Time Series Forecasting

Sumeet Gupta

Outline

- Introducing a Time Series
- Forecasting a time series
  - Dataset – Monthly Ridership Data for Amtrak Trains

Introduction

- Time series is a data set collected from a process with equally spaced periods of time
- Sampling adjacent points in time introduce a correlation

Examples

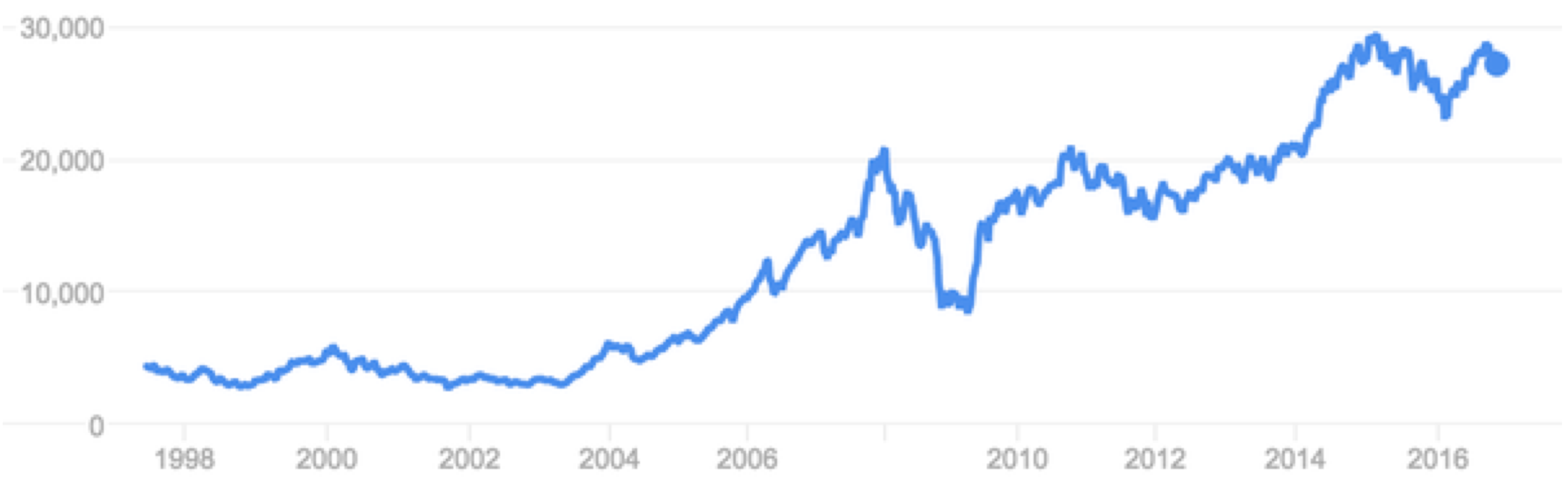
- Stock Market Data
- Dow Jones Industrial Averages
- Daily data on sales
- Monthly Inventory
- Daily Customers
- Monthly Interest rates

BSE SENSEX

INDEXBOM: SENSEX - 4 Nov, 3:40 PM IST

27,274.15 ↓156.13 (0.57%)

1 day | 5 day | 1 month | 3 months | 1 year | 5 years | max



Goals of Time Series Analysis

- Identifying the nature of the phenomenon represented by the sequence of observations
  - Useful for Policy Decision Making
- Forecasting (predicting future values of the time series variable)
  - Useful for predicting future values in a time series

Components of a Time Series

- Trend (T) – Gradual long term movement (up or down). Easiest to Detect. E.g., Population Growth in India
- Cyclical Patterns (S) – Results form events recurrent but not periodic in nature. Any up and down repetitive movement in demand. Repeat itself over a long period of time. E.g., Economic Recession
- Seasonal Patterns (S) – Results from events that are periodic and recurrent in nature. Any up and down repetitive movement within a trend occurring periodically. E.g., Sales during festive season
- Irregular Component (Noise) – Disturbance or residual variation that remains after all other behaviors have been accounted for. Erratic movements that are not predictable because they do not follow a pattern. E.g., Earthquake

Methodologies for Forecasting a Time Series

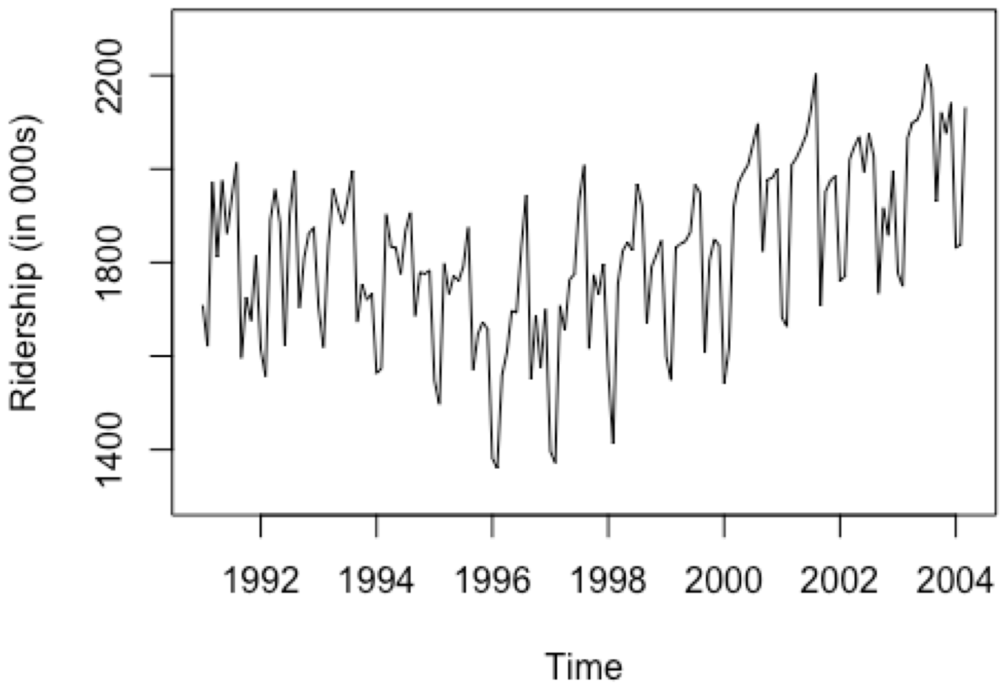
- Using Regression Models
  - Simple Linear Regression
  - Polynomial Trend
  - Trend +Season
- Using Smoothing Methods
  - Moving Average
  - Exponential Smoothing

