Inspire…Educate…Transform.

Supervised models

Time Series Forecasting

Dr. Anand Jayaraman
ajayaraman@gmail.com

Jan 29, 2017

Thanks to Dr.Sridhar Pappu for the material

"Prediction is very difficult, especially if it's about the future."

--Niels Bohr, Nobel laureate in Physics

What is Time Series data?

- A sequence of data points in successive order, indexed by time.

$$y_t,\ y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, \cdots$$

- Eg: Population of the country listed yearwise, Temperature in the city listed by the hour, Number of iPhones sold listed for each quarter

Forecasting

- Factors needed to forecast the stock price of Tata Motors:
  - Current Sales, Revenue and profit data
  - Sales trend
  - Level debt carried by the company
  - Competition
  - Import/export rules
  - Interest rate environment
  - US/INR exchange rate
  - Tax rates
  - Crack down on black money?
  - Cost of steel?
  - Number of smart phones sold?

Forecasting

$$\hat{y}_{t+1} = g(t, x_1, x_2, x_3 \dots, y_t, y_{t-1}, y_{t-2}, \dots)$$

g might be some complex linear or nonlinear function.

Time series forecasting attempts to do same forecast just using the past data of y, without relying on any other external predictors (x_i).

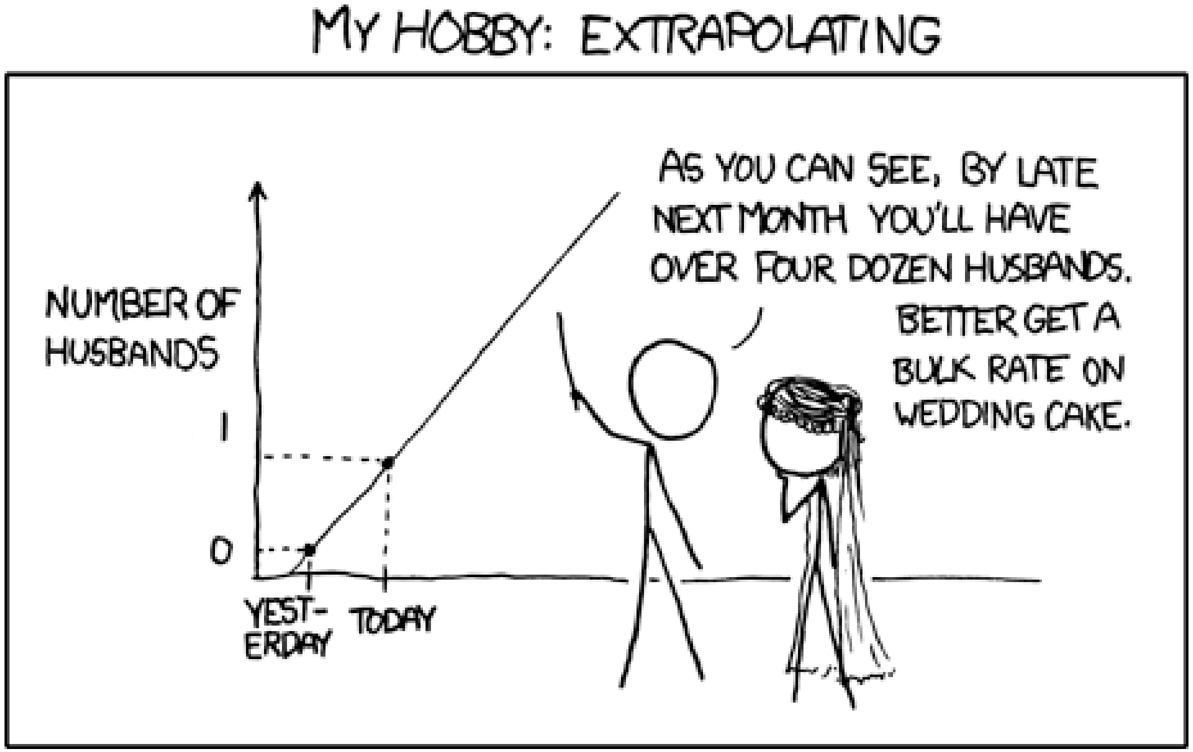
Typical Time Series

$$\hat{y}_{t+1} = f(t, y_t, y_{t-1}, y_{t-2} \cdots)$$

f can be linear or nonlinear function

Why Time Series forecasting?

- Causal independent variables are
  - Unknown to us
  - Not available
  - Might not fit the data well
  - Difficult to forecast



FORECASTING THROUGH TREND ANALYSIS

CSE 7202c

Regression with time

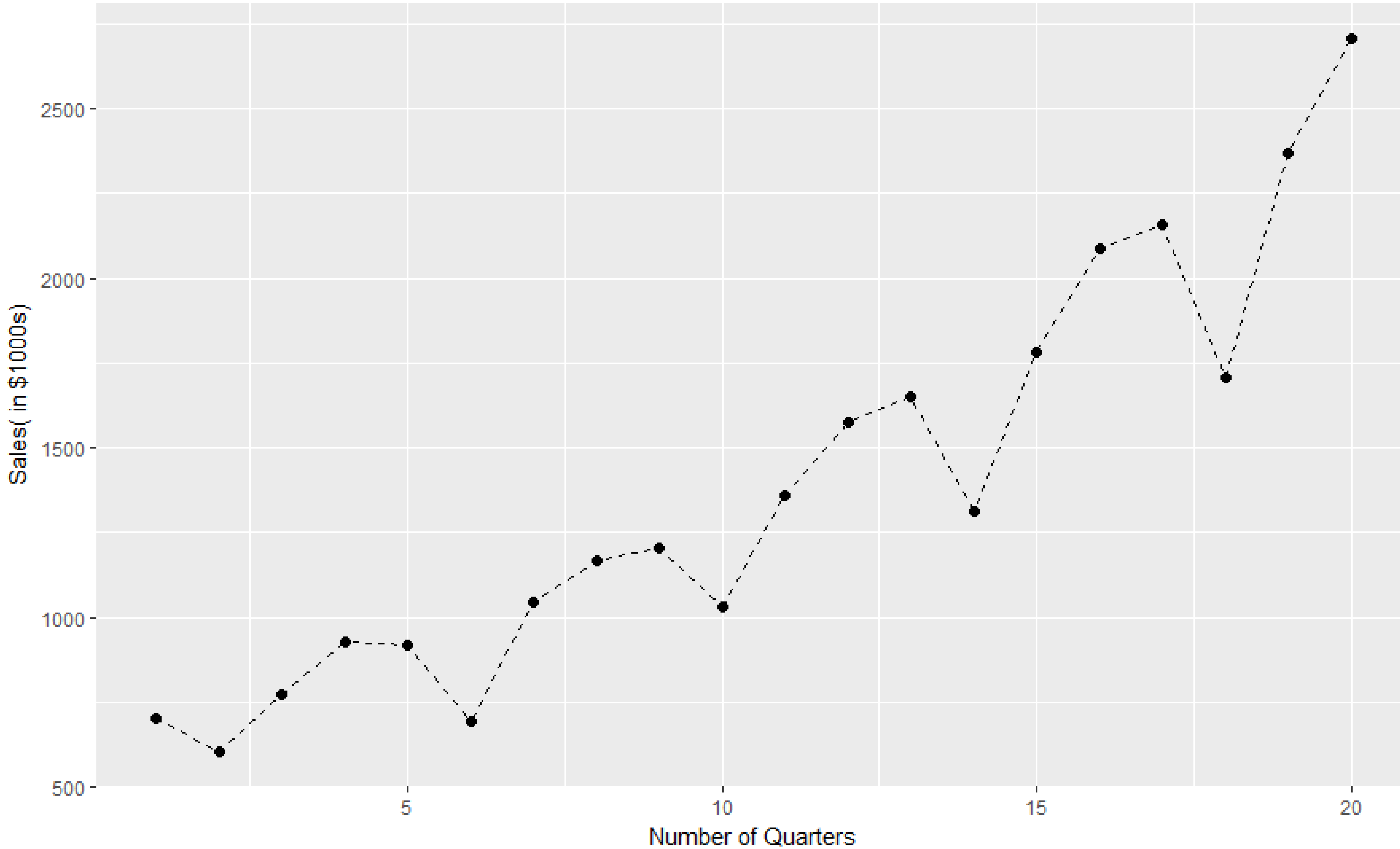
$$\hat{y}_{t+1} = f(t)$$

Regression on Time

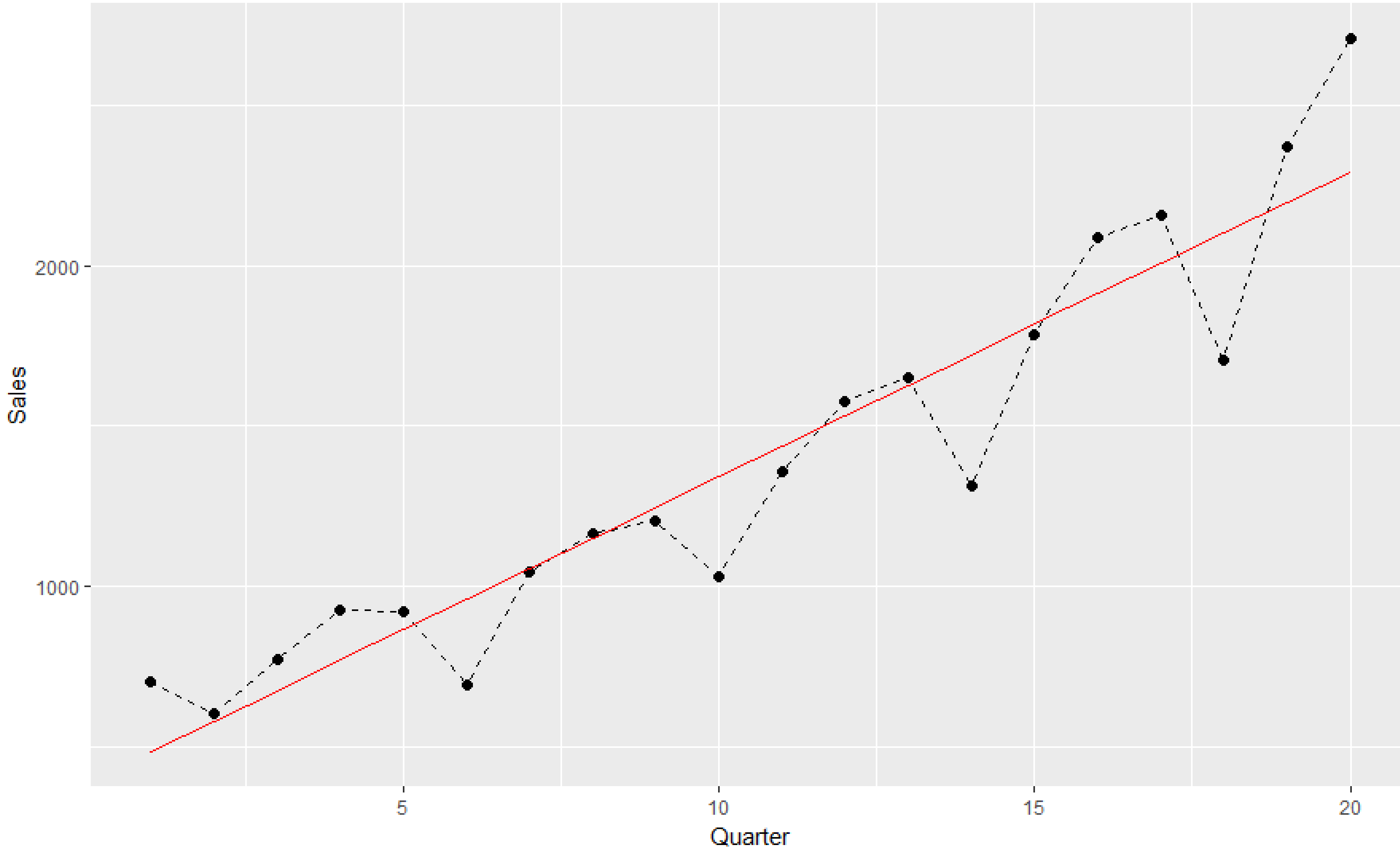
- Use when trend is the most pronounced

- ACF decays exponentially and PACF has very few spikes

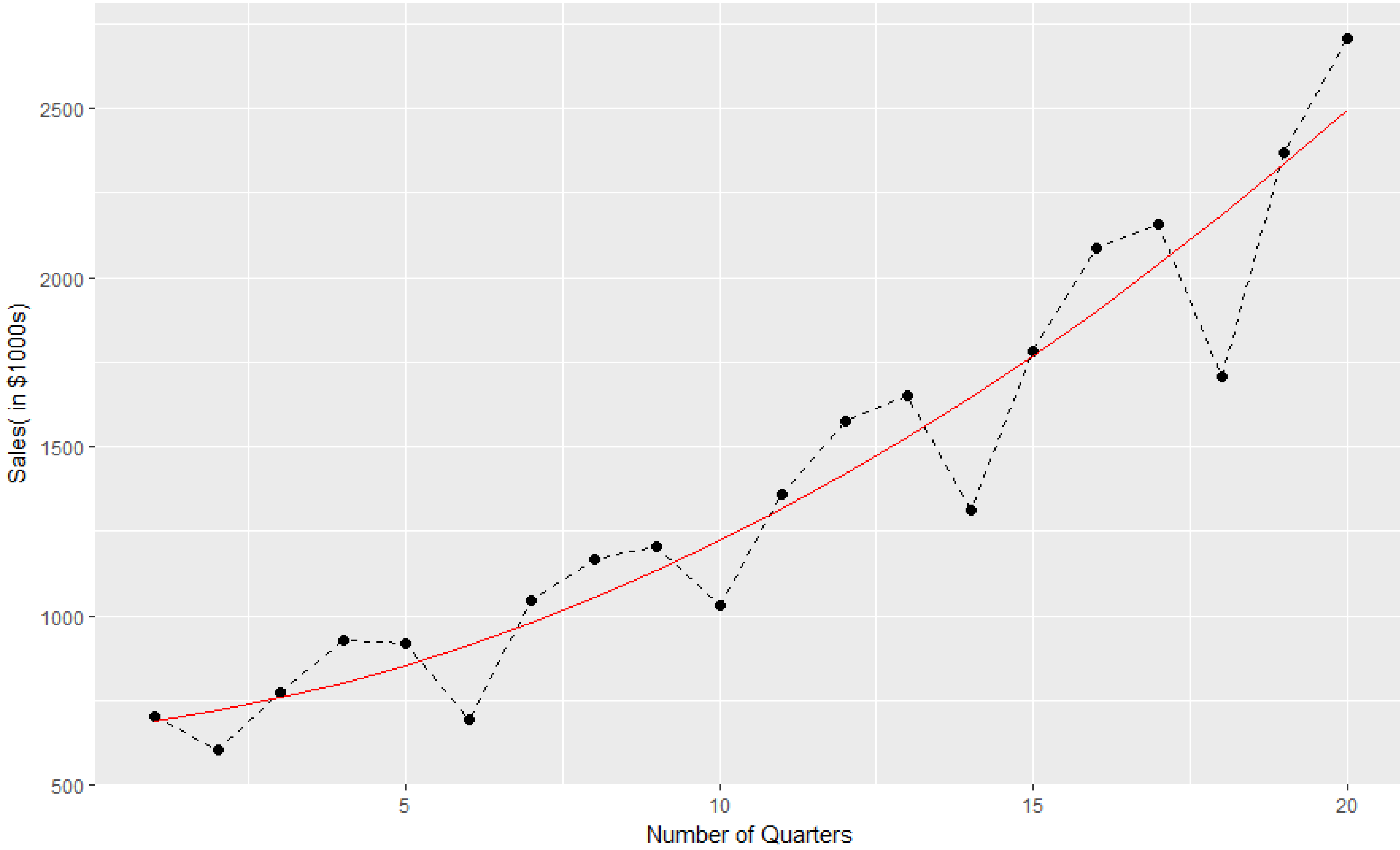
Seasonality



Regression Analysis – Linear fit



Quadratic Trend



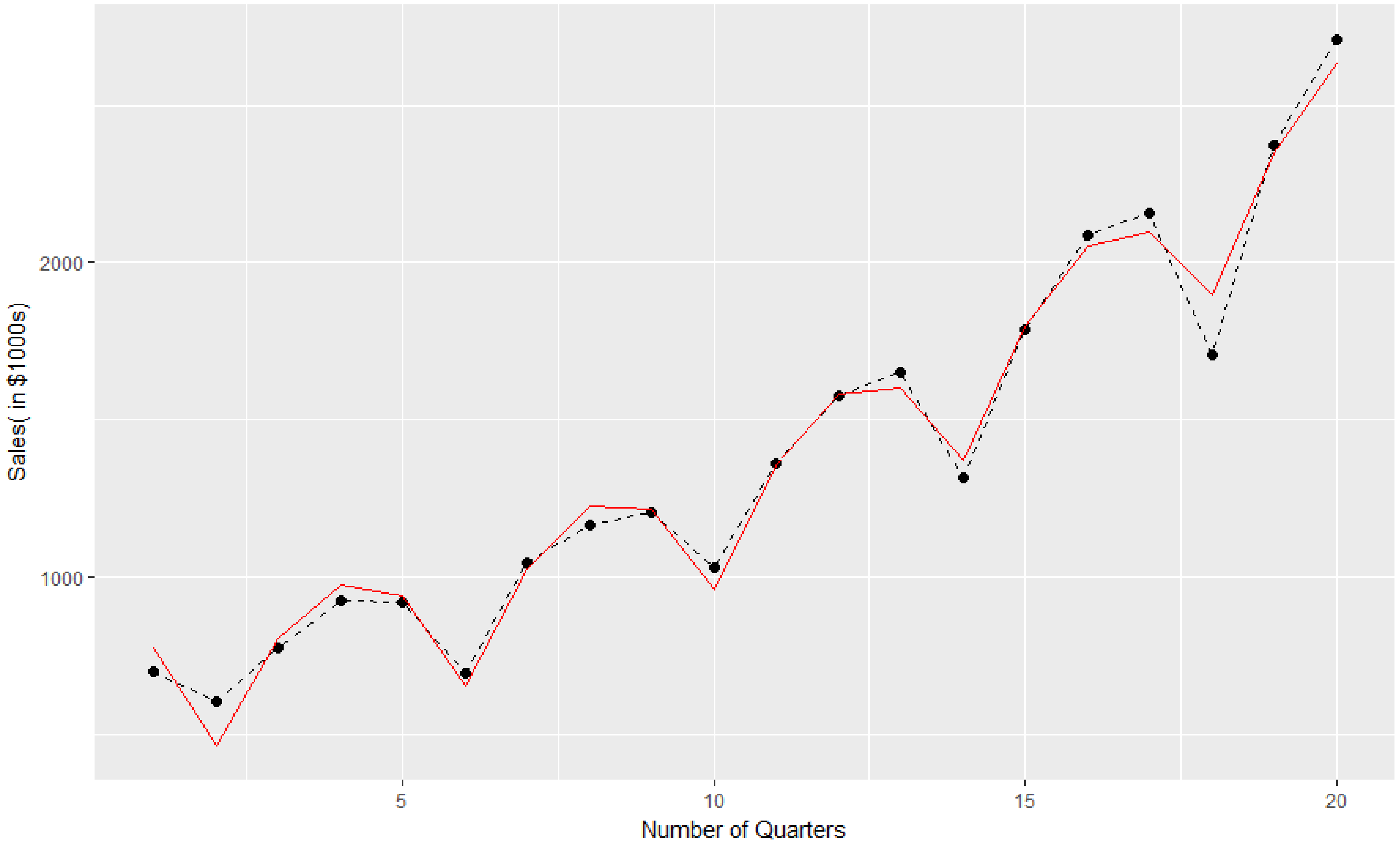
Seasonal Regression Models

	Value of		
Quarter	X_{3_t}	X_{4_t}	X_{5_t}
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

$$Y_t = \beta_0 + \beta_1 X_{1_t} + \beta_2 X_{2_t} + \beta_3 X_{3_t} + \beta_4 X_{4_t} + \beta_5 X_{5_t} + \varepsilon_t$$

