.715
33329	1811.665
33359	1974.964
33390	1862.356
33420	1939.86
33451	2013.264
33482	1595.657
33512	1724.924
33543	1675.667
33573	1813.863
33604	1614.827
33635	1557.088
33664	1891.223
33695	1955.981
33725	1884.714
33756	1623.042
33786	1903.309
33817	1996.712
33848	1703.897
33878	1810
33909	1861.601
33939	1875.122

For un-partitioned data, we get forecasts of future





Roll-forward forecasts: Ideally, the training period would roll forward



NAÏVE FORECASTS

OCCAM'S RAZOR

Sure there are simpler ways to catch that bird,
but the complicated ones kick ass.

Naïve k-step ahead forecast:

$$F_{t+k} = y_t$$

For a seasonal series (with M seasons)

Naïve k-step ahead forecast:

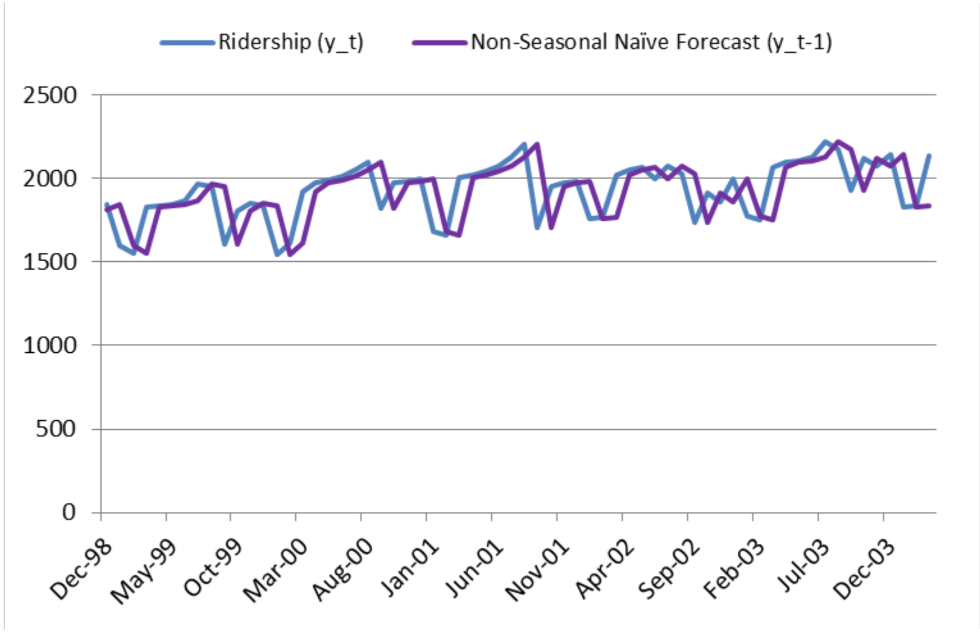
$$F_{t+k} = y_{t+k-M}$$

Evaluating predictive accuracy

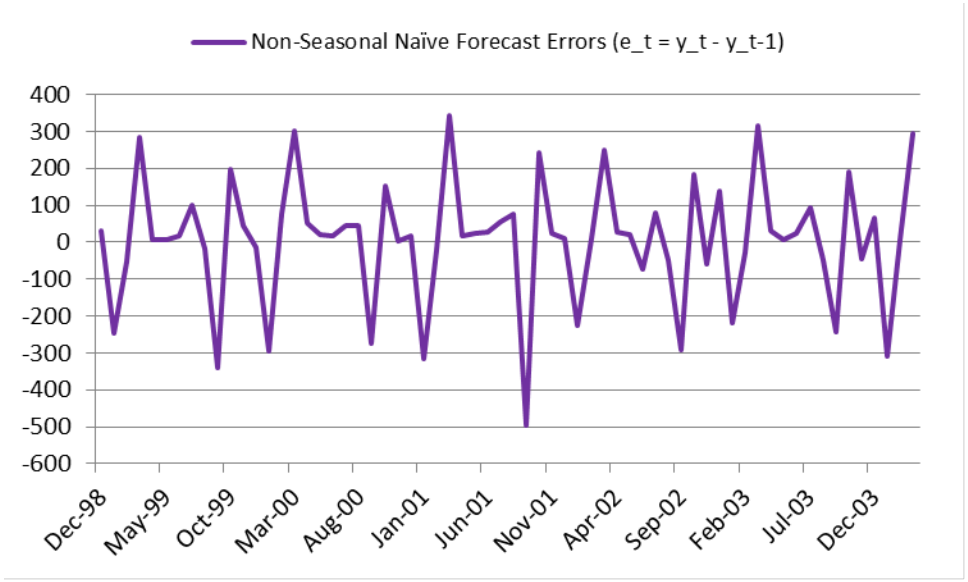
Look at **validation period**

Compare actual (y_t) and forecasted (F_t) values

Examine forecast errors ($e_t = y_t - F_t$)