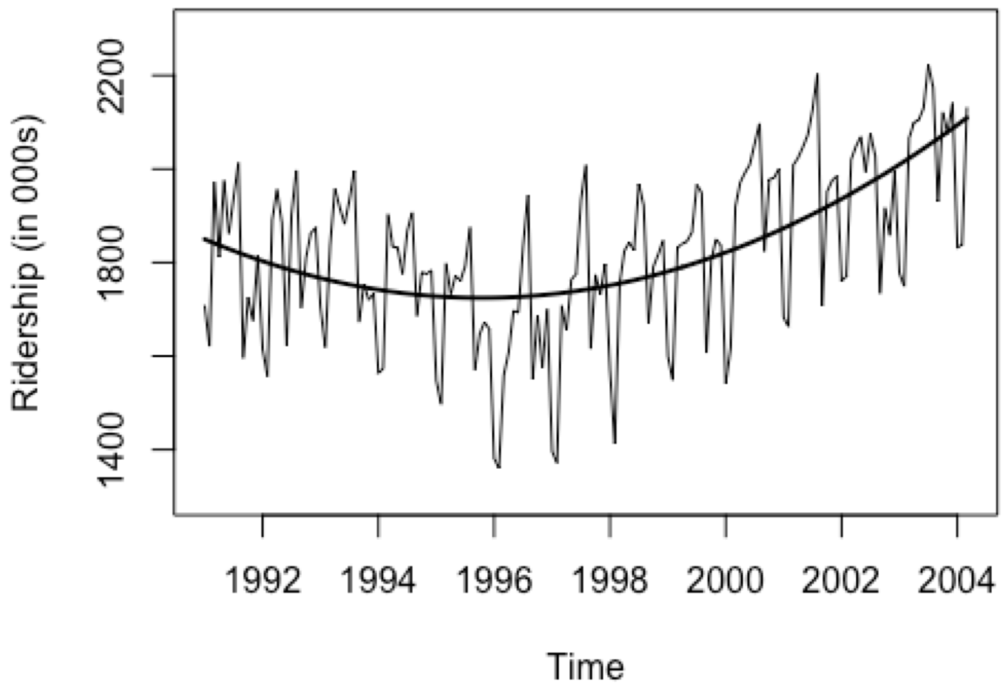
Methodology

- Divide the series into training and validation series
- Build a suitable model (Trend, Polynomial, Season, Trend + Season)
- Check the RMSE / MAPE of each model. The model with lowest RMSE / MAPE is a good model
- Check the Autocorrelation (Acf) in the Model for any significant lags as well as in the residuals.
- Use the residuals to build a forecasting model depending upon the lag in the ACF plot.
- Finally check the predictability of a time-series. If the series is a random walk, it cannot be predicted

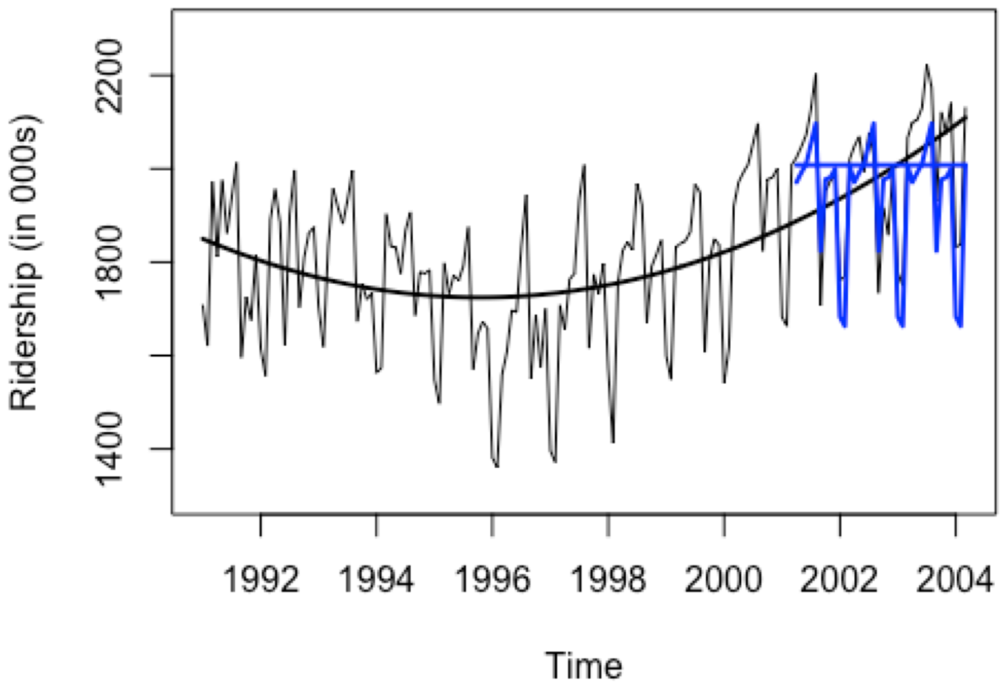
Time Series Forecasting Using Regression

Amtrak Data: Fitting a Linear Trend





Partitioning a Time Series



- Partitioning is not done at random as in Regression
  - A part of the time series from the beginning is considered as the Training Data
  - Rest of the data comprises Test Data
- Performance of the model on the Test Dataset determines the accuracy of our model

Setting the Baseline

- Setting the Baseline
  - Naive Forecast

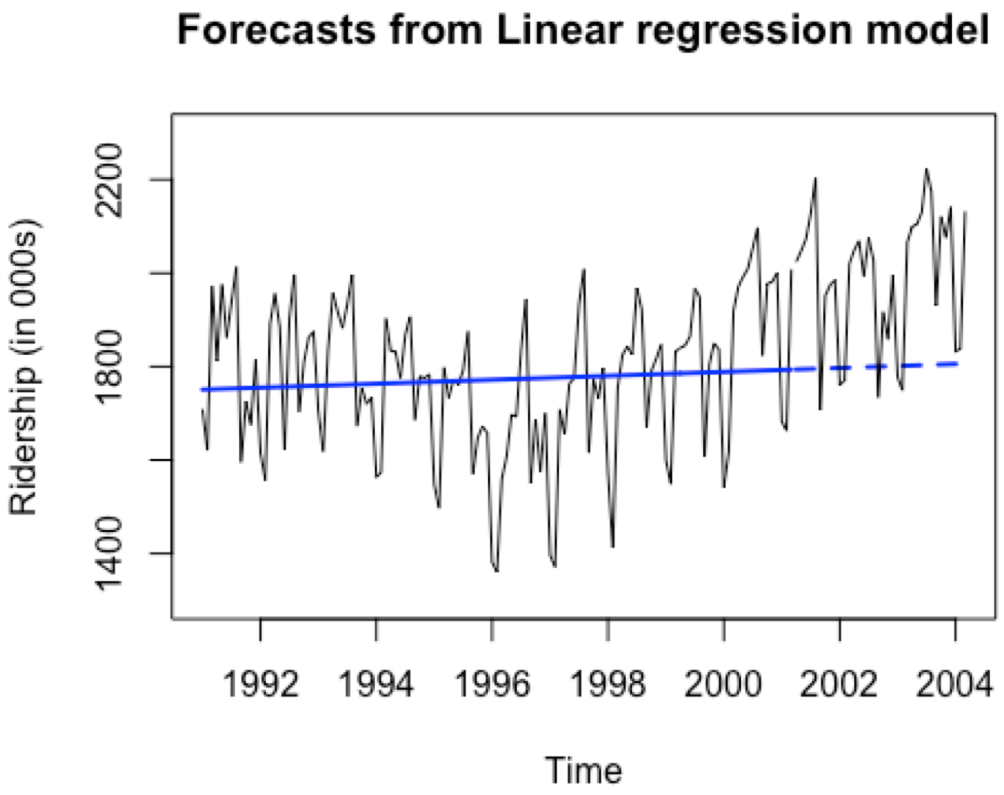
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	2.45082	168.1426	125.3033	-0.3459458	7.271436	1.519408	-0.2472638	NA
Test set	-14.77778	142.7186	115.8889	-1.2776915	6.019722	1.405251	0.2761603	0.8345144