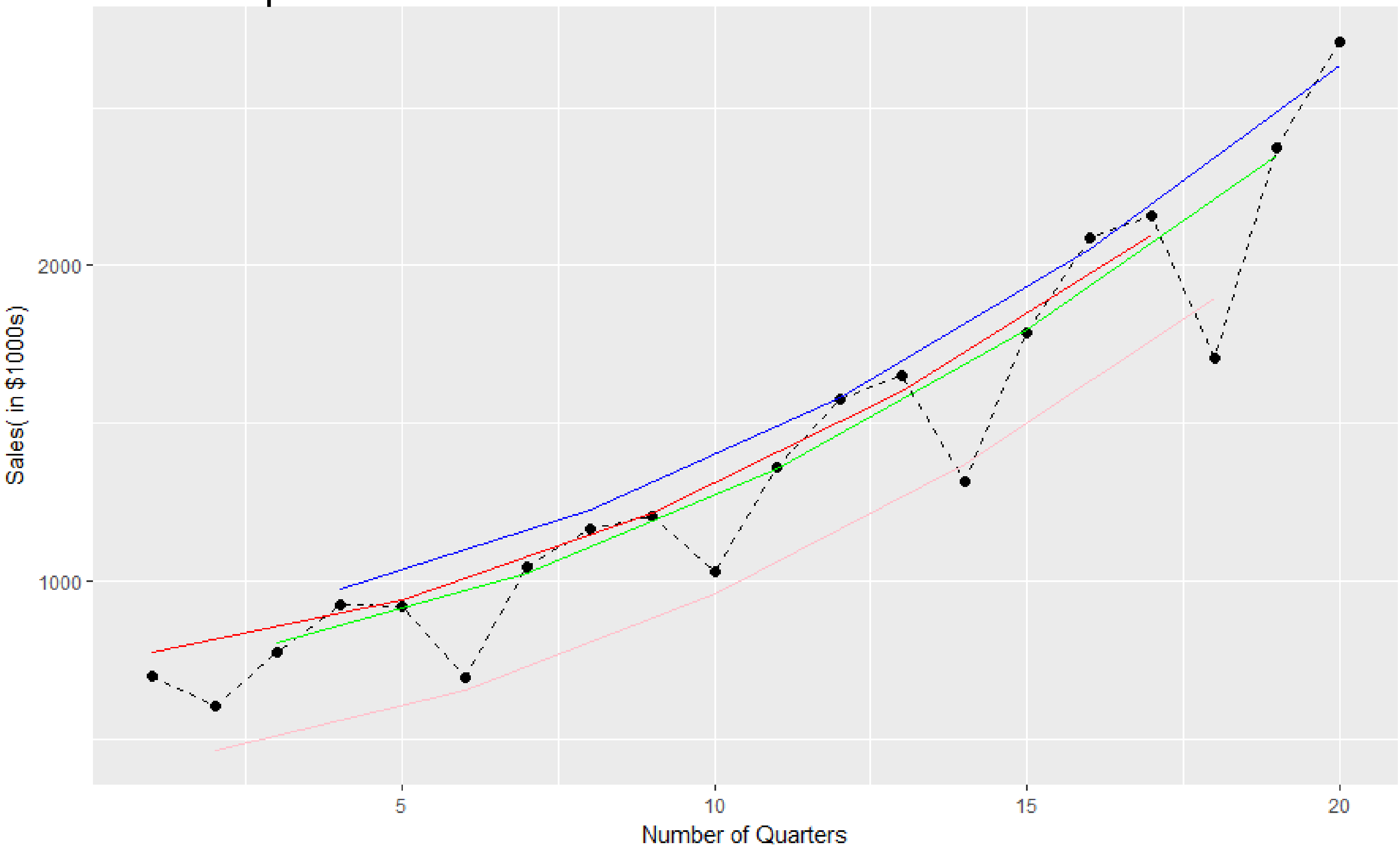
where, $X_{1t} = t$ and $X_{2t} = t^2$.

Seasonal Regression Models



Quadratic fit with seasonality

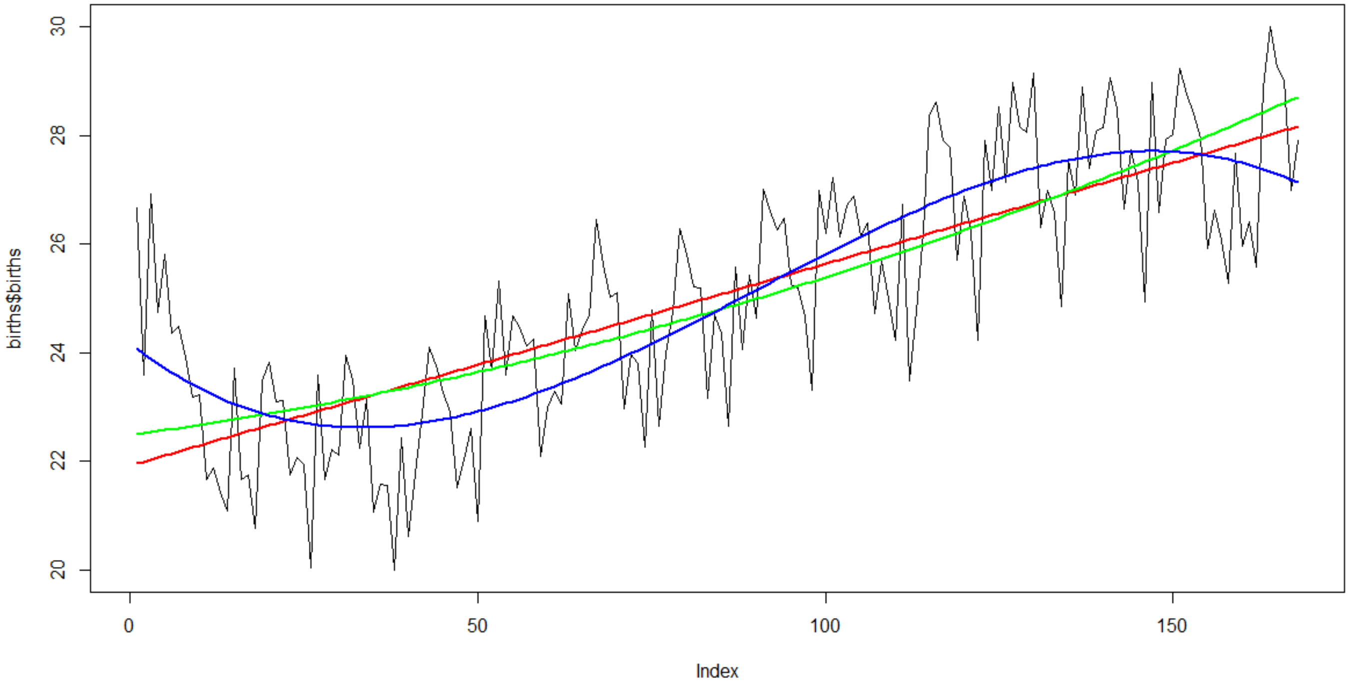
Plotting the fitted data-points separately for each quarter, shows how R manages to do such a good fit. Its basically fitting a quadratic line for each quarter with a different intercept.



CSE 7202c

Seasonal Regression Models

Seasonal Regression Models - Births

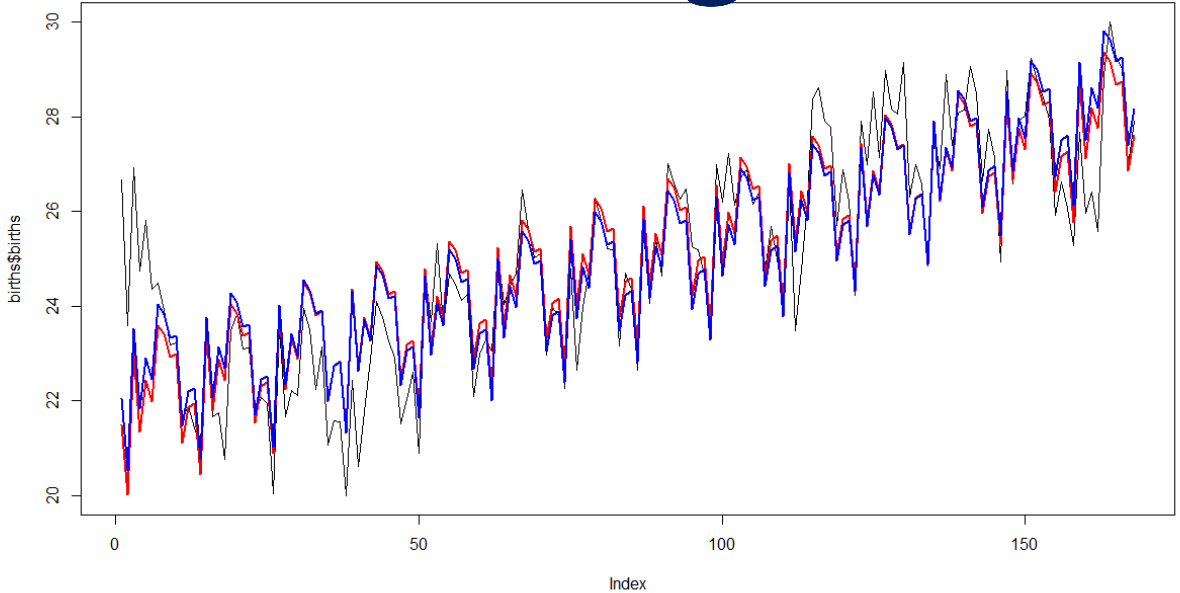


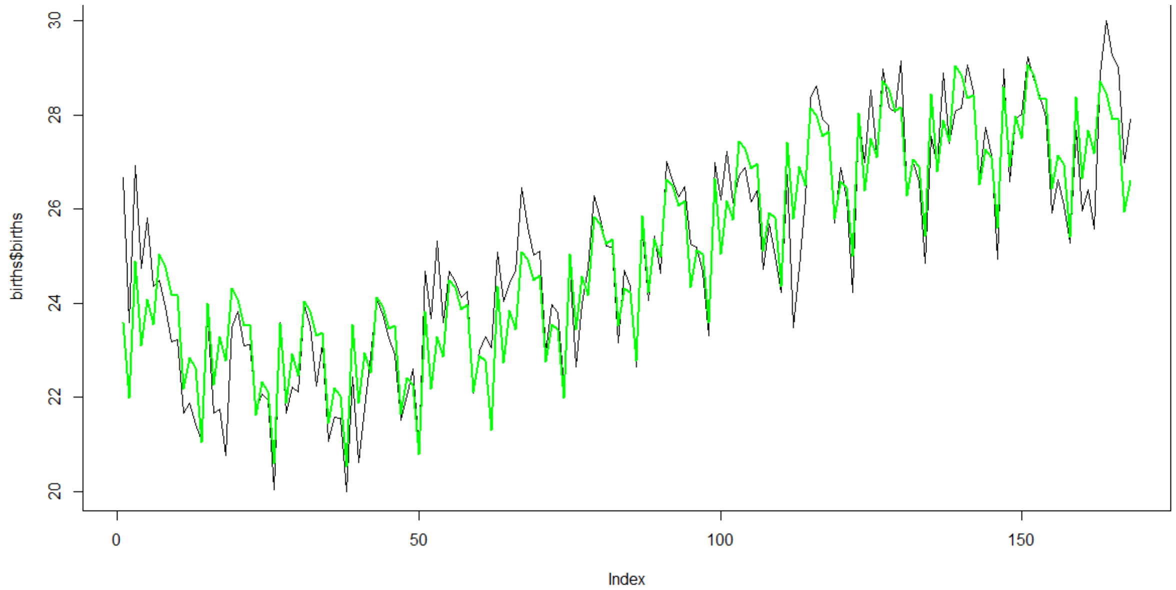
Data Editor

File Edit Help

	births	time	var3	var4	var5	var6
1	26.663	1				
2	23.598	2				
3	26.931	3				
4	24.74	4				
5	25.806	5				
6	24.364	6				
7	24.477	7				
8	23.901	8				
9	23.175	9				
10	23.227	10				
11	21.672	11				
12	21.87	12				
13	21.439	13				
14	21.089	14				
15	23.709	15				
16	21.669	16				
17	21.752	17				
18	20.761	18				
19	23.479	19				

Seasonal Regression Models - Births





Data Editor

File Edit Help

	births	time	seasonal	var4	var5
1	26.663	1	1		
2	23.598	2	2		
3	26.931	3	3		
4	24.74	4	4		
5	25.806	5	5		
6	24.364	6	6		
7	24.477	7	7		
8	23.901	8	8		
9	23.175	9	9		
10	23.227	10	10		
11	21.672	11	11		
12	21.87	12	12		
13	21.439	13	1		
14	21.089	14	2		
15	23.709	15	3		
16	21.669	16	4		
17	21.752	17	5		
18	20.761	18	6		
19	23.479	19	7		

Another Crude Way of Incorporating Seasonality

- Take the trend prediction and actual prediction.
- Depending on additive or multiplicative model compute the deviation and map it as seasonality effect for each prediction.
- Take averages of the seasonality value. Use this to make future predictions.

Case

Year	Quarter	Time variable (this is created)	Revenues (in $M)
2008	I	1	10.2
	II	2	12.4
	III	3	14.8
	IV	4	15
2009	I	5	11.2
	II	6	14.3
	III	7	18.4
	IV	8	18

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5595 -0.9384  0.4405  1.3265  1.9286

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.0393     1.5531   6.464  0.00065 ***
x             0.9440     0.3076   3.069  0.02196 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.993 on 6 degrees of freedom
Multiple R-squared: 0.6109,	Adjusted R-squared: 0.5461
F-statistic: 9.422 on 1 and 6 DF,  p-value: 0.02196
```

What is the Regression equation?

$$y = 10.0393 + 0.9440x$$

Seasonality: Multiplicative

Time	Observed values TSI* (assuming no impact of cyclicality)	Predicted values (per the regression) T*	SI* = TSI/T
1	10.2	10.983	0.929
2	12.4	11.927	1.040
3	14.8	12.871	1.150
4	15.0	13.815	1.086
5	11.2	14.759	0.759
6	14.3	15.703	0.911
7	18.4	16.647	1.105
8	18.0	17.591	1.023

* T: Trend; S: Seasonal; I: Irregular

Quarterly Seasonality

Time	Average seasonality factor
Q1	$0.844 \left(= \frac{0.929+0.759}{2}\right)$
Q2	0.975
Q3	1.127
Q4	1.054

Time	Observed values TSI* (assuming no impact of cyclicality)	Predicted values (per (per the regression)	SI* = TSI/T
1	10.2	10.983	0.929
2	12.4	11.927	1.040
3	14.8	12.871	1.150
4	15.0	13.815	1.086
5	11.2	14.759	0.759
6	14.3	15.703	0.911
7	18.4	16.647	1.105
8	18.0	17.591	1.023

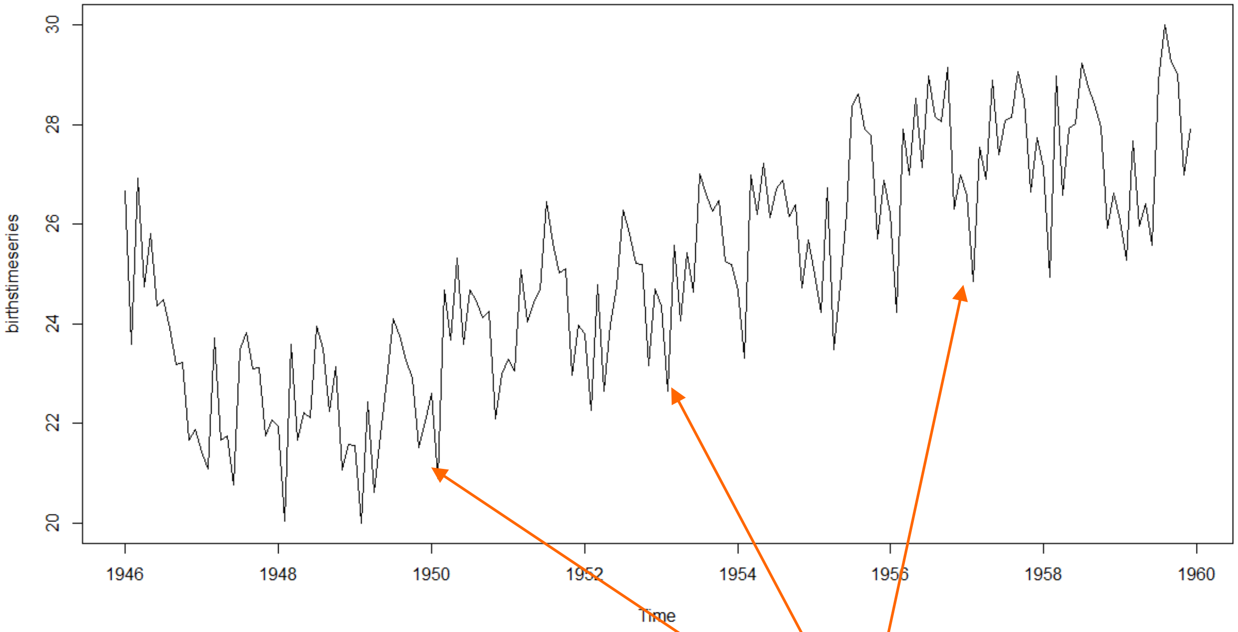
Computations

- Trend Y_9 =10.039 + 0.944(9) = 18.535

- Corrected for seasonality and randomness: 18.535 * 0.844 = 15.643

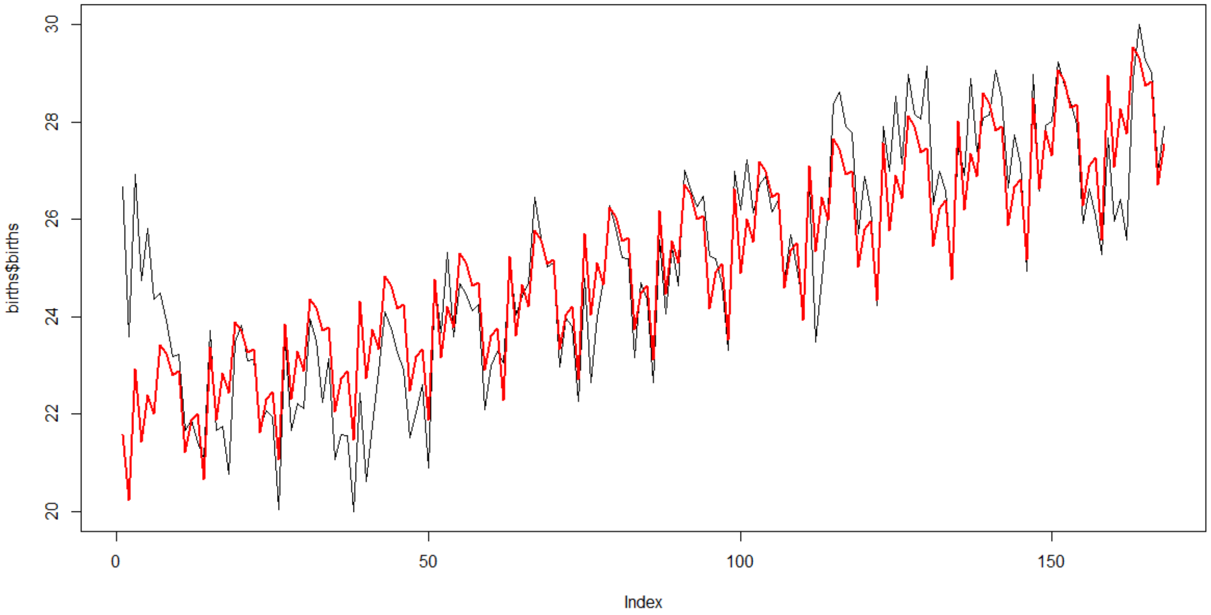
CSE 7202c

Births in NY