Ridership (y_t)
Non-Seasonal Naïve Forecast (y_t-1)
2500
2000
1500
1000
500
0
Dec-98
May-99
Oct-99
Mar-00
Aug-00
Jan-01
Jun-01
Nov-01
Apr-02
Sep-02
Feb-03
Jul-03
Dec-03



Non-Seasonal Naïve Forecast Errors (e_t = y_t - y_t-1)
400
300
200
100
0
-100
-200
-300
-400
-500
-600
Dec-98
May-99
Oct-99
Mar-00
Aug-00
Jan-01
Jun-01
Nov-01
Apr-02
Sep-02
Feb-03
Jul-03
Dec-03

Common Predictive Accuracy Measures

Compute some function of the forecast error:

Then, average across all records:

e_t	***Average Error***
$\|e_t\|$	***Mean absolute error*** **(MAE)**
$(e_t)^2$	***Mean Squared Error*** **(MSE)** **or take a root (RMSE)**
$\|e_t / y_t\| \times 100\%$	***Mean absolute % error*** **(MAPE)**

How are the measures different?

	Avg Error	MAE	MSE	MAPE
Easily interpreted	√	√		√
Sensitive to direction	√			
Sensitive to large errors			√	√
Zero counts acceptable	√	√	√	
Scale-independent				√
Symmetric in over/under forecast	√	√	√	heavier penalty on over-forecasts for count data

See more here: http://otexts.com/fpp/2/5/

Amtrak Ridership: consider SOME forecasting model

Single Validation Period

Month	Forecast	Actual Value	Forecast Error
Apr 2003	2114.958	2098.899	-16.059
May 2003	2153.025	2104.911	-48.114
June 2003	2118.499	2129.671	11.172
July 2003	2229.781	2223.349	-6.432
Aug 2003	2281.653	2174.360	-107.293
Sept 2003	1955.451	1931.406	-24.045
Oct 2003	2101.786	2121.470	19.684
Nov 2003	2098.774	2076.054	-22.720
Dec 2003	2149.743	2140.677	-9.066
Jan 2004	1920.407	1831.508	-88.899
Feb 2004	1890.080	1838.006	-52.074
Mar 2004	2197.968	2132.446	-65.522

$$MAE = \frac{1}{12}(|-16.059| + \cdots + |-65.522|) = 39.26$$

$$Average\,error = \frac{1}{12}(-16.059 + \cdots + (-65.522)) = -34.14$$

$$MAPE = \frac{1}{12}\left(\left|\frac{-16.059}{2098.899}\right| + \cdots + \left|\frac{-65.522}{2132.446}\right|\right) \times 100\% = 1.9\%$$

$$RMSE = \sqrt{\frac{1}{12}((-16.059)^2 + \cdots + (-65.522)^2)} = \sqrt{2560.2} = 50.6$$

Model Performance

Training Error Measures

Mean Absolute Percentage Error (MAPE)	7.354745417
Mean Absolute Deviation (MAD)	125.7275474
Mean Square Error (MSE)	27642.2836
Tracking Signal Error (TSE)	0.85738569
Cumulative Forecast Error (CFE)	107.797
Mean Forecast Error (MFE)	1.134705263

Data Fit

Validation Error Measures

Mean Absolute Percentage Error (MAPE)	8.696001034
Mean Absolute Deviation (MAD)	171.1979688
Mean Square Error (MSE)	40220.55472
Tracking Signal Error (TSE)	42.1540165
Cumulative Forecast Error (CFE)	7216.682
Mean Forecast Error (MFE)	112.7606563

Predictive Accuracy

Challenges

Missing values

Compute average metrics excluding missing values

Zero counts

MAE/RMSE: no problem

Cannot compute MAPE

- Exclude zero counts
- Use alternative (MASE)- see textbook

Forecast Accuracy vs. Profitability

Forecast errors can have asymmetric costs

- Large vs. small errors (or cross a threshold)
- Positive vs. negative errors

Under-estimate vs. over-estimate demand

Buy/sell stock: profitability depends on y_t not on e_t
Methods that shine in extreme periods but random at normal periods are **good**

Series: Bad science

Predictions are fine, but there are better ways to protect a population

Last year's earthquake in Abruzzo in Italy shows it is impossible to predict certain tragedies – but that hasn't stopped the seismologists being blamed

Ben Goldacre
The Guardian, Saturday 19 June 2010
Article history

Santa Maria churchin L'Aquila after the earthquake. Photograph: Christian Sinib the Guardian

22 October 2012 Last updated at 19:06 GMT

L'Aquila quake: Italy scientists guilty of manslaughter

The BBC's Alan Johnston in Rome says the prosecution argued that the scientists were "just too reassuring"

Six Italian scientists and an ex-government official have been sentenced to six years in prison over the 2009 deadly earthquake in L'Aquila.