  - Seasonal Naïve Forecast

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	13.90991	99.20586	82.46847	0.5835948	4.714103	1.000000	0.6403247	NA
Test set	54.88889	95.73285	84.16667	2.6609067	4.251513	1.020592	0.6359060	0.553814

A Model with Linear Trend

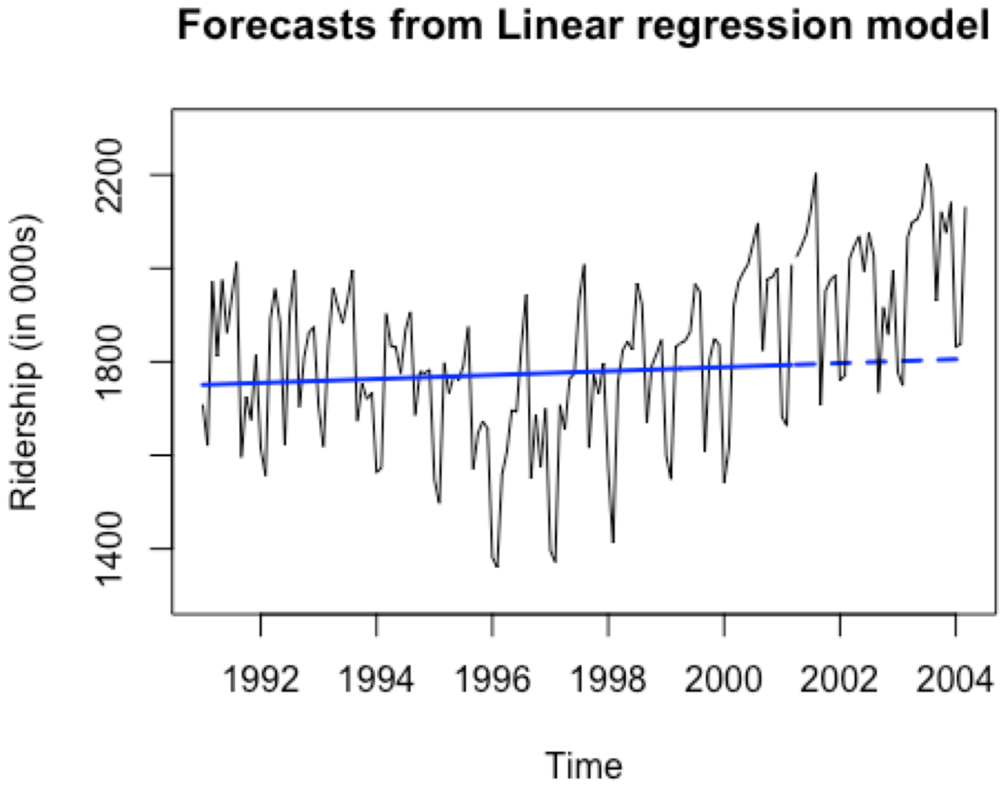
$Y_t = \beta_0 + \beta_1 t + \epsilon,$ level (β_0), trend (β_1), and noise (ϵ)

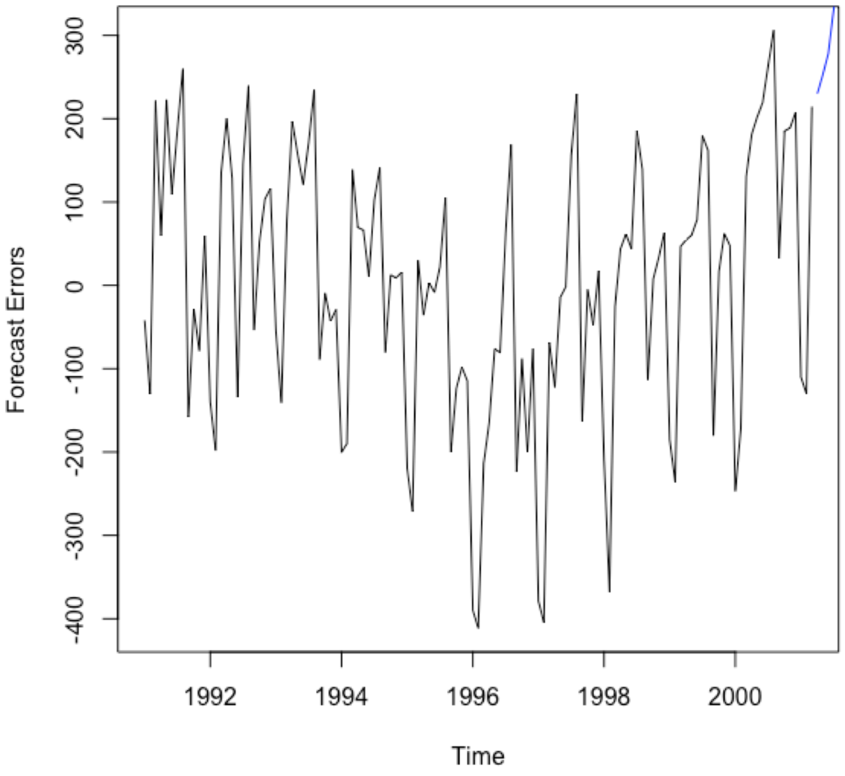



```
> summary(train.lm)

Call: lm(formula = formula, data = "train.ts", na.action = na.exclude)

Residuals:
    Min      1Q  Median      3Q     Max
-411.29 -114.02   16.06  129.28  306.35

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1750.3595    29.0729  60.206   <2e-16 ***
trend          0.3514     0.4069   0.864     0.39
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 160.2 on 121 degrees of freedom
Multiple R-squared:  0.006125,       Adjusted R-squared:  -0.002089
F-statistic: 0.7456 on 1 and 121 DF,  p-value: 0.3896
```