```
> birthstimeseries <- ts(births,
+                        frequency=12,
+                        start=c(1946,1))
> birthstimeseries
        Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct    Nov    Dec
1946 26.663 23.598 26.931 24.740 25.806 24.364 24.477 23.901 23.175 23.227 21.672 21.870
1947 21.439 21.089 23.709 21.669 21.752 20.761 23.479 23.824 23.105 23.110 21.759 22.073
1948 21.937 20.035 23.590 21.672 22.222 22.123 23.950 23.504 22.238 23.142 21.059 21.573
1949 21.548 20.000 22.424 20.615 21.761 22.874 24.104 23.748 23.262 22.907 21.519 22.025
1950 22.604 20.894 24.677 23.673 25.320 23.583 24.671 24.454 24.122 24.252 22.084 22.991
1951 23.287 23.049 25.076 24.037 24.430 24.667 26.451 25.618 25.014 25.110 22.964 23.981
1952 23.798 22.270 24.775 22.646 23.988 24.737 26.276 25.816 25.210 25.199 23.162 24.707
1953 24.364 22.644 25.565 24.062 25.431 24.635 27.009 26.606 26.268 26.462 25.246 25.180
1954 24.657 23.304 26.982 26.199 27.210 26.122 26.706 26.878 26.152 26.379 24.712 25.688
1955 24.990 24.239 26.721 23.475 24.767 26.219 28.361 28.599 27.914 27.784 25.693 26.881
1956 26.217 24.218 27.914 26.975 28.527 27.139 28.982 28.169 28.056 29.136 26.291 26.987
1957 26.589 24.848 27.543 26.896 28.878 27.390 28.065 28.141 29.048 28.484 26.634 27.735
1958 27.132 24.924 28.963 26.589 27.931 28.009 29.229 28.759 28.405 27.945 25.912 26.619
1959 26.076 25.286 27.660 25.951 26.398 25.565 28.865 30.000 29.261 29.012 26.992 27.897
>
```

Seasonality: Multiplicative

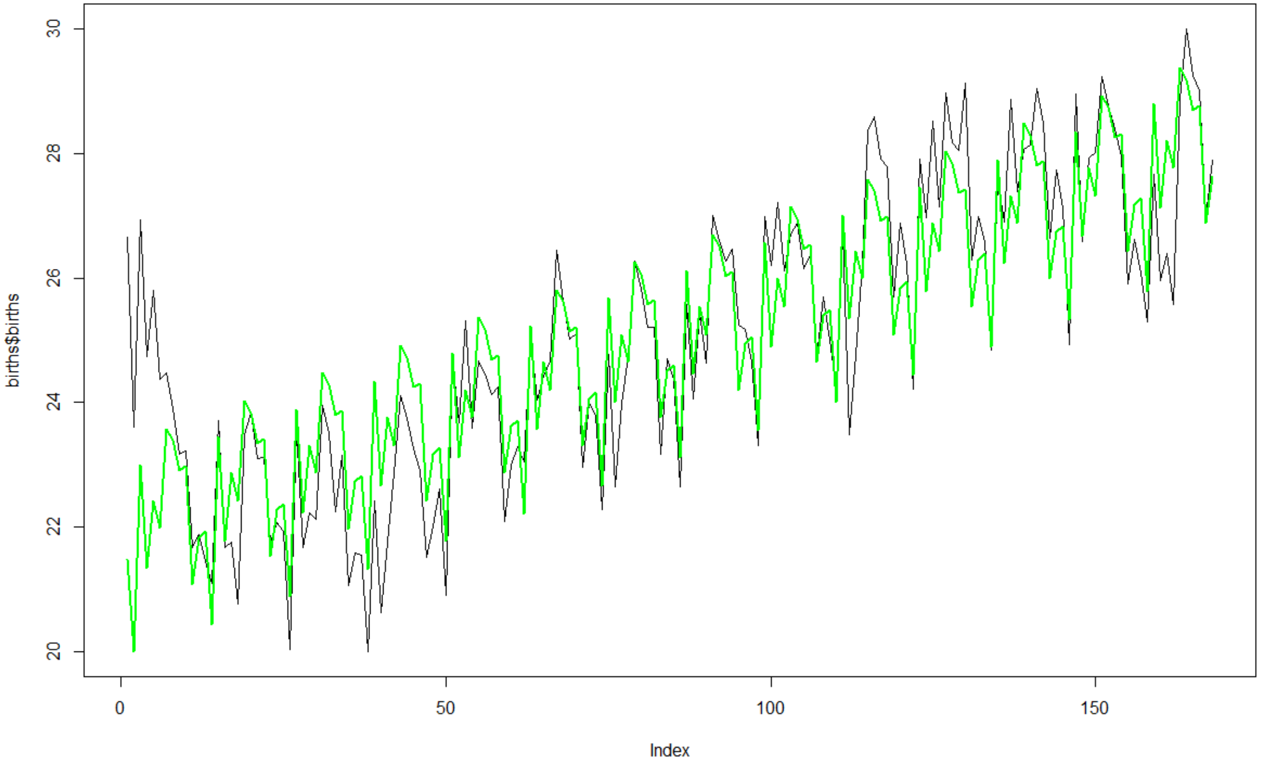


Data Editor

File Edit Help

	births	time	seasonal	mae	var5
1	26.663	1	1	1.214304	
2	23.598	2	2	1.072901	
3	26.931	3	3	1.222374	
4	24.74	4	4	1.121036	
5	25.806	5	5	1.167374	
6	24.364	6	6	1.100294	
7	24.477	7	7	1.103546	
8	23.901	8	8	1.075775	
9	23.175	9	9	1.041357	
10	23.227	10	10	1.041954	
11	21.672	11	11	0.9705802	
12	21.87	12	12	0.9778208	
13	21.439	13	1	0.9569611	
14	21.089	14	2	0.93978	
15	23.709	15	3	1.054788	
16	21.669	16	4	0.9624398	
17	21.752	17	5	0.9645349	
18	20.761	18	6	0.9190777	
19	23.479	19	7	1.037695	

Seasonality: Additive

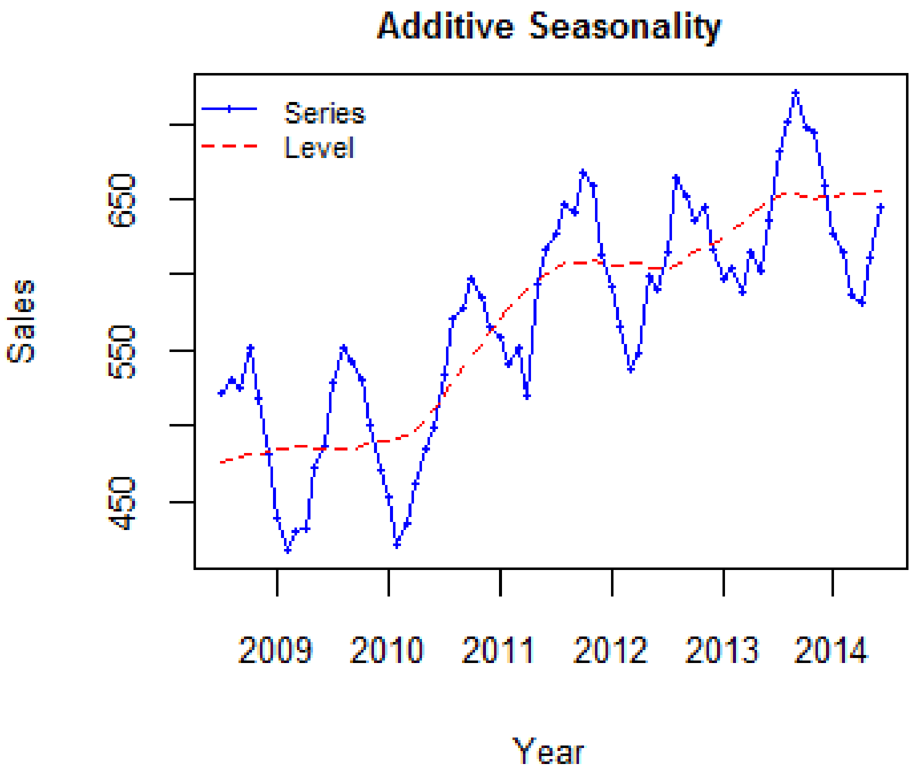


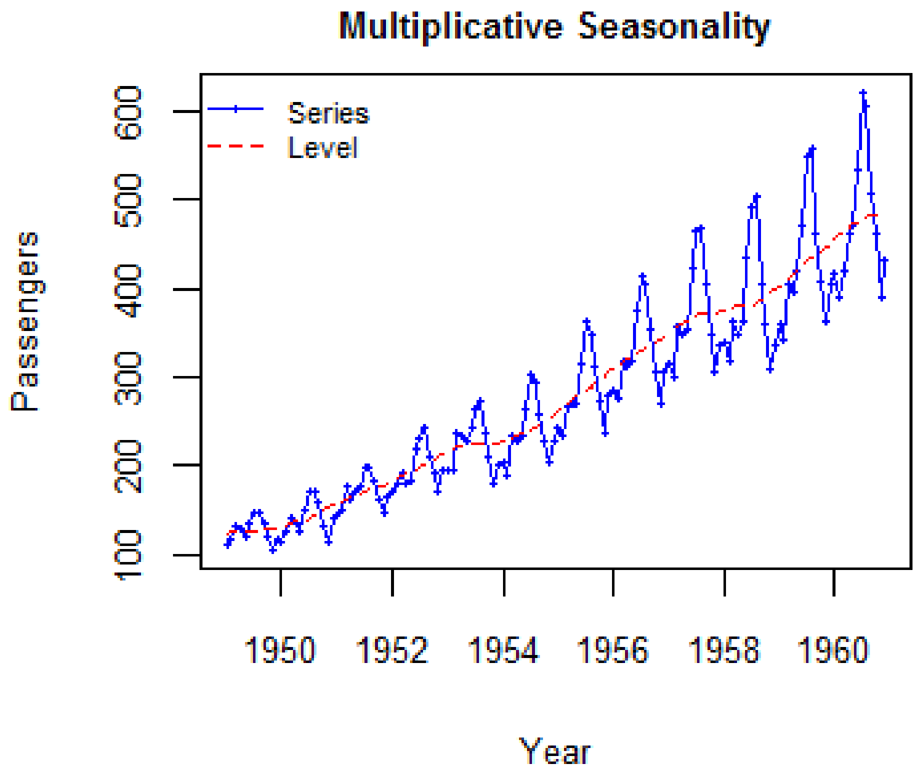
Data Editor

File Edit Help

	births	time	seasonal	mae
1	26.663	1	1	4.70557
2	23.598	2	2	1.603422
3	26.931	3	3	4.899274
4	24.74	4	4	2.671125
5	25.806	5	5	3.699977
6	24.364	6	6	2.220829
7	24.477	7	7	2.29668
8	23.901	8	8	1.683532
9	23.175	9	9	0.920384
10	23.227	10	10	0.9352357
11	21.672	11	11	-0.6569126
12	21.87	12	12	-0.4960608
13	21.439	13	1	-0.9642091