Related Storie

L'Aquila quake

Beyond point forecasts

MEASURING
FORECAST UNCERTAINTY

Forecast/Prediction Interval

"Probability of 90% that the value will be in the range [a,b]"

Theoretical Formula

If the forecast errors are **normal**, prediction interval is

$$\mathbf{F_{t+k} \pm k\, \sigma}$$

σ = estimated standard deviation of forecast errors
= sqrt(MSE) from training set

k = some multiple
(k=1.645 gives 90% prediction interval)

Challenges to Formula

- Errors often non-normal
- If model is biased (over/under-forecasts), symmetric interval around F_{t+k}?
- Estimating the error standard deviation is tricky

One solution is transforming errors to normal

Empirical Solution

To construct prediction interval for 1-step-ahead forecasts

1. Create roll-forward forecasts (F_{t+1}) on validation period
2. Compute forecast errors
3. Compute percentiles of error distribution
   ($e^{(5)}$=5th percentile; $e^{(95)}$=95th percentile)
4. Prediction interval:

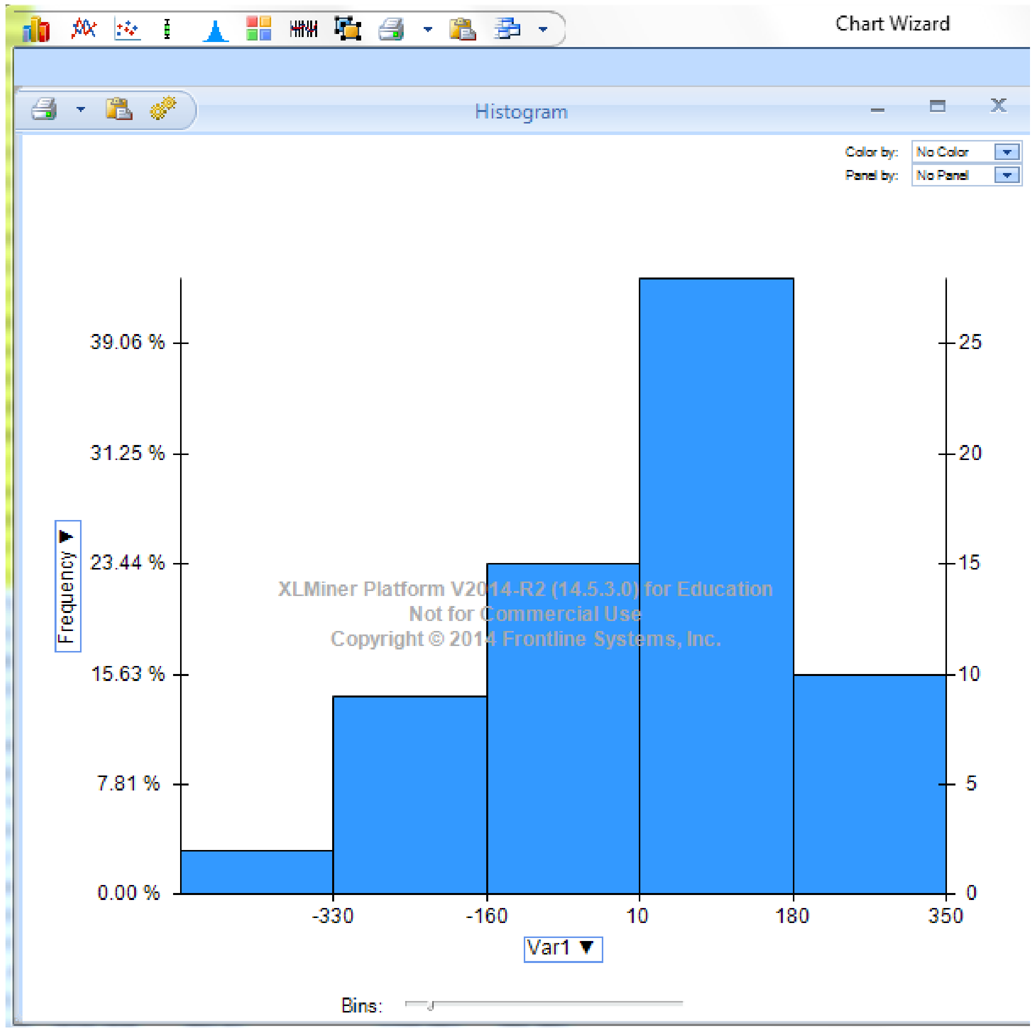
$$\left[\mathbf{F_{t+1} + e^{(5)}\ ,\ F_{t+1} + e^{(95)}} \right]$$

Distribution of roll-forward 1-step ahead forecast errors (validation period)

Not normal!

5^{th} percentile = -317
95^{th} percentile = 294

But it should be
$1.645*\sigma = 282$
where $\sigma = 171$.



90% prediction interval for 1-step ahead forecast F_{t+1}:

$$[F_{t+1} - 317 , F_{t+1} + 294]$$

For next class

Assignment #1: Visualization and Evaluating Naïve Forecasts

- Chap 2: Prob 5
- Chap 3: Prob 1
- Chap 9.1: Tips & Suggested Steps 1-4

Hints:

- For Tip 1 and 3, use Tableau's aggregation feature to explore different frequencies. Recall how YEAR(Month) in last class's demo summed up the months within each year.
- For Tip 3, think about how you would use a naïve forecast to capture multiple seasons.

Also for next class

Choose a time series & a goal for your team's project