A Model with Linear Trend

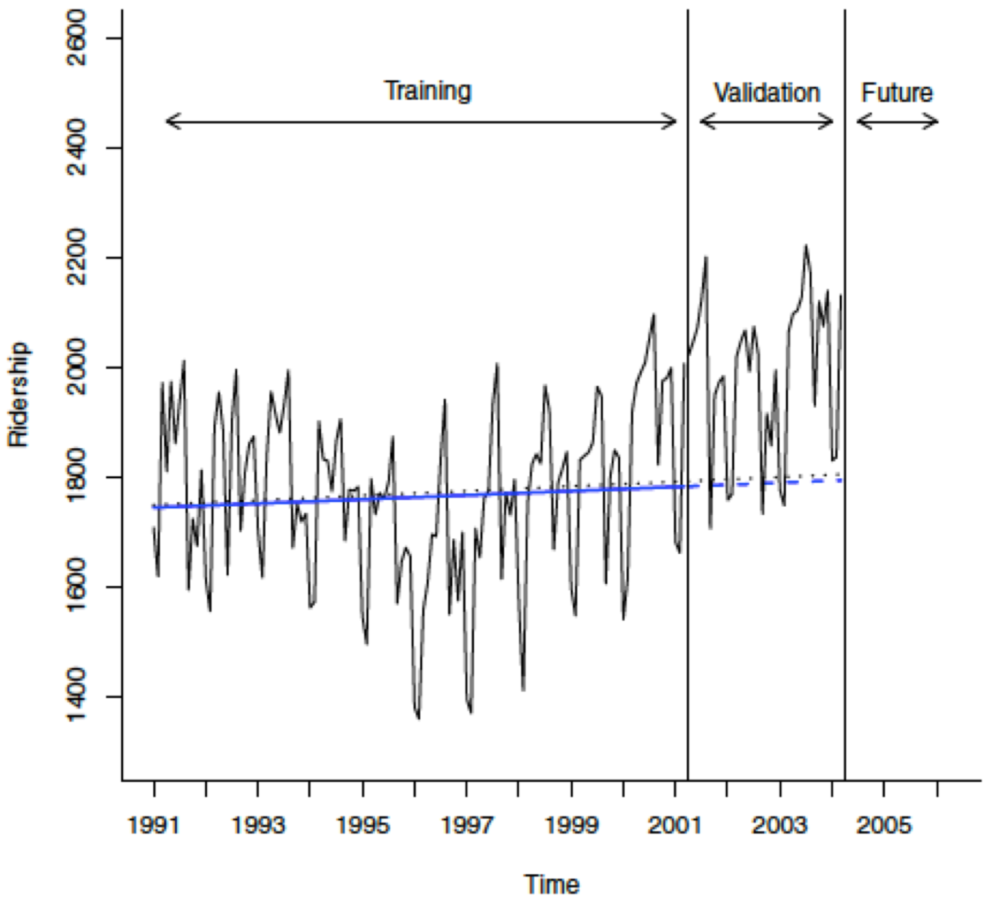
$Y_t = \beta_0 + \beta_1 t + \epsilon$, level ($\beta_0$), trend ($\beta_1$), and noise ($\epsilon$)





	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	7.367977e-15	158.8978	129.6662	-0.8536018	7.535088	1.572313	0.4370243	NA
Test set	1.931962e+02	239.5155	209.4629	9.2133989	10.148966	2.539914	0.2731996	1.35863

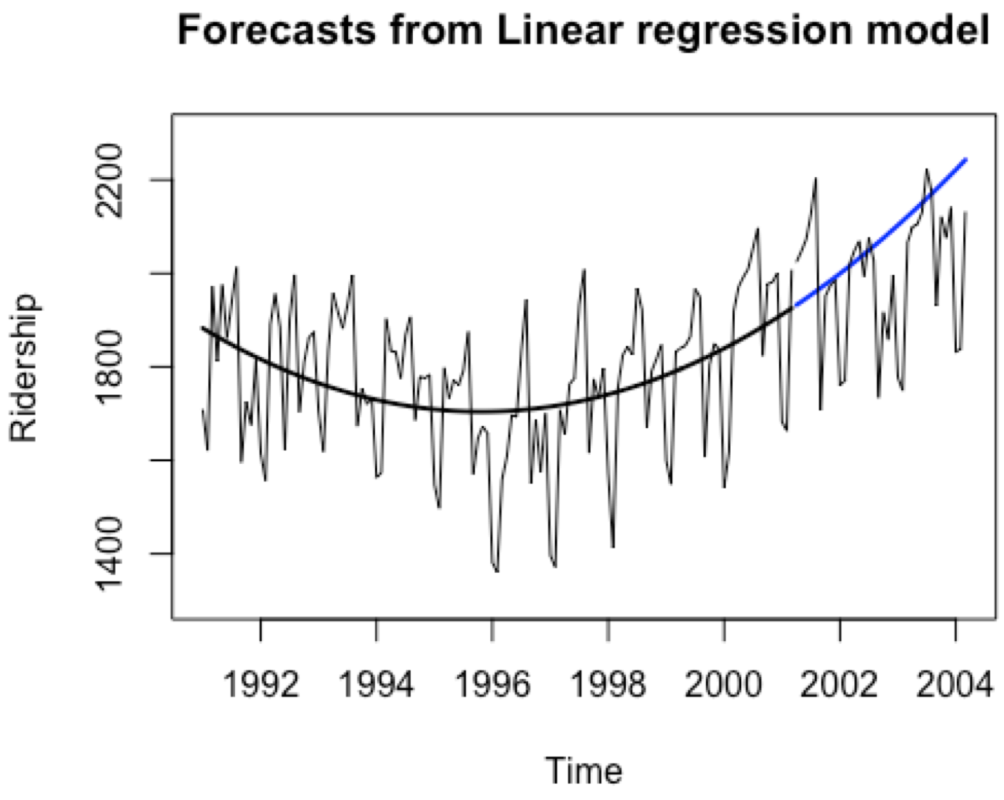
A Model with Exponential Trend



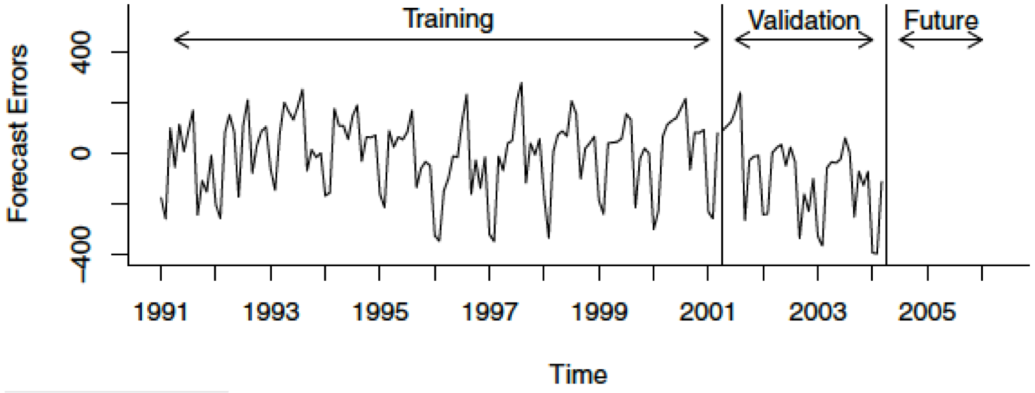
$(Y_t = ce^{\beta_1 t+\epsilon})$ $(\log Y_t = \beta_0+\beta_1 t+\epsilon)$

```
                       ME     RMSE      MAE        MPE      MAPE     MASE      ACF1
Training set -1.825888e-15 158.8978 129.6662 -0.8536018  7.535088 1.572313 0.4370243
Test set      1.931962e+02 239.5155 209.4629  9.2133989 10.148966 2.539914 0.2731996
             Theil's U
Training set        NA
Test set       1.35863
> accuracy(train.lm.expo.trend.pred, valid.ts)
                     ME     RMSE      MAE        MPE      MAPE     MASE      ACF1
Training set   7.371984 159.0621 130.5554 -0.4343999  7.553287 1.583094 0.4367924
Test set     203.331919 247.7809 216.2570  9.7244828 10.469088 2.622299 0.2734325
             Theil's U
Training set        NA
Test set      1.406161
```