14	21.089	14	2	-1.351357
15	23.709	15	3	1.231494
16	21.669	16	4	-0.8456539
17	21.752	17	5	-0.7998021
18	20.761	18	6	-1.82795
19	23.479	19	7	0.8529014

Additive or Multiplicative





Source: http://www.forsoc.net/2014/11/11/can-you-identify-additive-and-multiplicative-seasonality/

Goodness of Fit

- MSE (Mean square error)
- MAE (Mean absolute error)
- RMSE (Root mean square error)

- MAPE (Mean absolute percent error)

- NMSE (Normalized mean square error)
- NMAE (Normalized mean absolute error)
- NMAPE (Normalized mean absolute percent error)

Issues with Regressing on Time

- It is too much of a curve fit for a statistician to sleep well!
- If there is no trend or if seasonality and fluctuations are more important than trend, then the coefficients behave weirdly

TIME SERIES: AUTO REGRESSIVE METHODS

Auto Regressive Methods

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, y_{t-2} \cdots)$$

Identifying Techniques for Different Processes

- We use different techniques for different processes
  - Random stationary
  - Seasonal
  - Trend
- First we need to identify them

AUTOCORRELATION AND PARTIAL AUTOCORRELATION

Time Series Descriptive Statistics

- In descriptive statistics covered earlier (central tendencies, measures of variability, skewness, kurtosis, distributions, correlations, etc.), the order of observations in the data was of no consequence.

- In time series descriptive statistics, order of observations is of primary importance and so autocorrelations, etc. play a vital role in identifying the models and their characteristics.

Autocorrelation (ACF) and Partial ACF (PACF)

- ACF: n^{th} lag of ACF is the correlation between a day and n days before that.

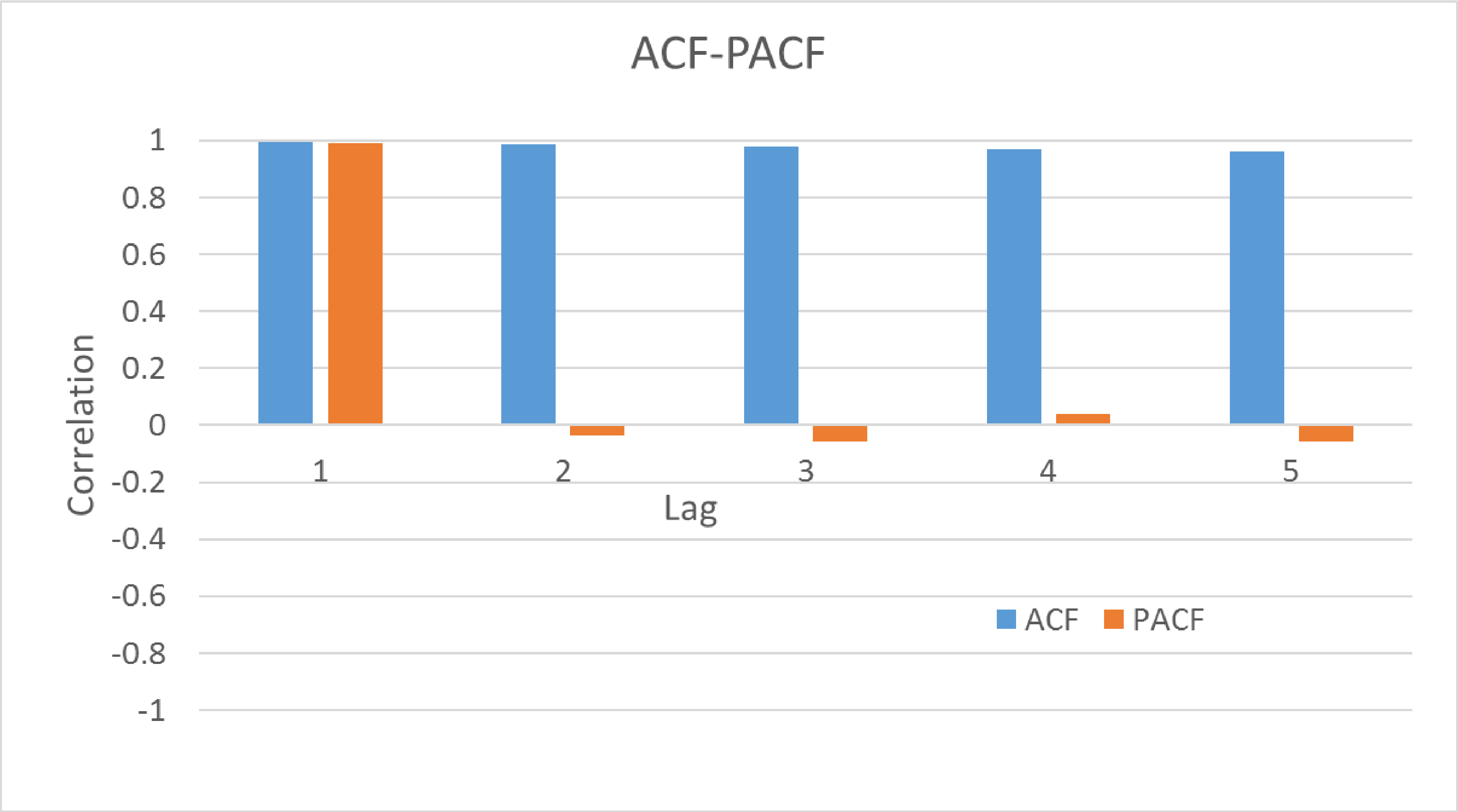
- PACF: The same as ACF with all intermediate correlations removed. It is the k_{th} coefficient of the ordinary least squares regression.

$$[y_t] = \beta_0 + \sum_{i=1}^{k} \beta_i [y_{t-i}] \text{ where}$$

$[y_t]$ is the input time series, k is the lag order and β_i is the i_{th} coefficient of the linear multiple regression.

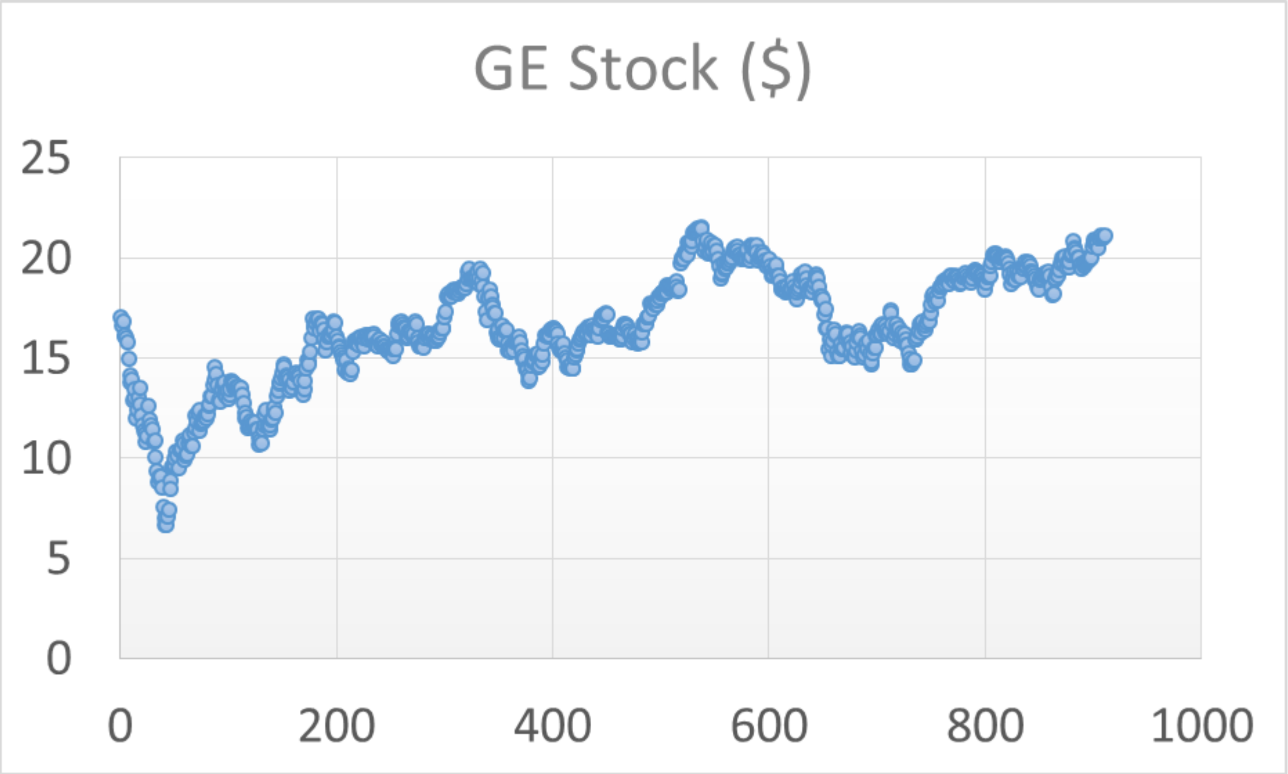
EXCEL ACTIVITY

Autocorrelation (ACF) and Partial ACF (PACF)

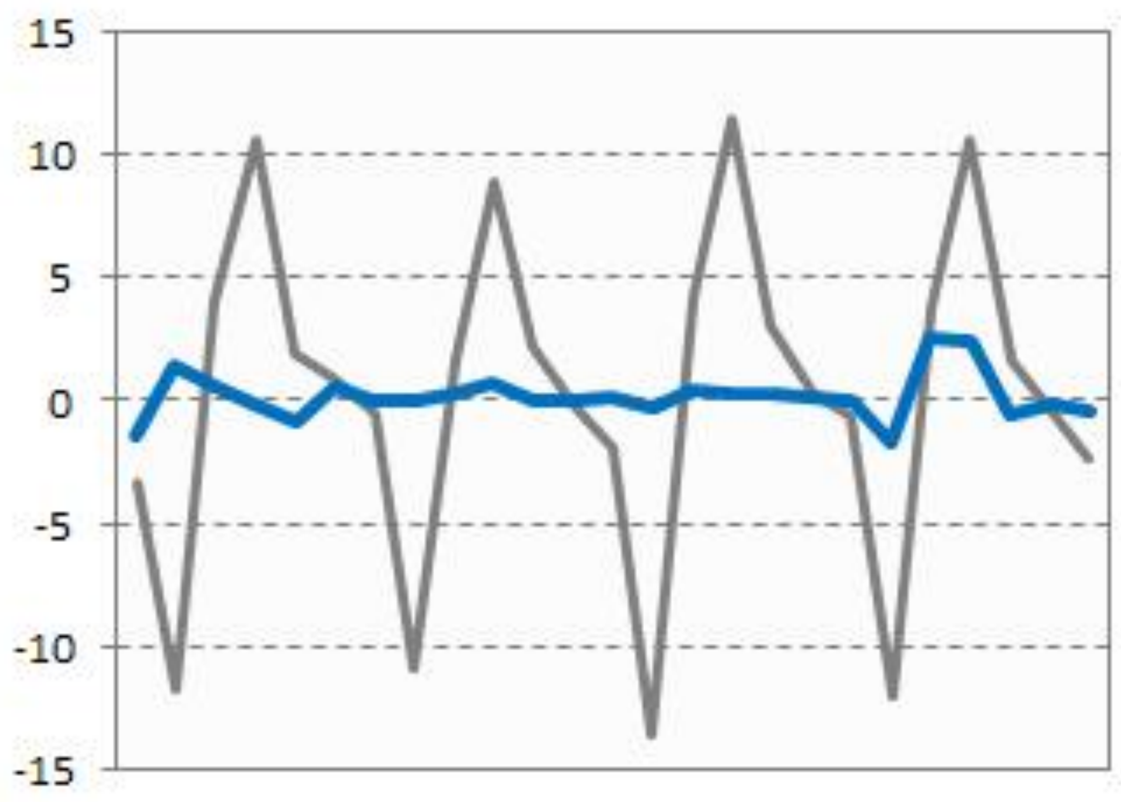


Components of Time Series

- **Trend**
- Seasonality
- Random component

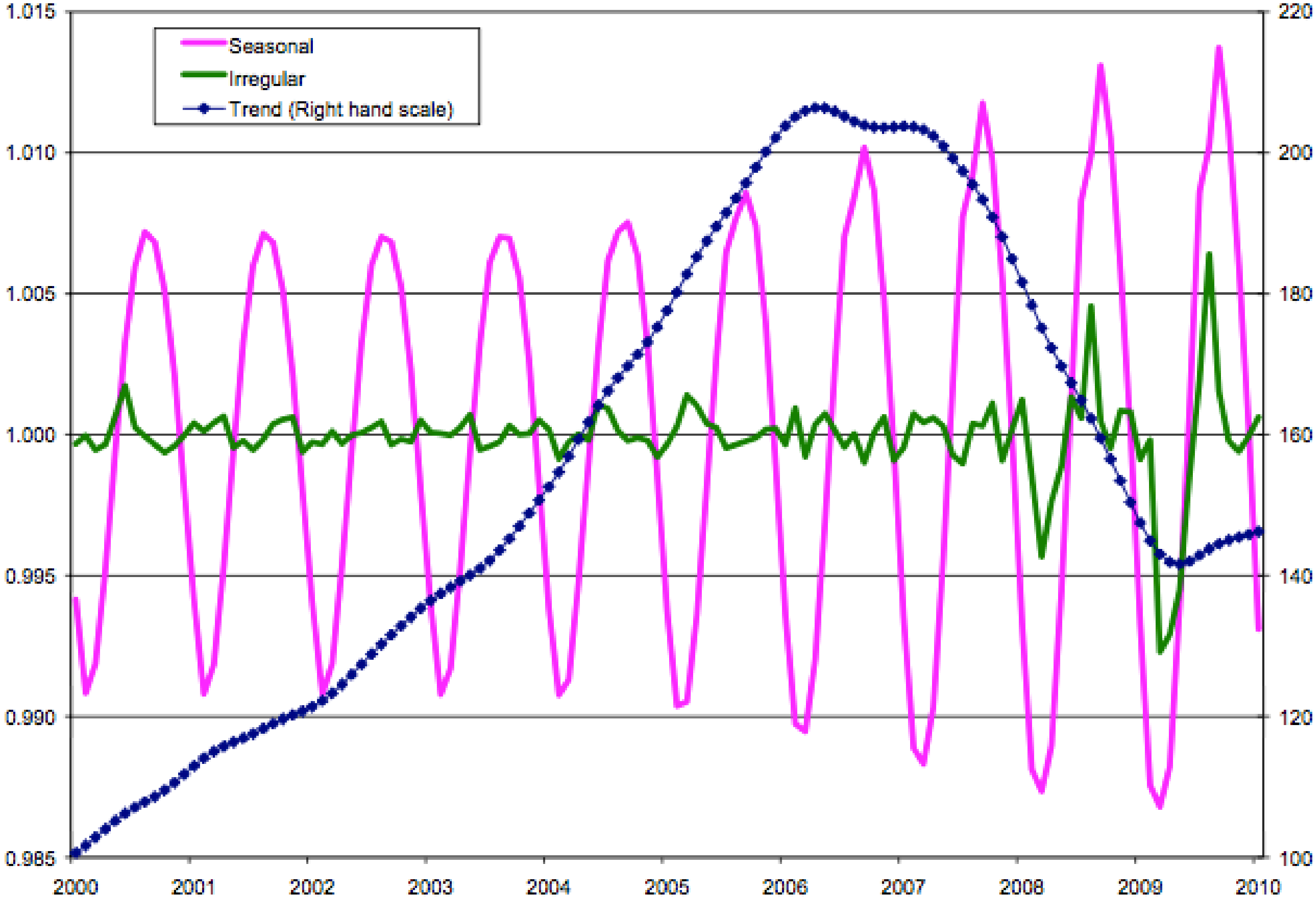


Seasonality



CSE 7202c

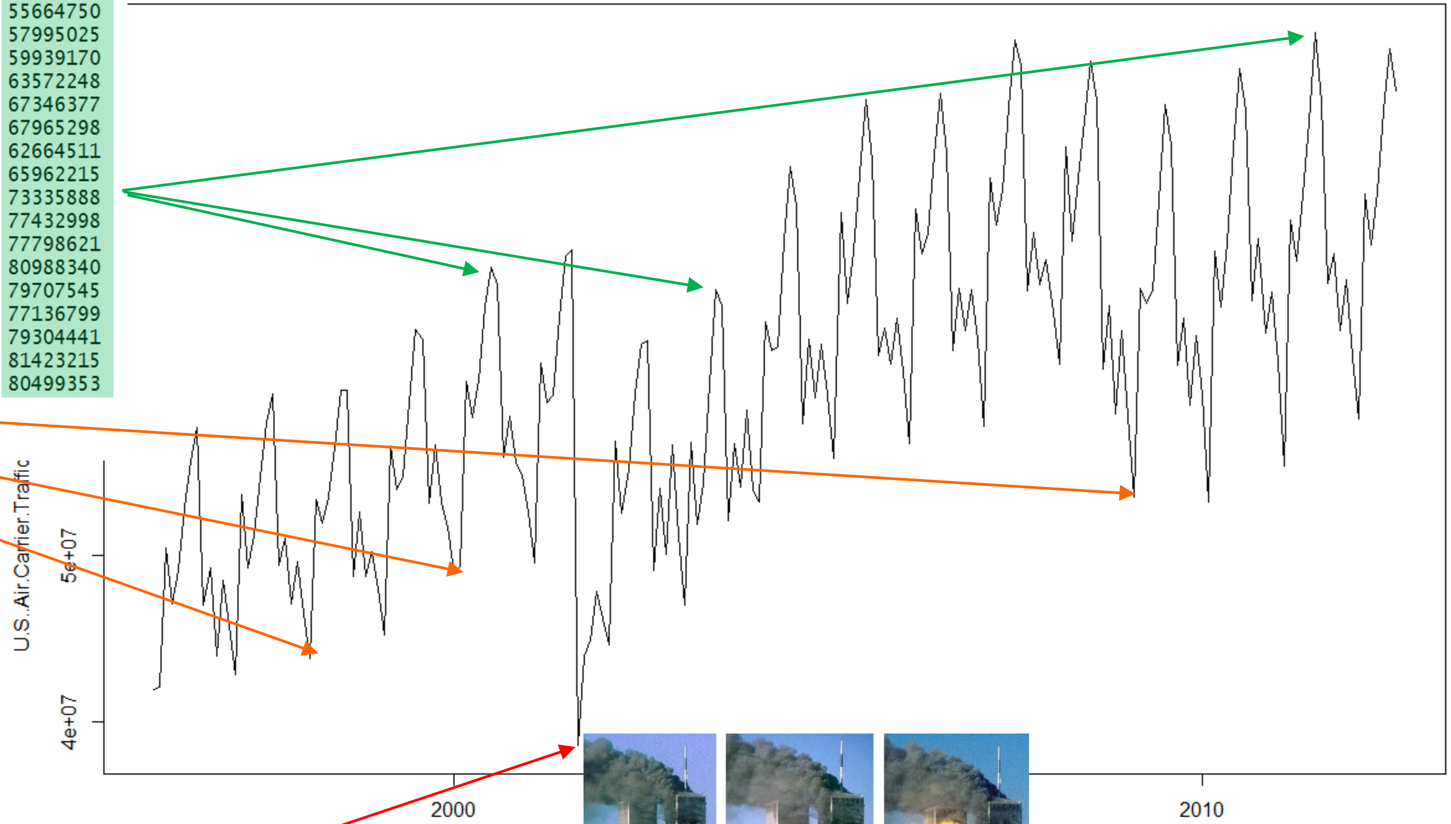
Trend, Seasonality and Randomness



CSE 7202c

US Air Carrier Traffic – Revenue Passenger Miles (‘000) *RPM*