A Model with Polynomial Trend

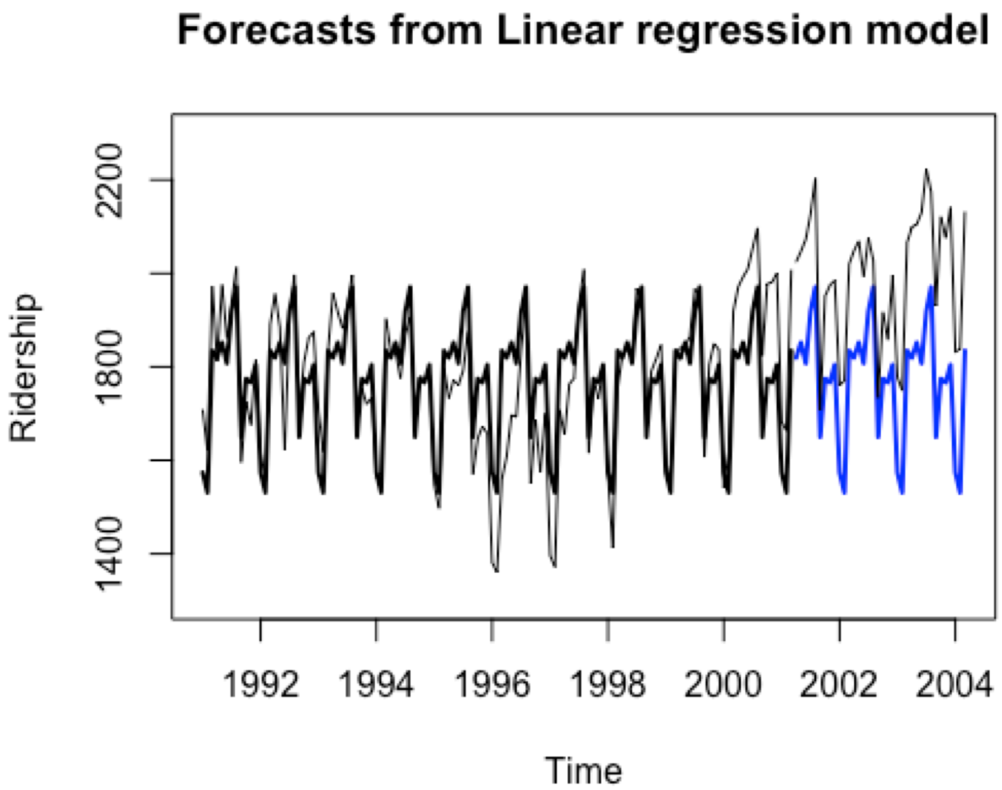


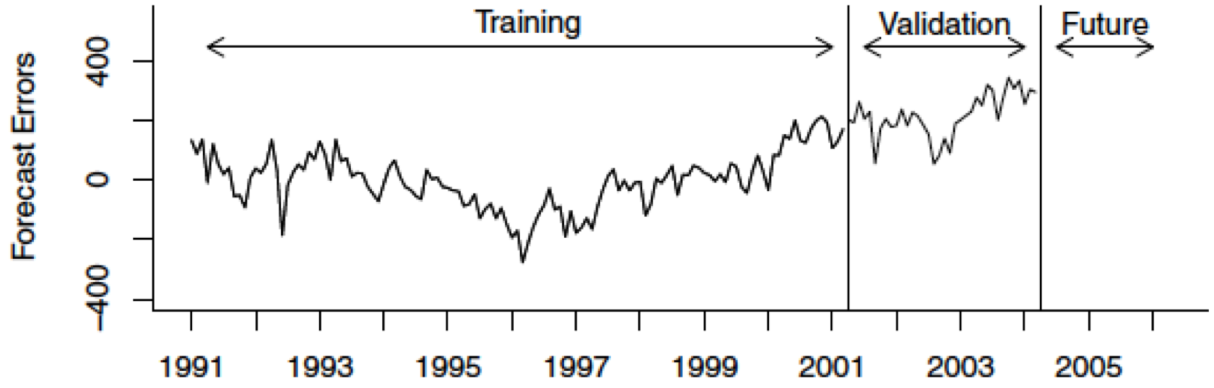
$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon.$$



	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-1.294397e-14	146.9583	120.3771	-0.734747	7.014330	1.459675	0.3439566	NA
Test set	-8.377213e+01	179.7701	133.6305	-4.715735	7.069862	1.620383	0.4171593	1.061603

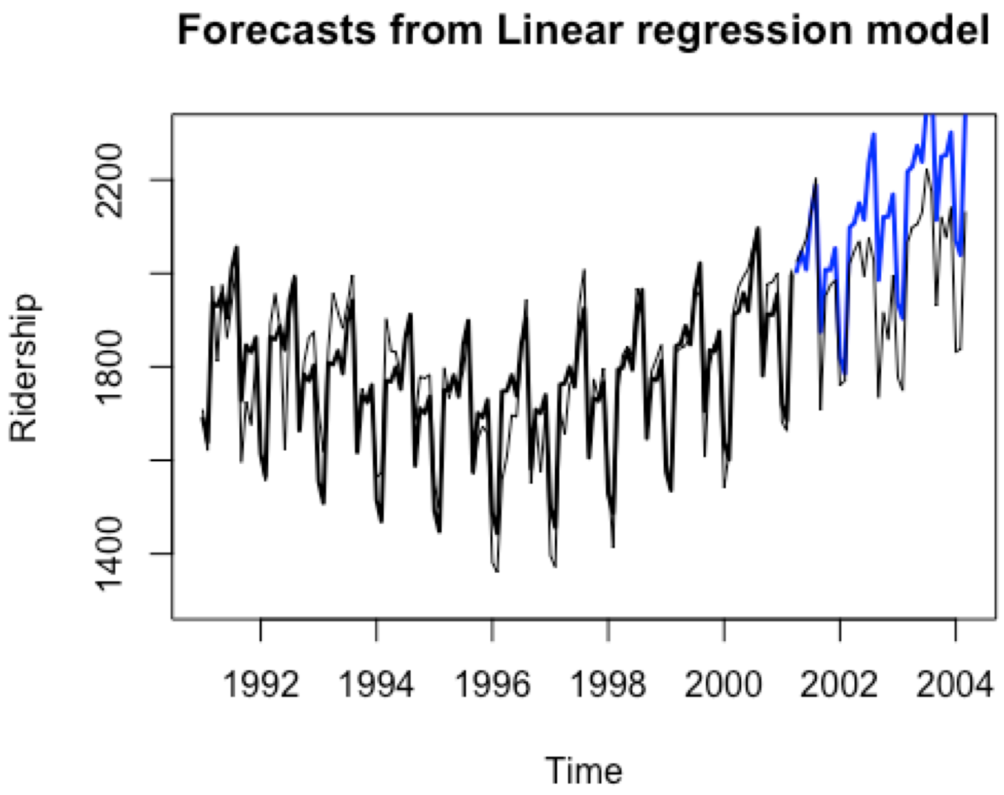
A Model with Seasonality

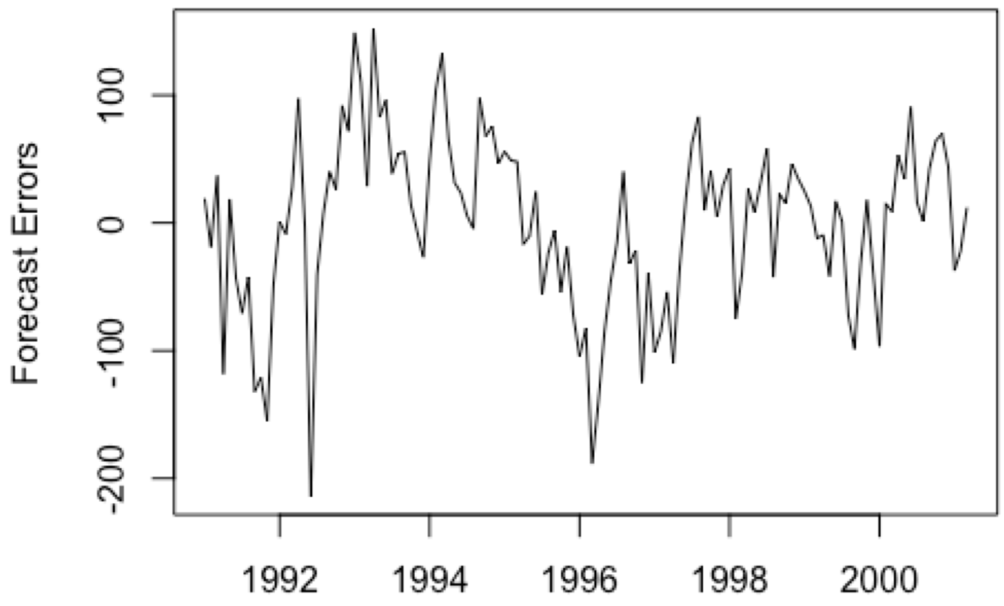




	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	5.540284e-15	96.28793	75.11101	-0.3095729	4.325658	0.9107846	0.7856258	NA
Test set	2.179056e+02	229.61573	217.90556	10.8636887	10.863689	2.6422893	0.6346724	1.330757