```
> milestimeseries <- ts(miles, frequency = 12, start = c(1996,1))
> milestimeseries
          Jan      Feb      Mar      Apr      May      Jun      Jul
1996 41972194 42054796 50443045 47112397 49118248 52880510 55664750
1997 45850623 42838949 53620994 49282817 51191842 54707221 57995025
1998 46514139 43769273 53361926 51968480 53515798 56460422 59939170
1999 47988560 45241211 56555731 53920855 54674958 59213000 63572248
2000 49045412 49306303 60443541 58286680 60533783 64903295 67346377
2001 52634354 49532578 61575055 59151645 59662416 64353323 67965298
2002 46224031 44615129 56897729 52542164 55116060 59745343 62664511
2003 51197175 47040806 56766580 51857453 54335598 60272900 65962215
2004 53979786 53179693 64035864 62340117 62530704 68866398 73335888
2005 59629608 55795165 70595861 65145552 68268899 72952959 77432998
2006 61035027 56729212 70799794 68120559 69352606 74099239 77798621
2007 63016013 57793832 72700241 69836156 71933109 76926452 80988340
2008 64667106 61504426 74575531 68906882 72725750 76162105 79707545
2009 58373786 53506580 66027341 65166300 65868254 71350227 77136799
2010 59651061 53240066 68307090 64953250 68850904 74474550 79304441
2011 61630362 55391206 70158268 67683558 71711448 76057910 81423215
2012 61940180 58243763 71696039 68669228 71887523 76760759 80499353
          Aug      Sep      Oct      Nov      Dec
1996 57723208 47035464 49263120 43937074 48539606
1997 59715433 49418190 51058879 47056048 49654209
```

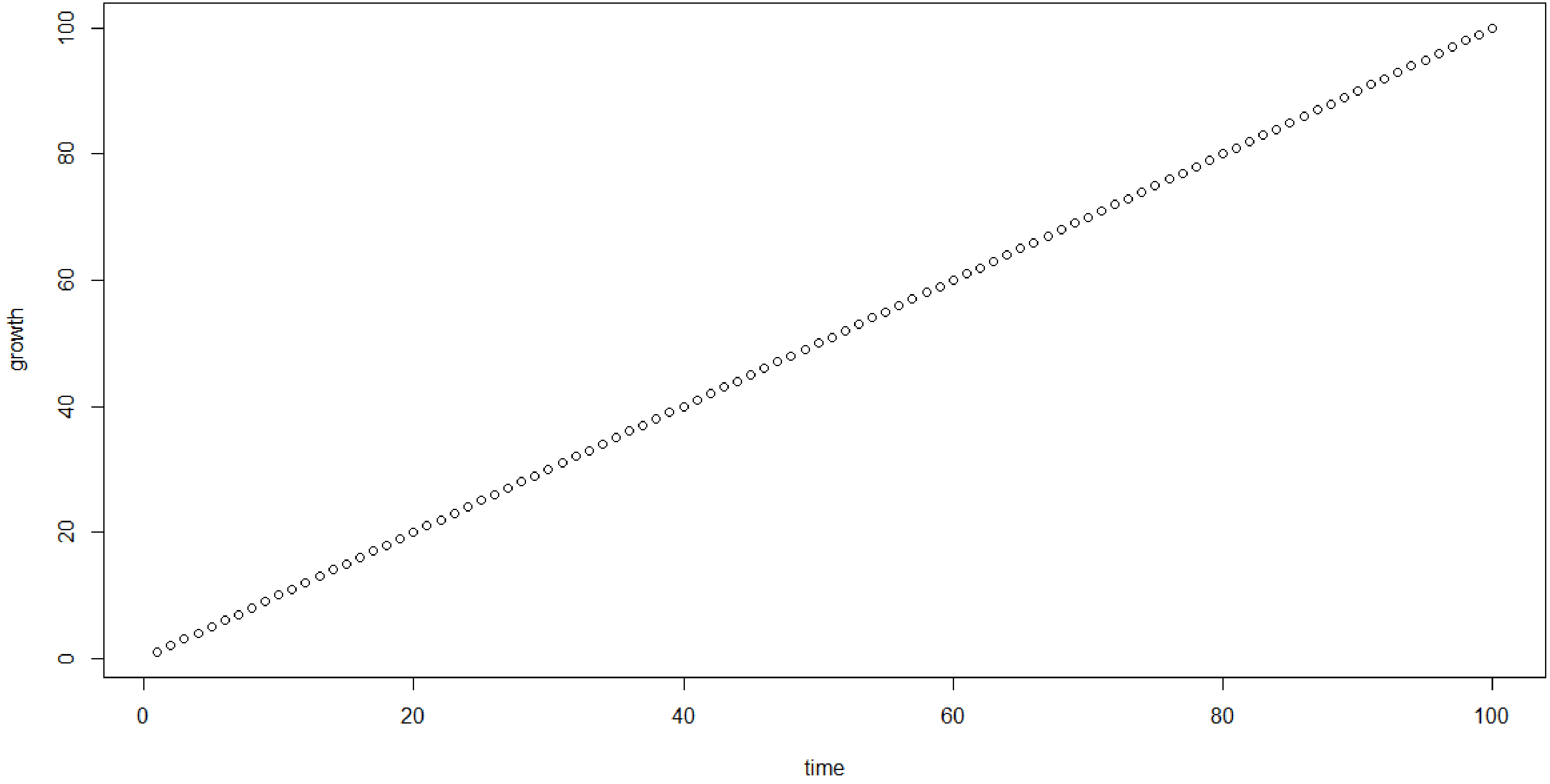


Data sources:
http://www.bts.gov/xml/air_traffic/src/index.xml