A Model with Trend and Seasonality

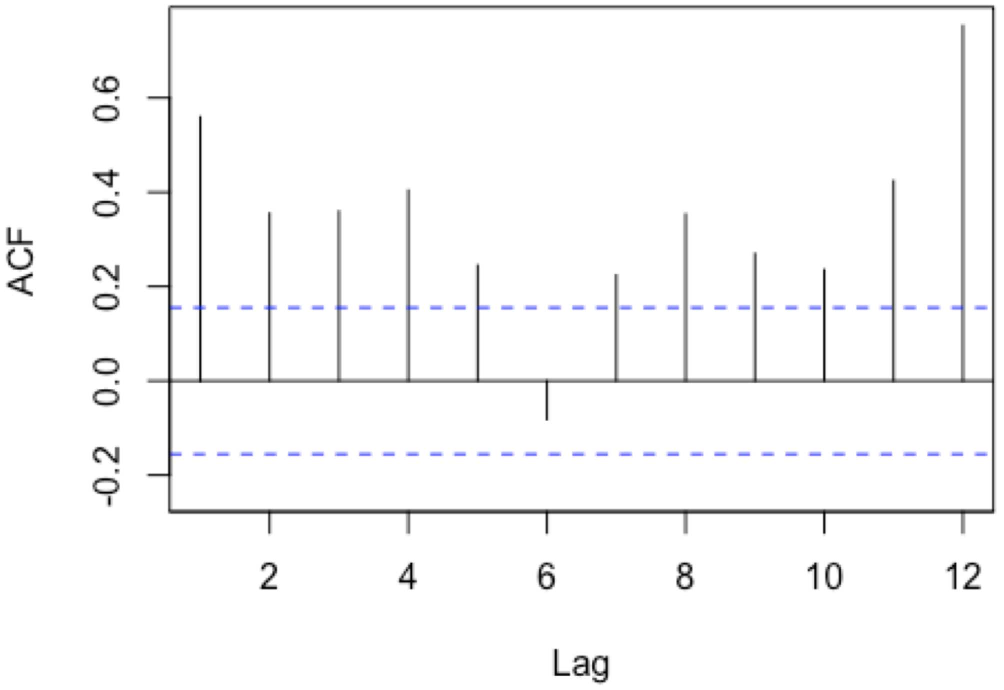




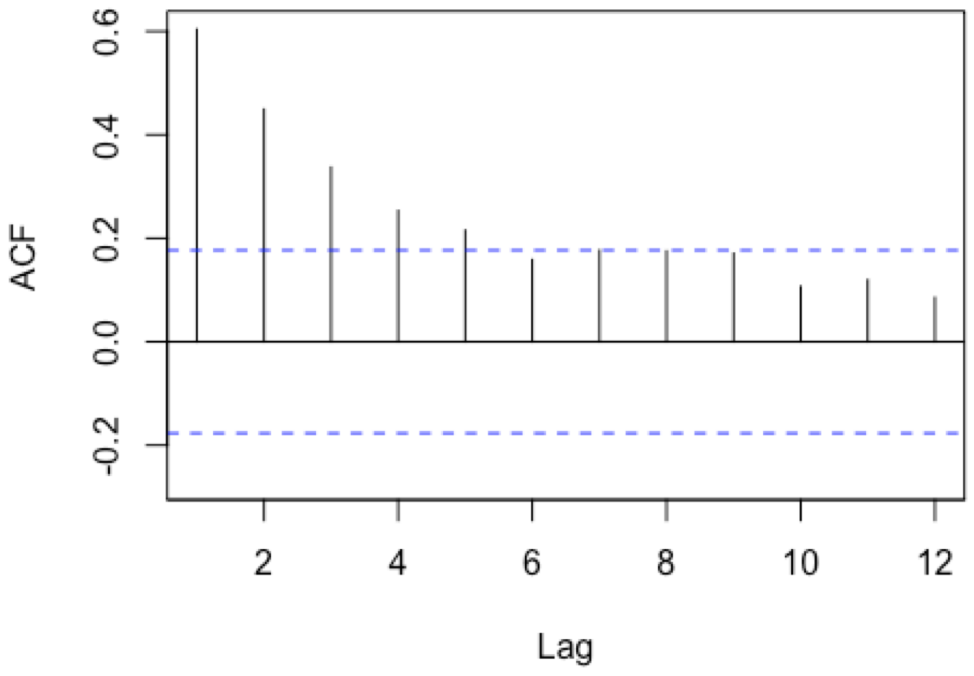
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	1.854553e-15	66.72437	51.93312	-0.1523749	3.014258	0.629733	0.6039636	NA
Test set	-1.259593e+02	153.12581	131.58966	-6.4211573	6.691712	1.595636	0.7073279	0.8955116

Time Series Forecasting: Autocorrelation and ARIMA Models

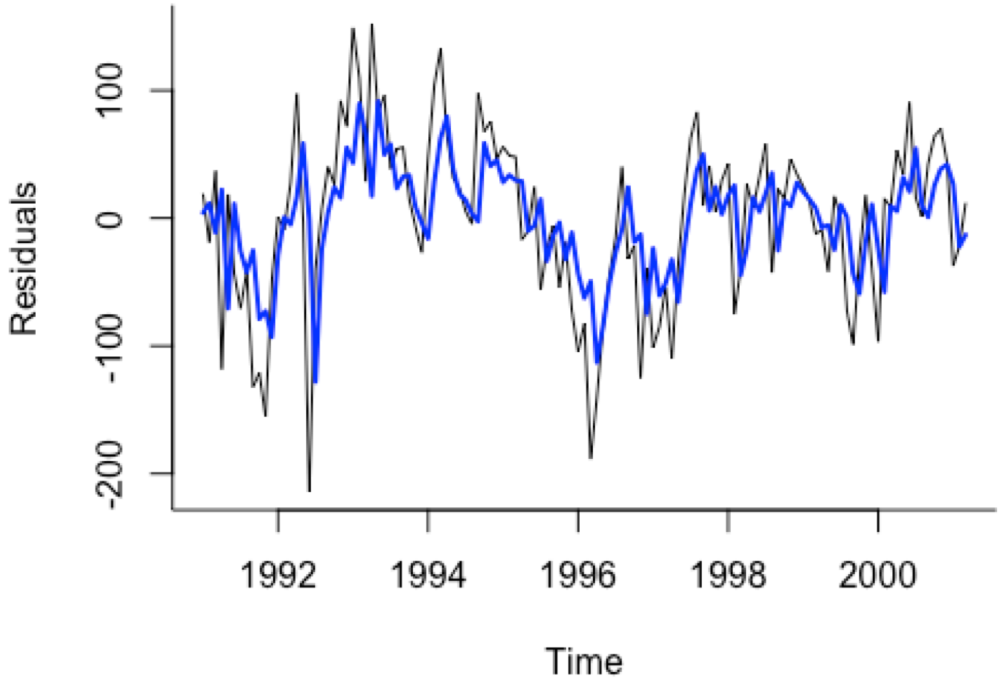
- Here the lag is present due to lag at 12 and negative correlation at 6

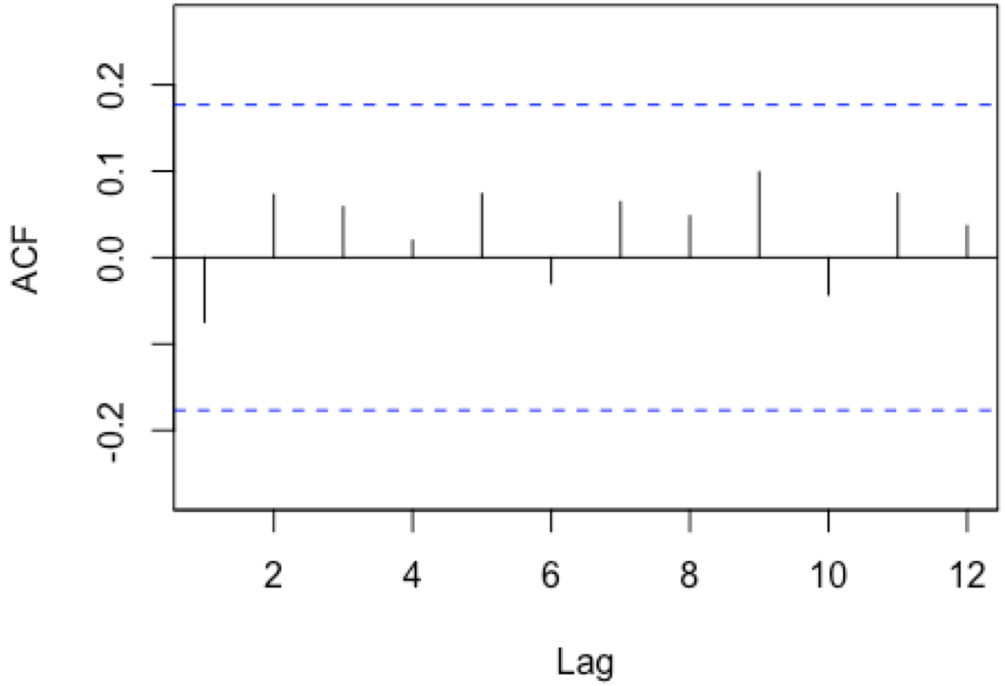


Here the lag is present at 1



Time Series Forecasting: Fitting AR(1) Model





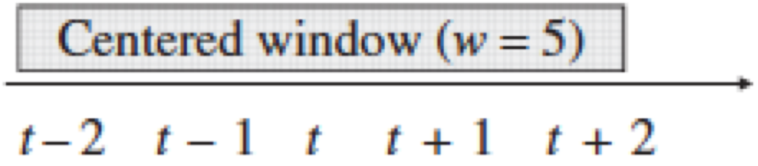
Evaluating Predictability of a Time Series

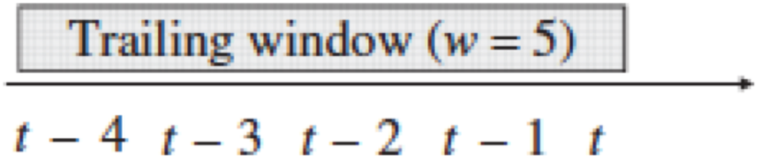
- Test whether the series is a random walk
- A random walk is a series in which changes from one time period to the next are random.
- A random walk is a special case of an AR(1) model, where the slope coefficient is equal to 1
- Forecasts from such a model are basically equal to the most recent observed value
- #(the naive forecast), reflecting the lack of any other information.
- #If a series is a random walk, it cannot be predicted (E.g., Stock Prices)