and https://datamarket.com/data/set/281x/us-air-carrier-traffic-statistics-revenue-passenger-miles
Last accessed: 31-Mar-2016

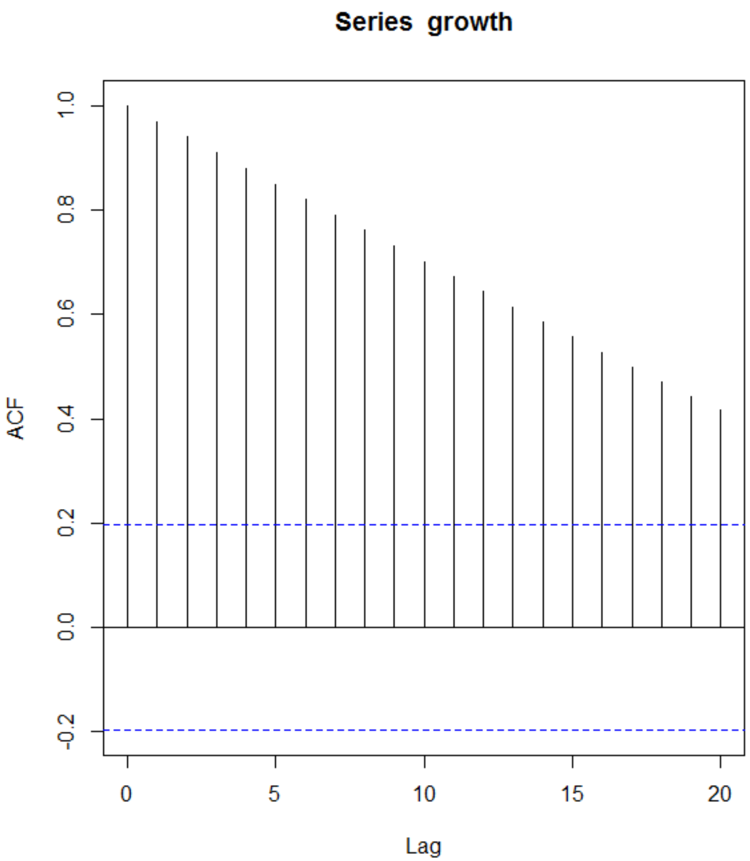
```
          Aug      Sep      Oct      Nov      Dec
1996 57723208 47035464 49263120 43937074 48539606
1997 59715433 49418190 51058879 47056048 49654209
1998 59927214 48751280 52578217 48734375 50208641
1999 63003663 53131972 56653901 53215500 51746821
2000 66256804 55900504 58373996 55590325 54822970
2001 68377080 38601868 43964788 44915764 47836501
2002 62944816 49096035 54019748 50106814 56656594
2003 64989766 52121480 56724551 54128776 58739845
2004 70961522 57881042 63021142 59453943 62680310
```

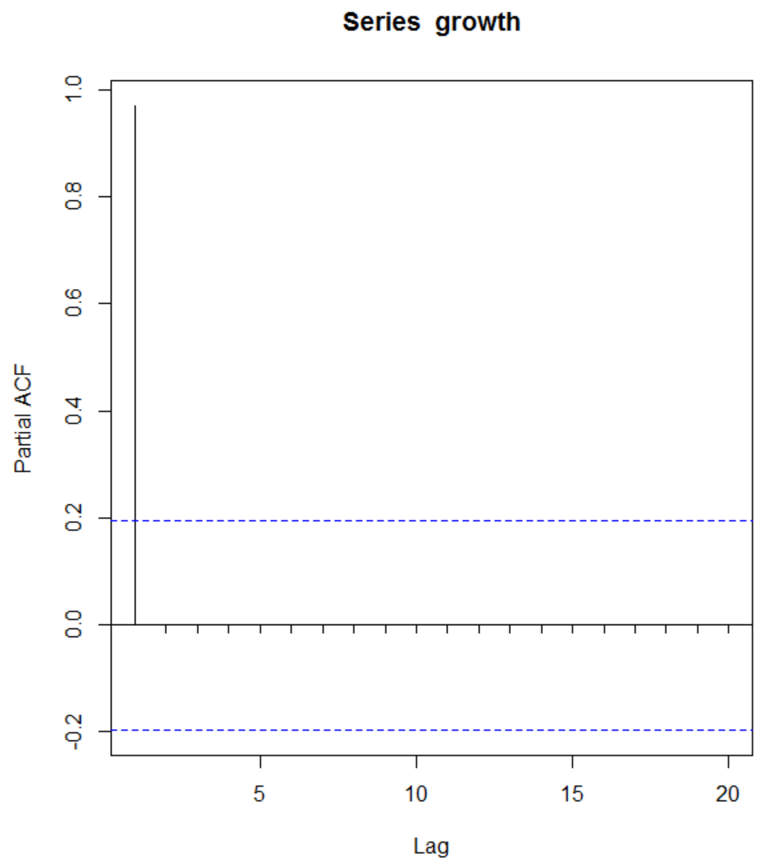
ACF and PACF – Idealized Trend, Seasonality and Randomness

ACF and PACF – Idealized Trend



ACF and PACF – Idealized Trend

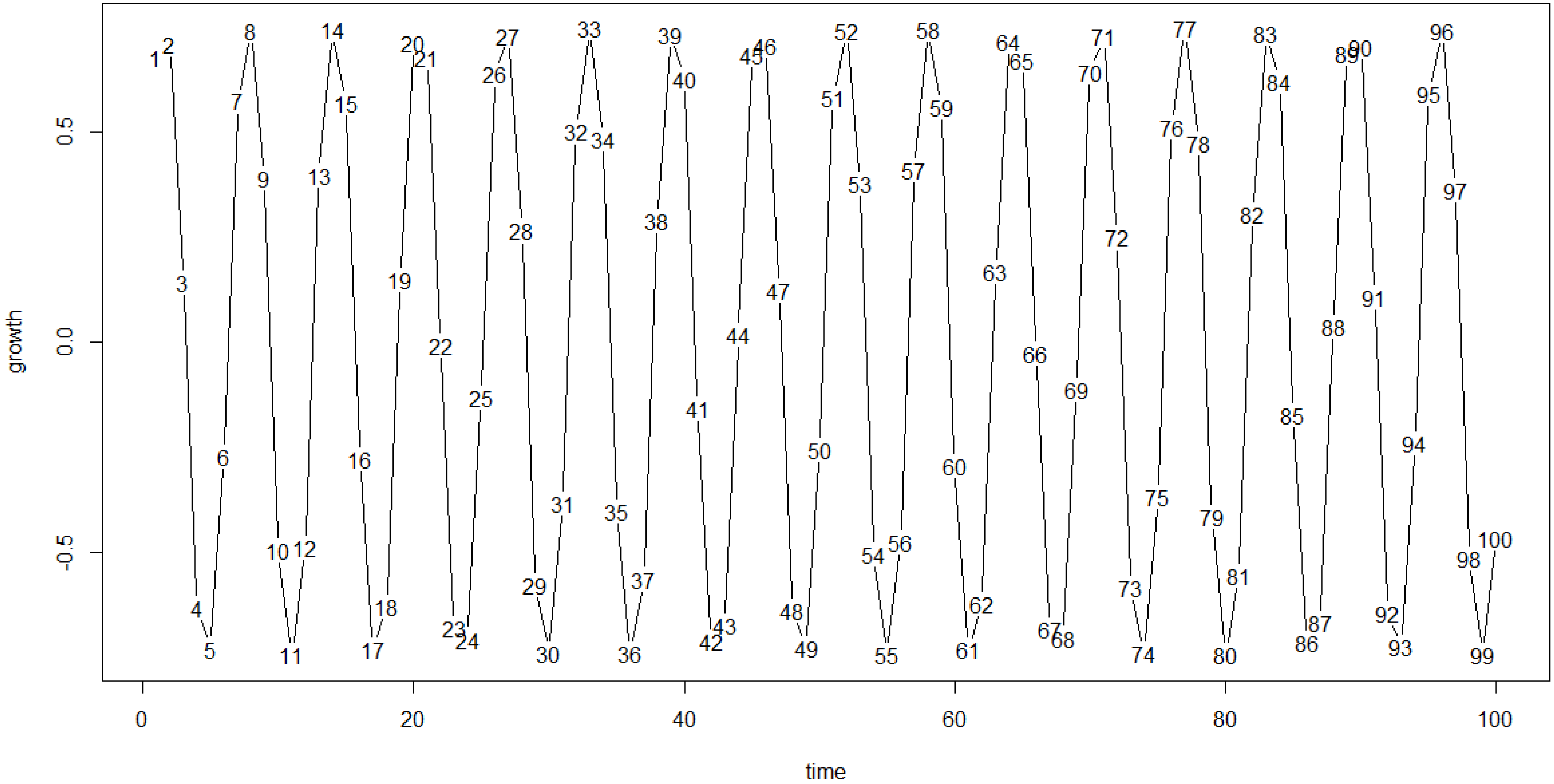




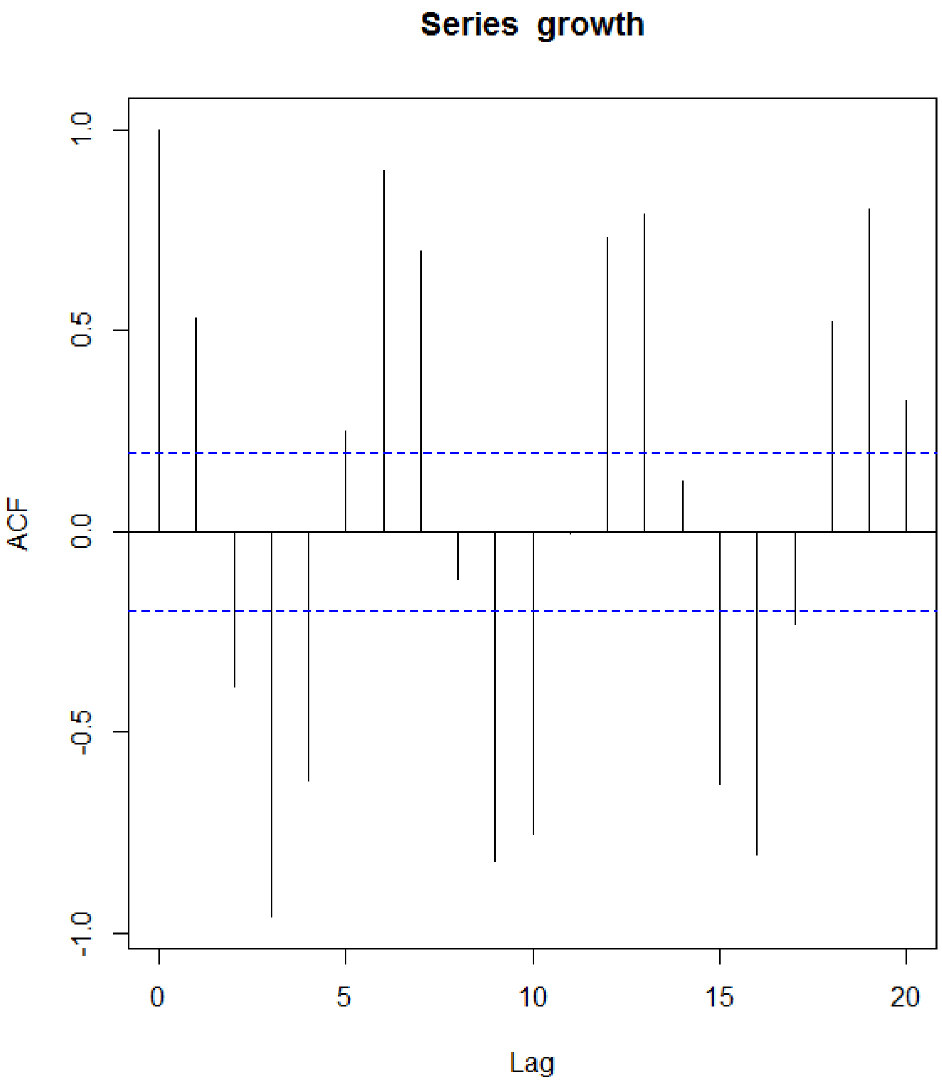
95% CI: $0 \pm \frac{1.96}{\sqrt{n}}$

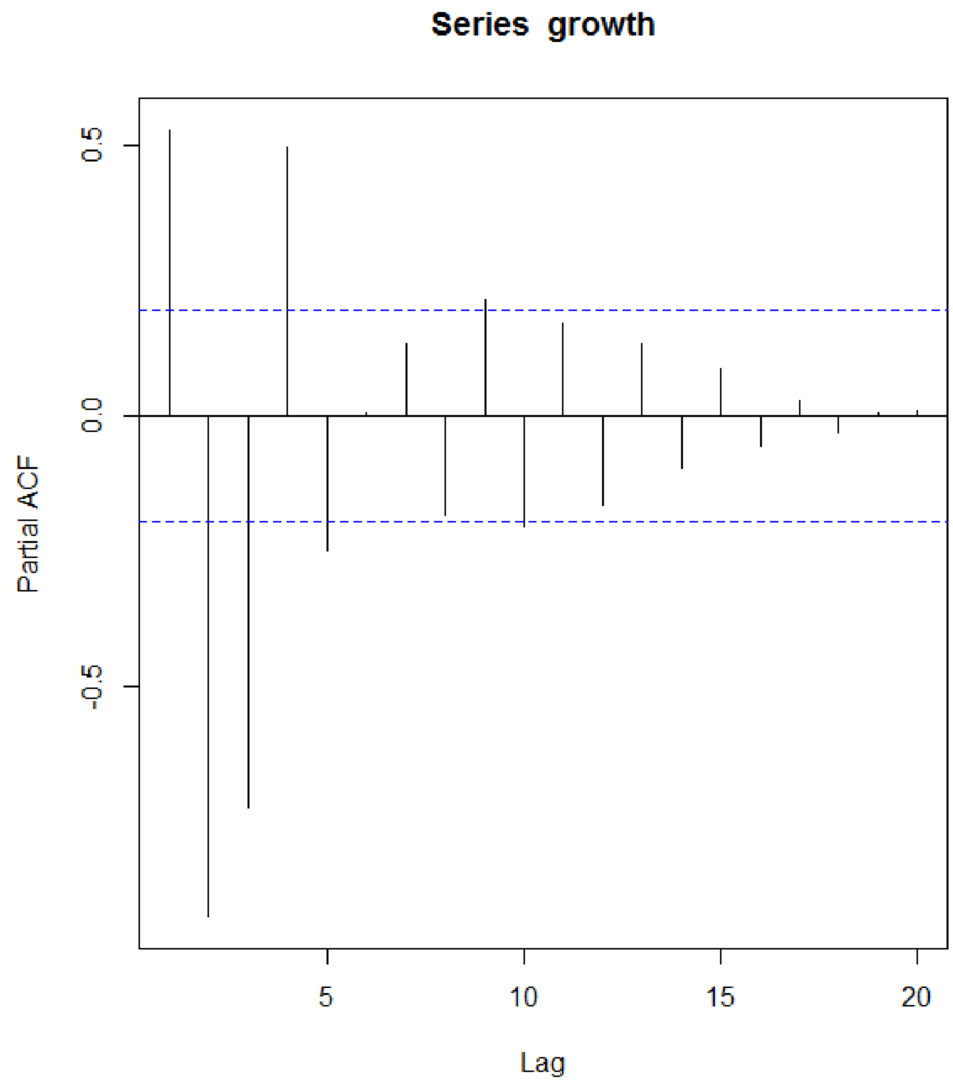
- ACF is a bar chart of correlation coefficients of the time series and its lags.
- PACF is a plot of the partial correlation coefficients of the time series and its lags.

ACF and PACF – Idealized Seasonality

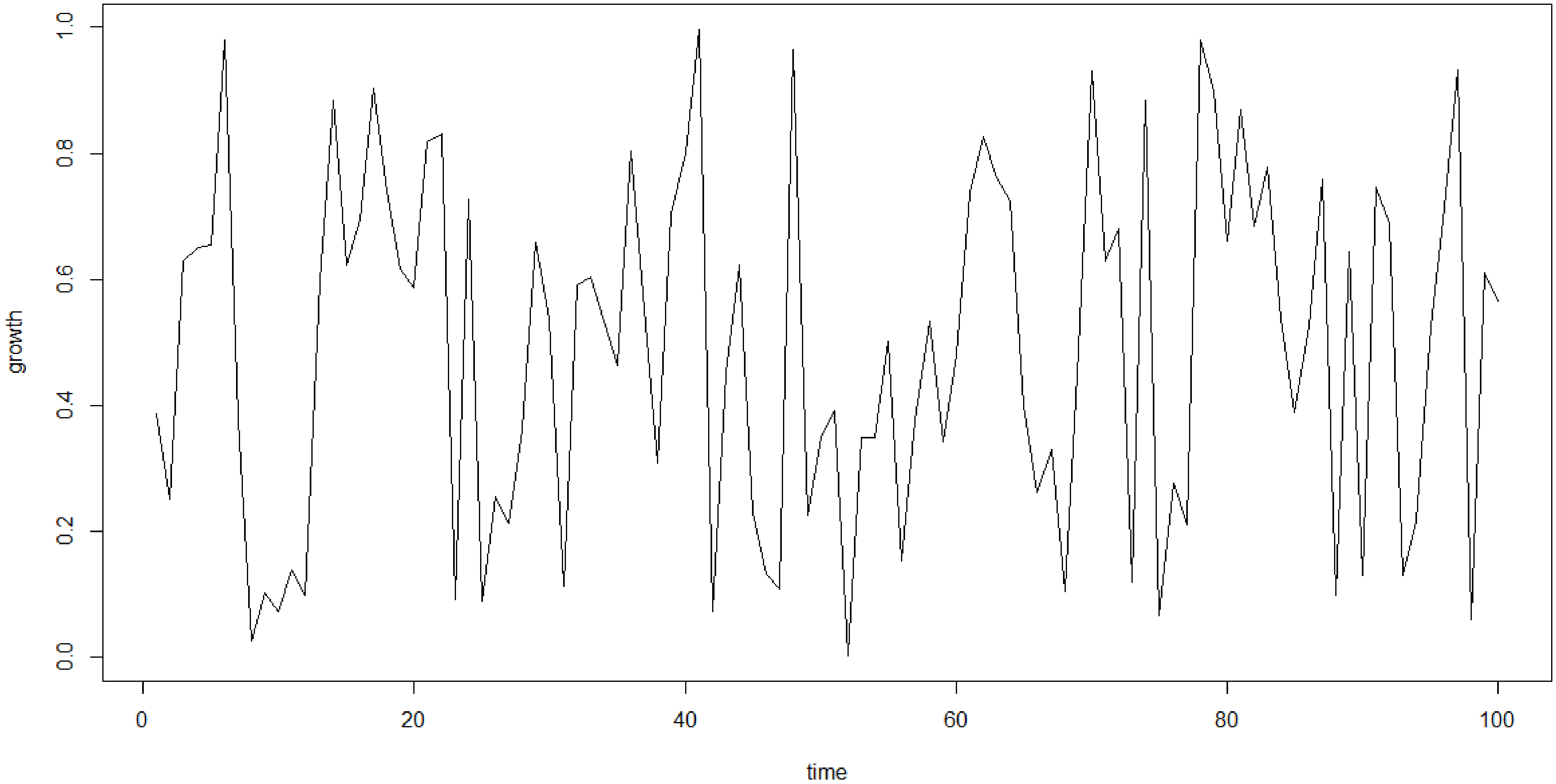


ACF and PACF – Idealized Seasonality

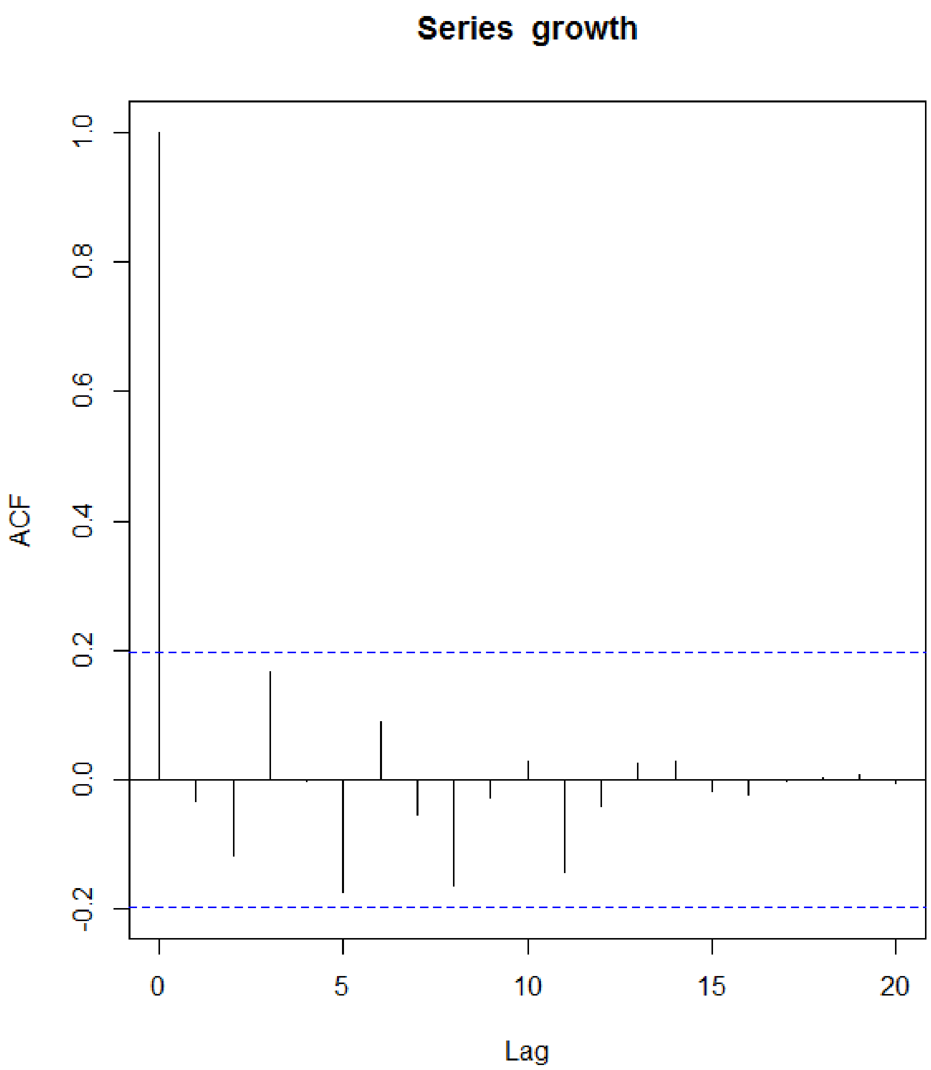


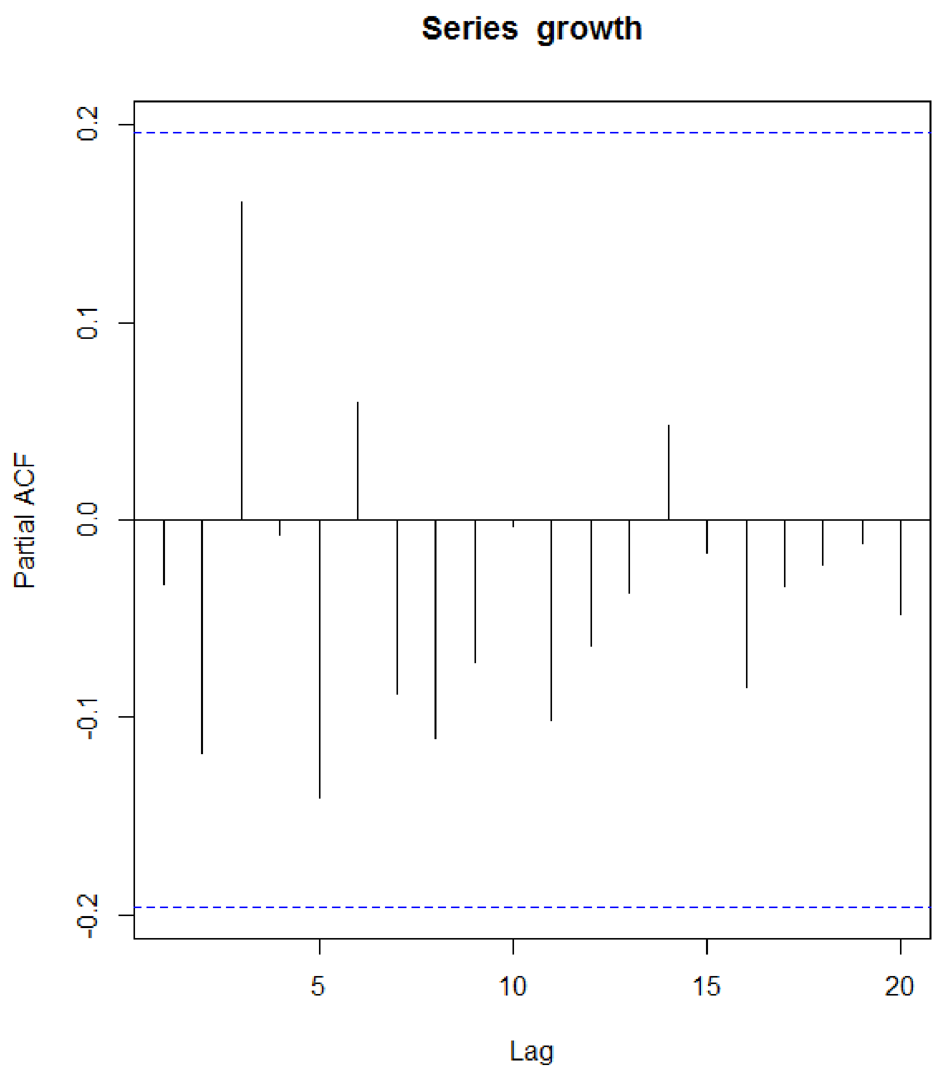


ACF and PACF – Idealized Randomness



ACF and PACF – Idealized Randomness



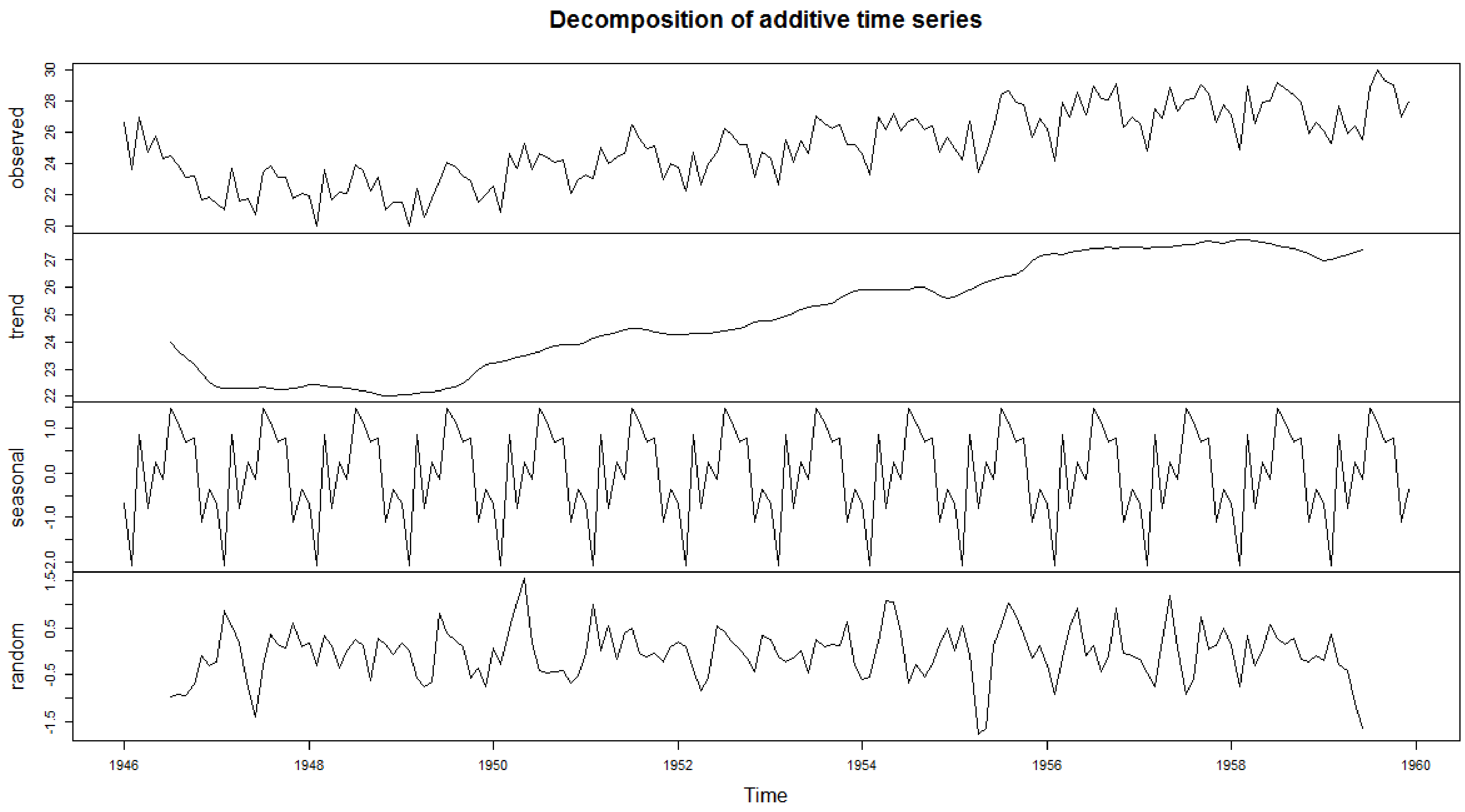


ACF and PACF – Idealized Trend, Seasonality and Randomness

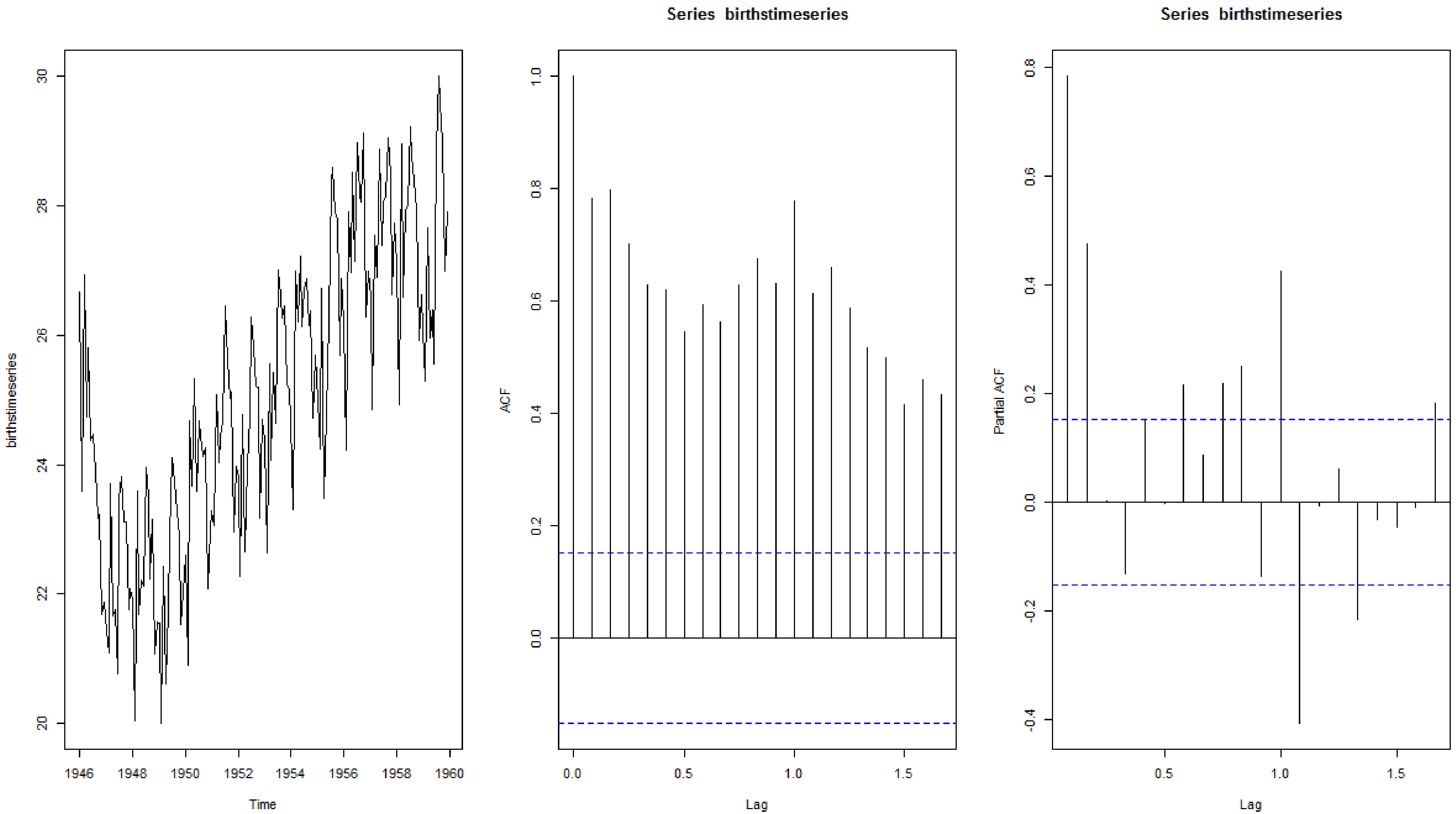
- Ideal Trend: Decreasing ACF and 1 or 2 lags of PACF
- Ideal Seasonality: Cyclicality in ACF and a few lags of PACF with some positive and some negative
- Ideal Random: A spike may or may not be present; even if present, magnitude will be small

ACF and PACF (Real-world): Decomposing Time Series into the 3 Components

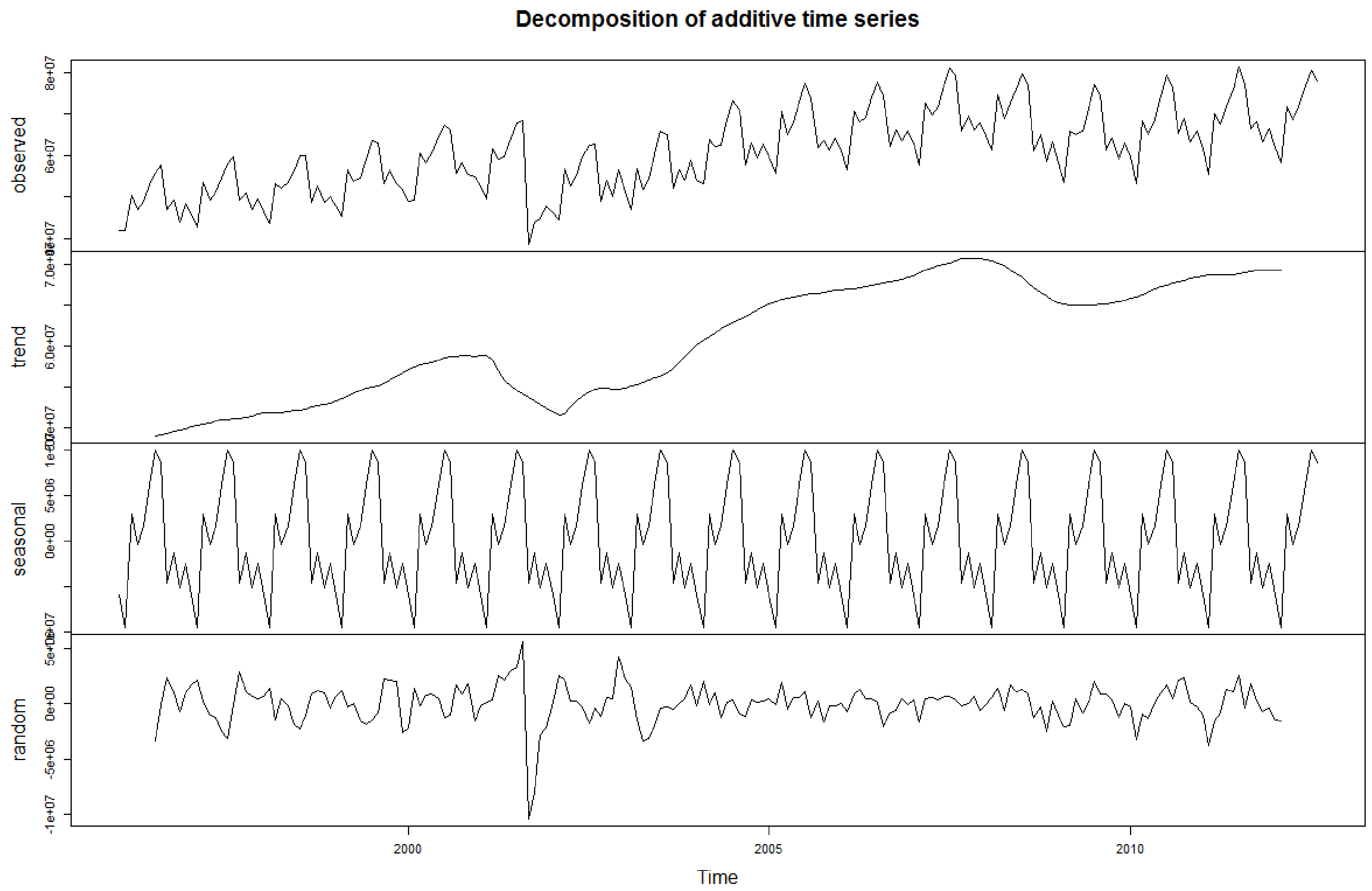
ACF and PACF (Real-world): Decomposing Time Series into the 3 Components – Births in NY



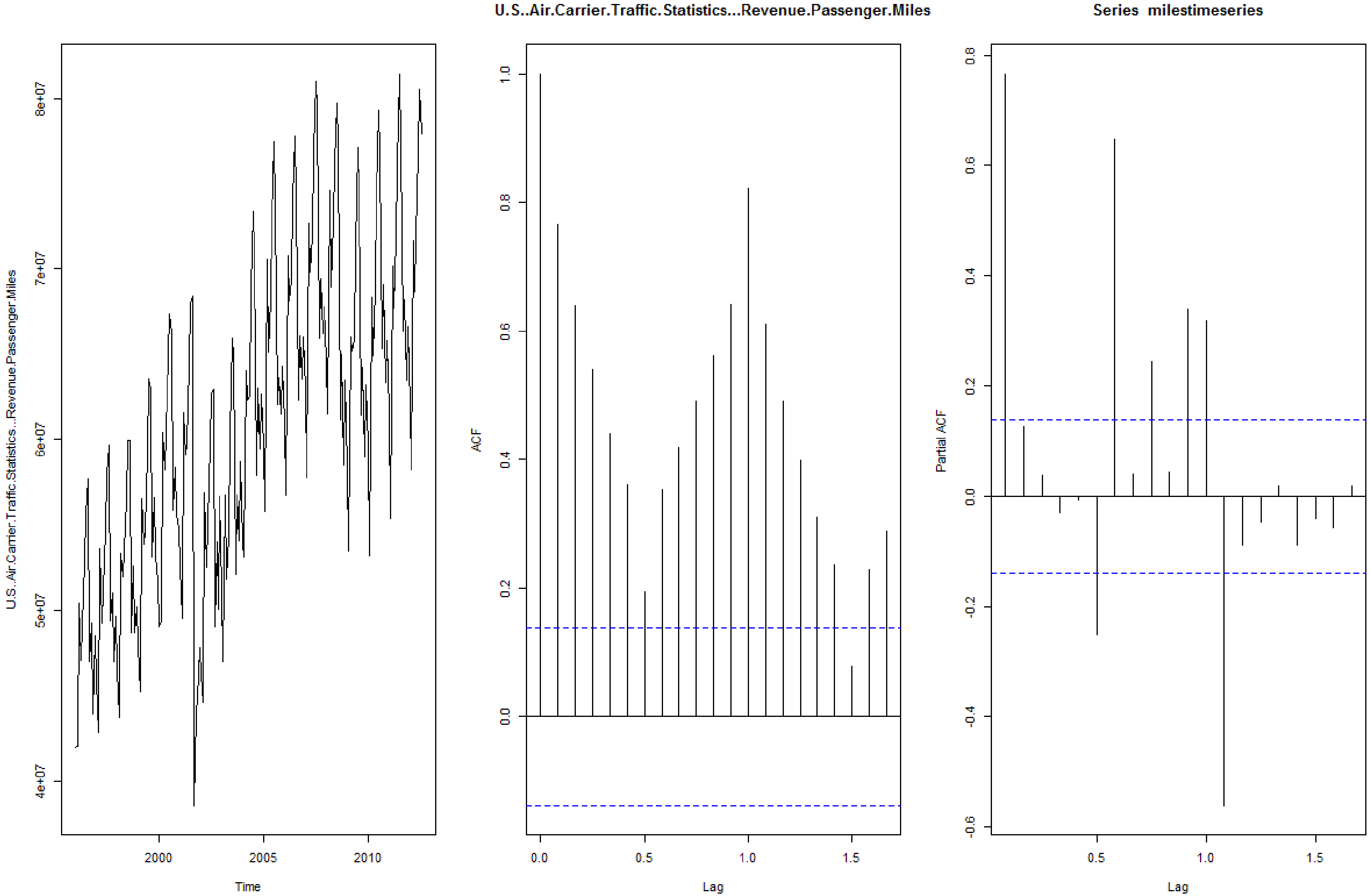
ACF and PACF (Real-world): Decomposing Time Series into the 3 Components – Births in NY



ACF and PACF (Real-world): Decomposing Time Series into the 3 Components – Revenue Passenger Miles (RPM)



ACF and PACF (Real-world): Decomposing Time Series into the 3 Components – RPM



Stationary and Non-Stationary

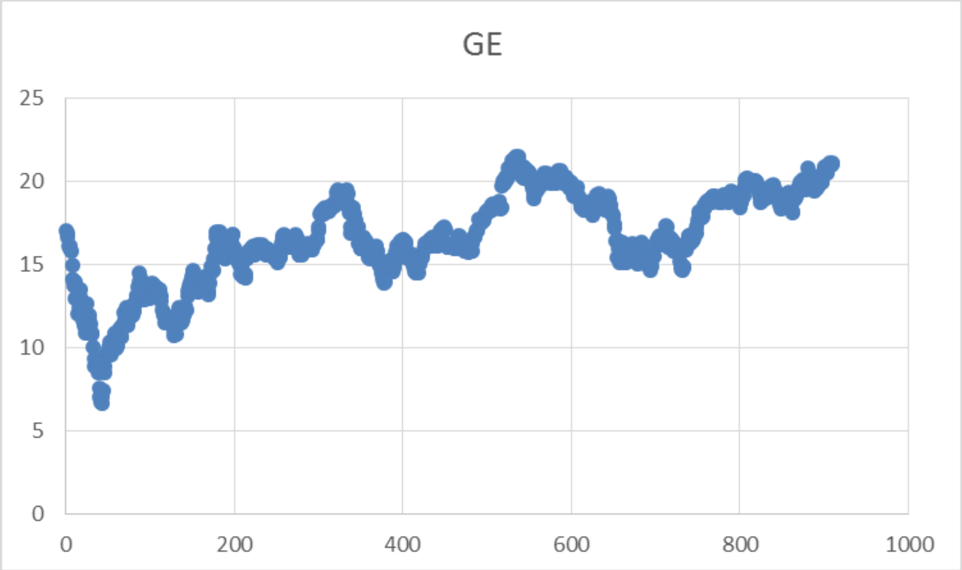
- Stationary data has constant statistical properties – mean, variance, autocorrelation, etc. – over time

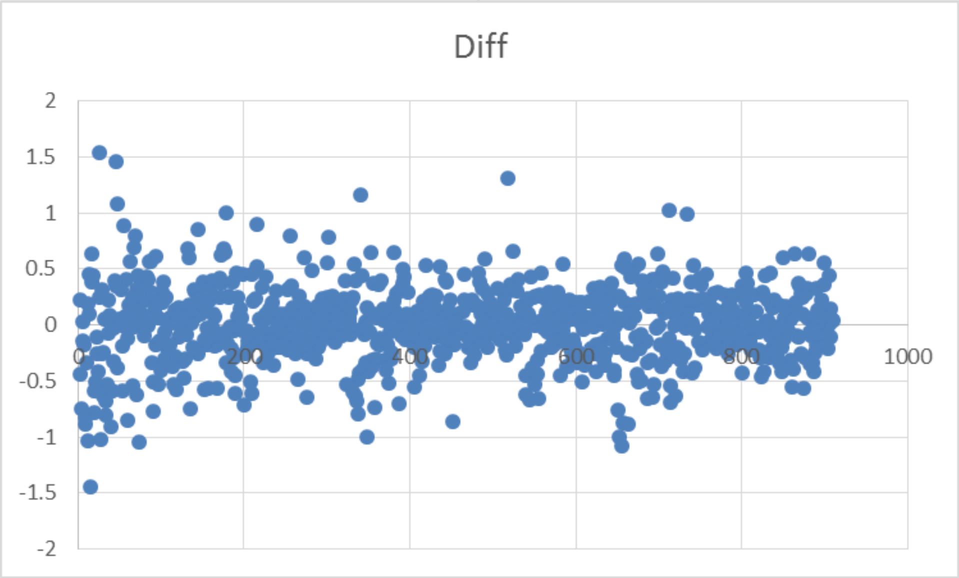
- If the data is stationary, forecasting is easier!

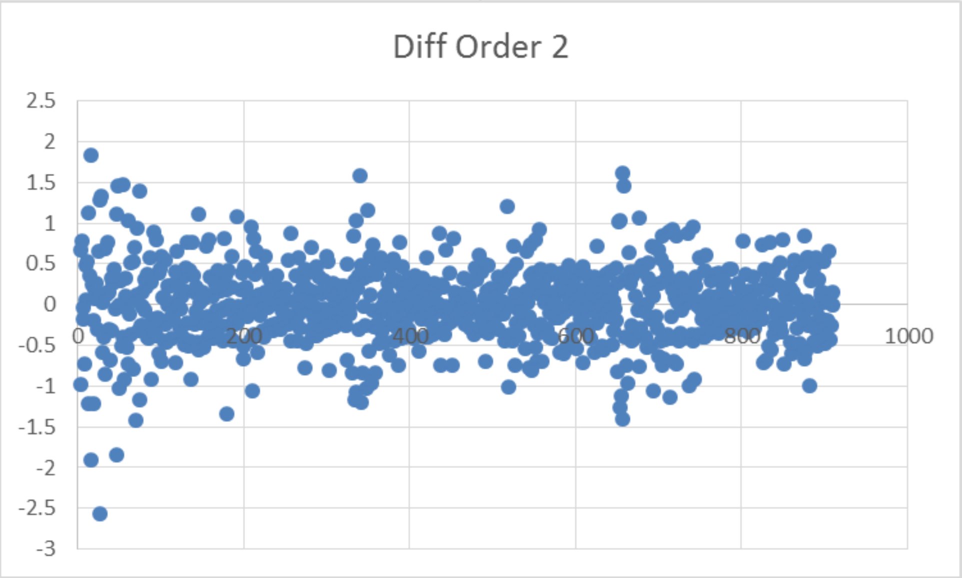
- Differencing to convert non-stationary to stationary

EXCEL ACTIVITY

Removing Trend from Data

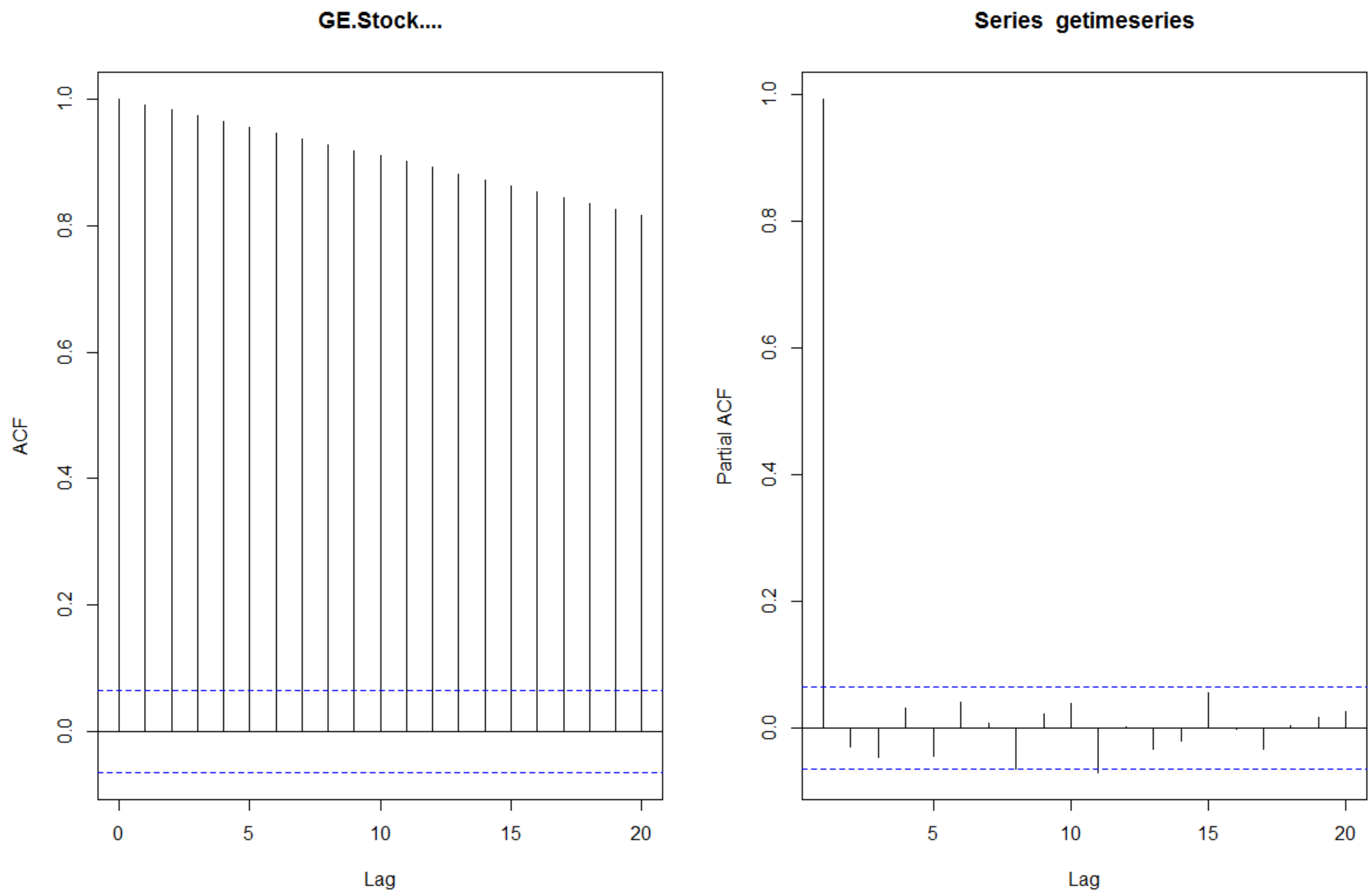






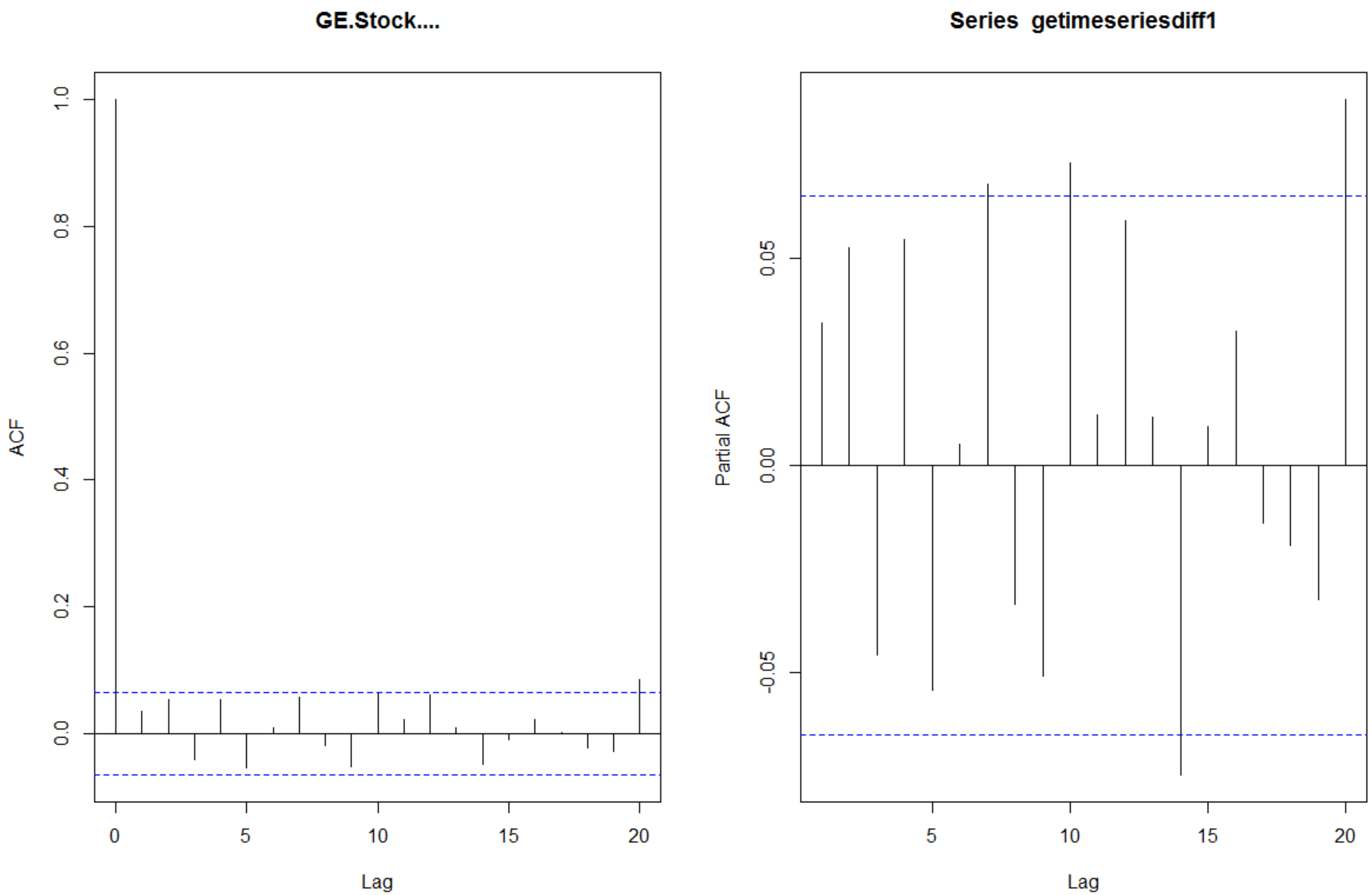
CSE 7202c

ACF and PACF of Stationary and Non-Stationary



Price of GE stock is highly correlated with the previous day's value.

ACF and PACF of Stationary and Non-Stationary



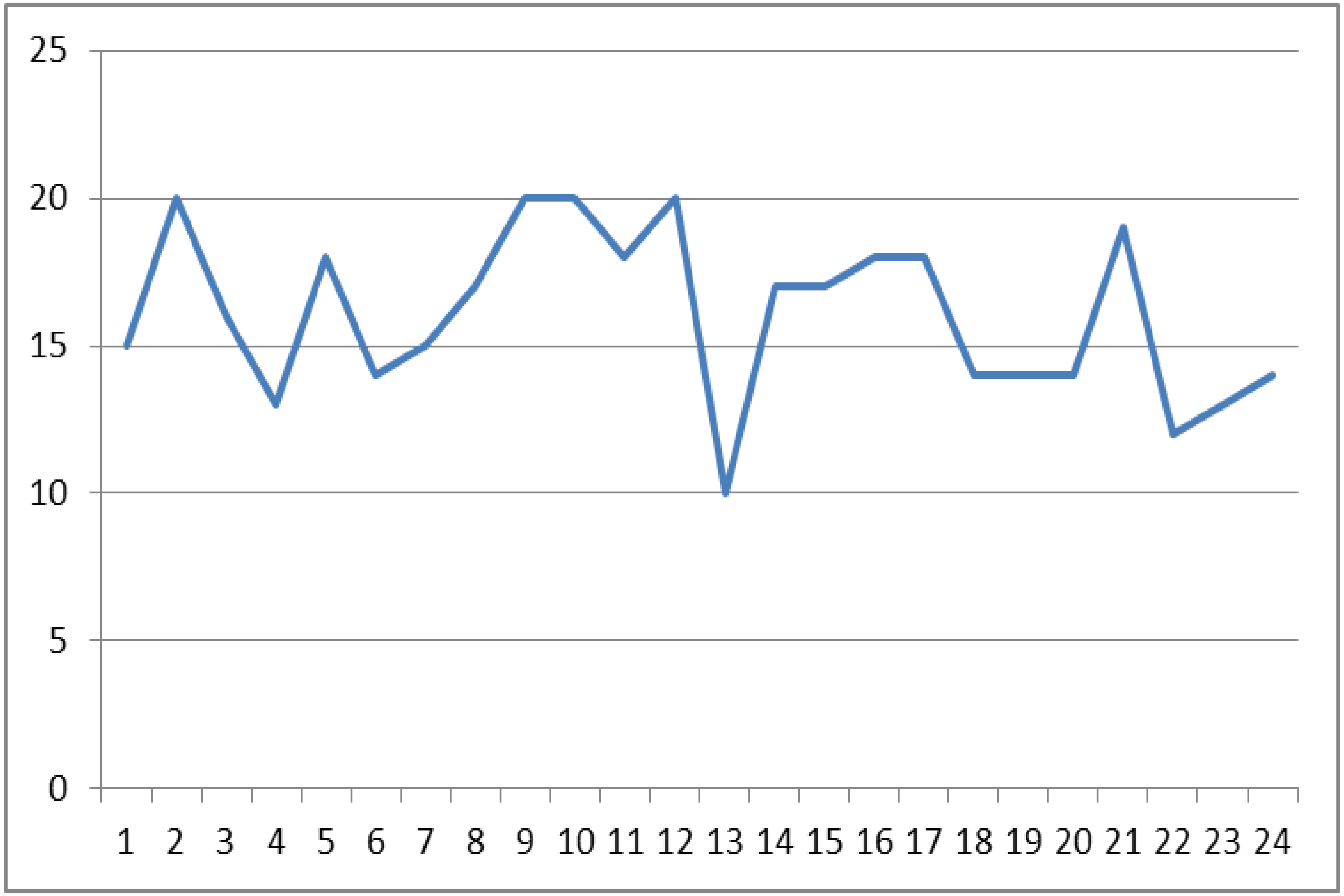
Daily changes in GE stock price are essentially random.

ACF and PACF of Stationary and Non-Stationary