Forecasting using Smoothing Methods

Moving Average: Centered Vs Trailing

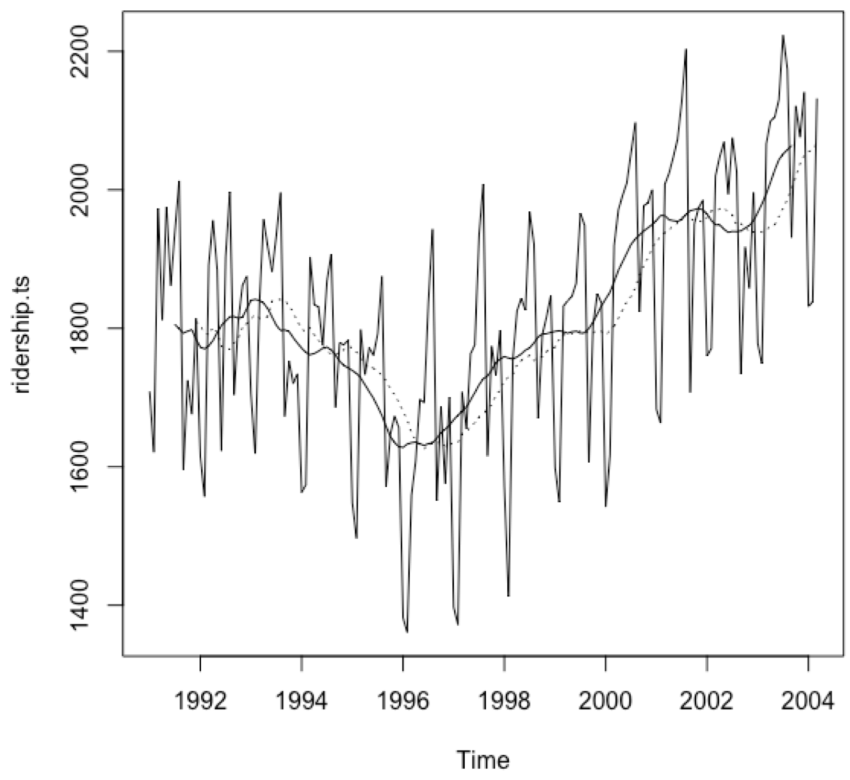
$$\text{MA}_t = \left(Y_{t-(w-1)/2} + \cdots + Y_{t-1} + Y_t + Y_{t+1} + \cdots + Y_{t+(w-1)/2}\right)/w.$$





$$F_{t+k} = \left(Y_t + Y_{t-1} + \cdots + Y_{t-w+1}\right)/w$$

Moving Average: Centered Vs Trailing



Exponential Smoothing

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \dots,$$

Advanced Exponential Smoothing

- Series with a Trend

$$F_{t+k} = L_t + kT_t$$

  - *Updating the Series*

$$\begin{aligned} L_t &= \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta\,(L_t - L_{t-1}) + (1-\beta)T_{t-1}. \end{aligned}$$

Advanced Exponential Smoothing

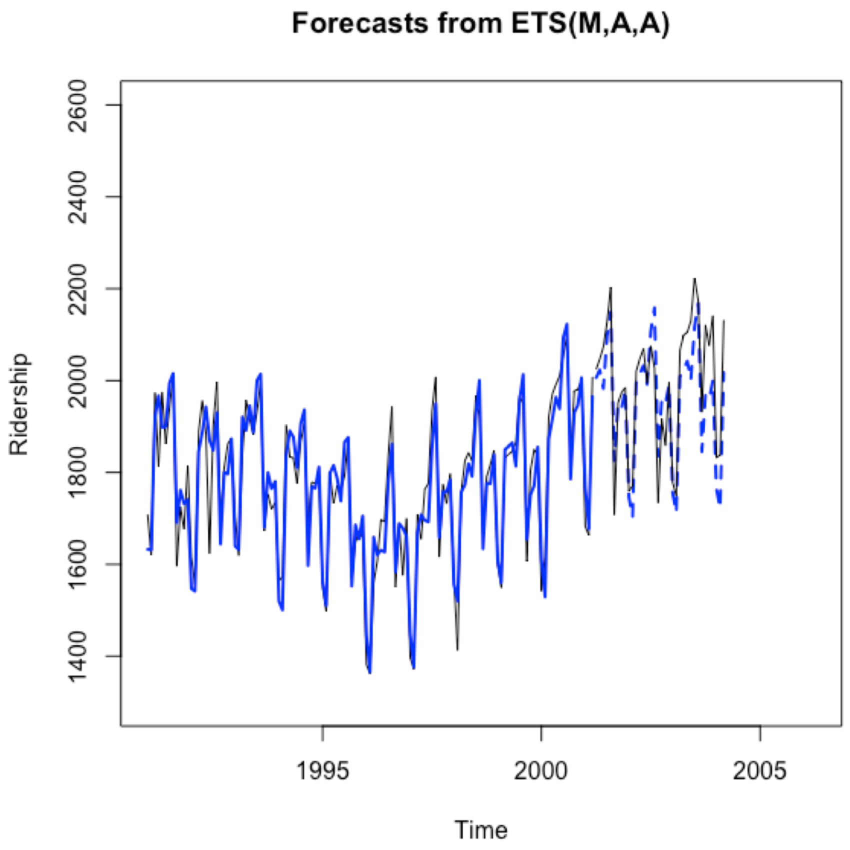
- Series with Seasonality (Run Holt Winter's Exponential Smoothing)

$$F_{t+k} = (L_t + kT_t)\, S_{t+k-M}$$

  - *Updating the Series*

$$\begin{aligned} L_t &= \alpha Y_t / S_{t-M} + (1-\alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \beta\,(L_t - L_{t-1}) + (1-\beta) T_{t-1} \\ S_t &= \gamma Y_t / L_t + (1-\gamma) S_{t-M}. \end{aligned}$$

Advanced Exponential Smoothing




```
ETS(M,A,A)
Call:
 ets(y = train.ts, model = "MAA")
  Smoothing parameters:
    alpha = 0.5483
    beta  = 1e-04
    gamma = 1e-04
  Initial states:
    l = 1881.6423
    b = 0.4164
    s=27.1143 -10.6847 -2.9465 -121.1763 201.1625 147.3359
           37.6688 75.8711 60.4021 44.4779 -252.047 -207.1783
  sigma:  0.0317
     AIC     AICc      BIC
1614.219 1619.351 1659.214
```

	IE	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0.17363	56.46385	45.15148	-0.07486935	2.574531	0.5474999	0.06005373	NA
Test set	33.55969	76.84633	62.29953	1.57731008	3.130149	0.7554345	0.61477472	0.4444268