- Non-stationary series have an ACF that remains significant for half a dozen or more lags, rather than quickly declining to zero.

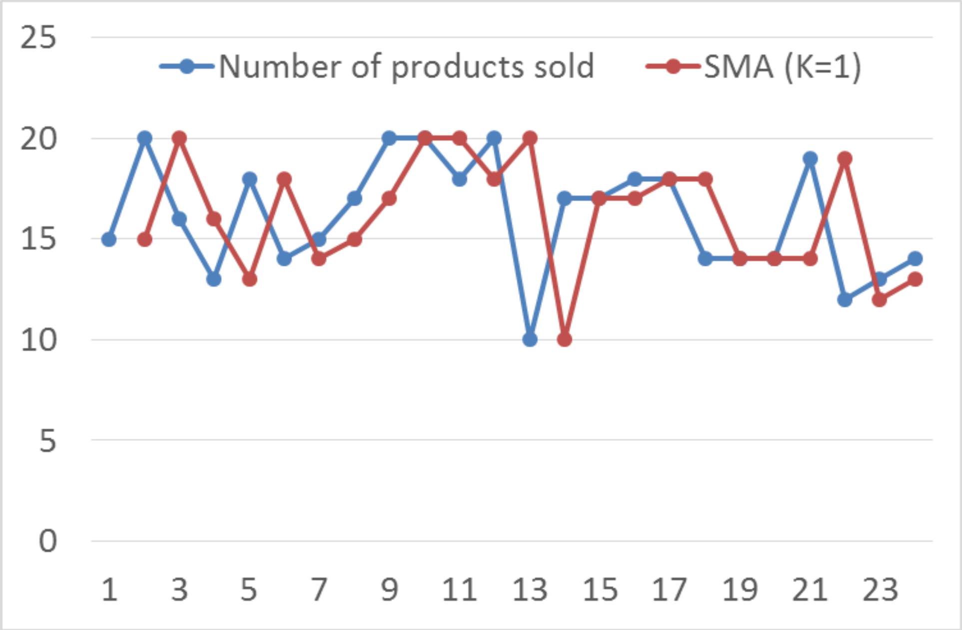
- You must difference such a series until it is stationary before you can identify the process.

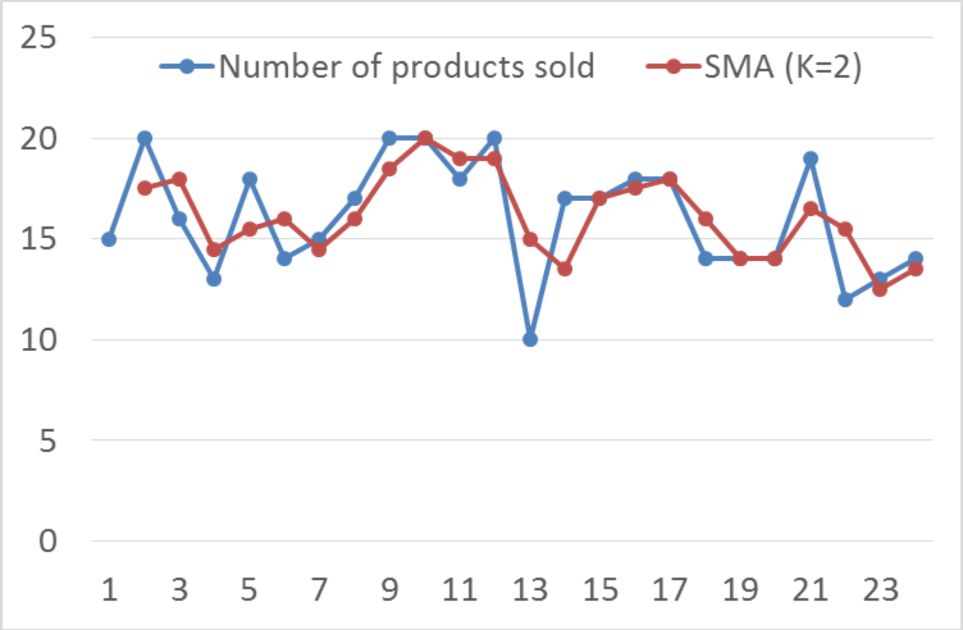
Stationary Model: Moving Averages

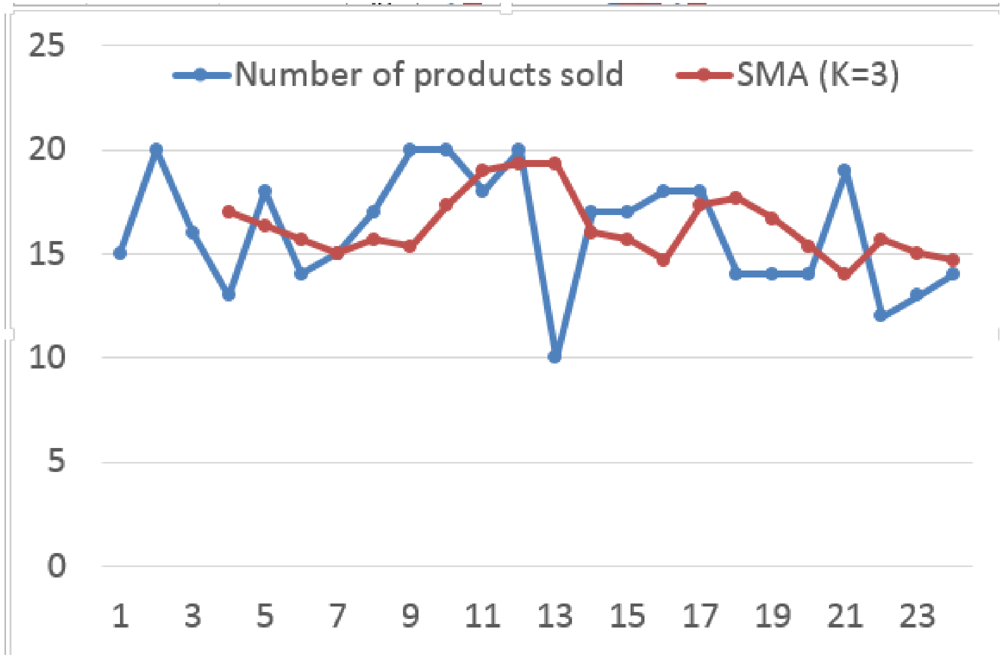


Stationary Model: Case 1 – Simple Moving Averages

Number of products sold	SMA (K=1)	Error	SMA (K=2)	Error	SMA (K=3)	Error
15						
20	15	5				
16	20	4	17.5	1.5		
13	16	3	18	5	17	4
18	13	5	14.5	3.5	16.333333	1.66667
14	18	4	15.5	1.5	15.666667	1.66667
15	14	1	16	1	15	0
17	15	2	14.5	2.5	15.666667	1.33333
20	17	3	16	4	15.333333	4.66667
20	20	0	18.5	1.5	17.333333	2.66667
18	20	2	20	2	19	1
20	18	2	19	1	19.333333	0.66667
10	20	10	19	9	19.333333	9.33333
17	10	7	15	2	16	1
17	17	0	13.5	3.5	15.666667	1.33333
18	17	1	17	1	14.666667	3.33333
18	18	0	17.5	0.5	17.333333	0.66667
14	18	4	18	4	17.666667	3.66667
14	14	0	16	2	16.666667	2.66667
14	14	0	14	0	15.333333	1.33333
19	14	5	14	5	14	5
12	19	7	16.5	4.5	15.666667	3.66667
13	12	1	15.5	2.5	15	2
14	13	1	12.5	1.5	14.666667	0.66667
		2.91304		**2.68182**		**2.49206**







Only decision point is K

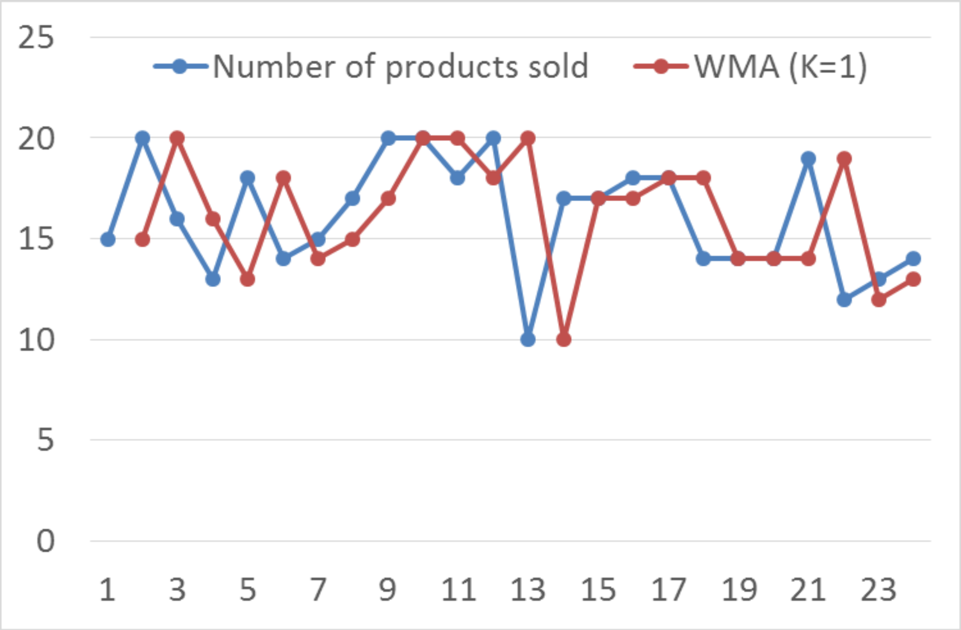
Stationary Model: Case 2 – Weighted Moving Averages

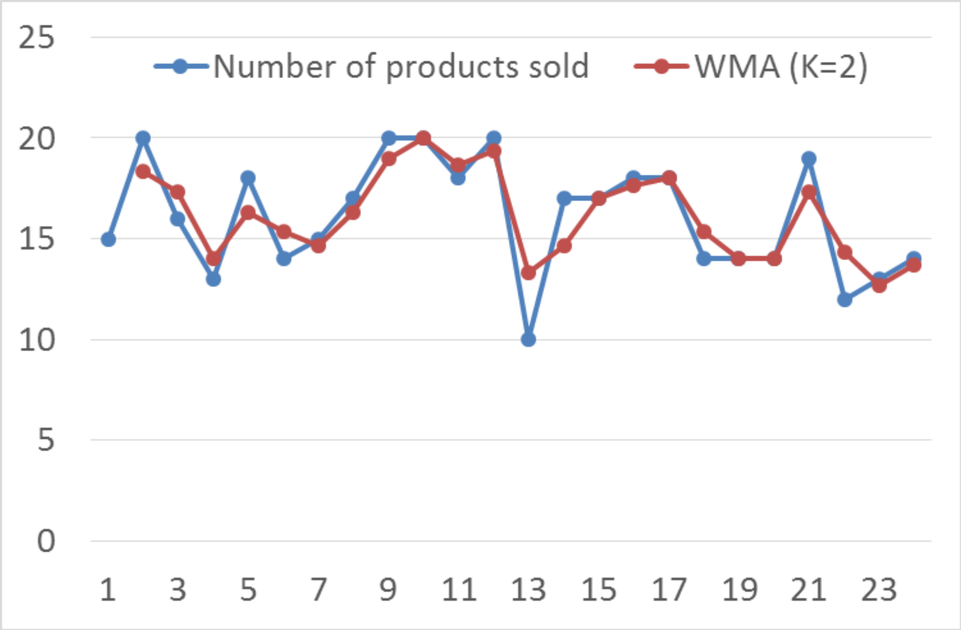
$$\hat{Y}_{t+1} = w_1 Y_t + w_2 Y_{t-1} + \cdots + w_k Y_{t-k+1}$$

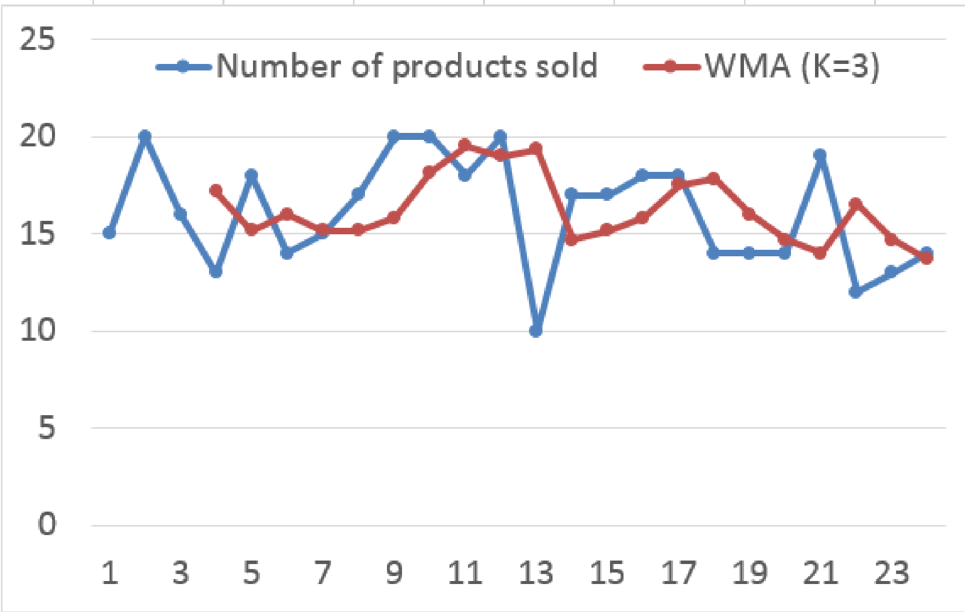
- Typically we choose a time period of moving average and weights are chosen such that the error is minimized

Stationary Model: Case 2 – Weighted Moving Averages

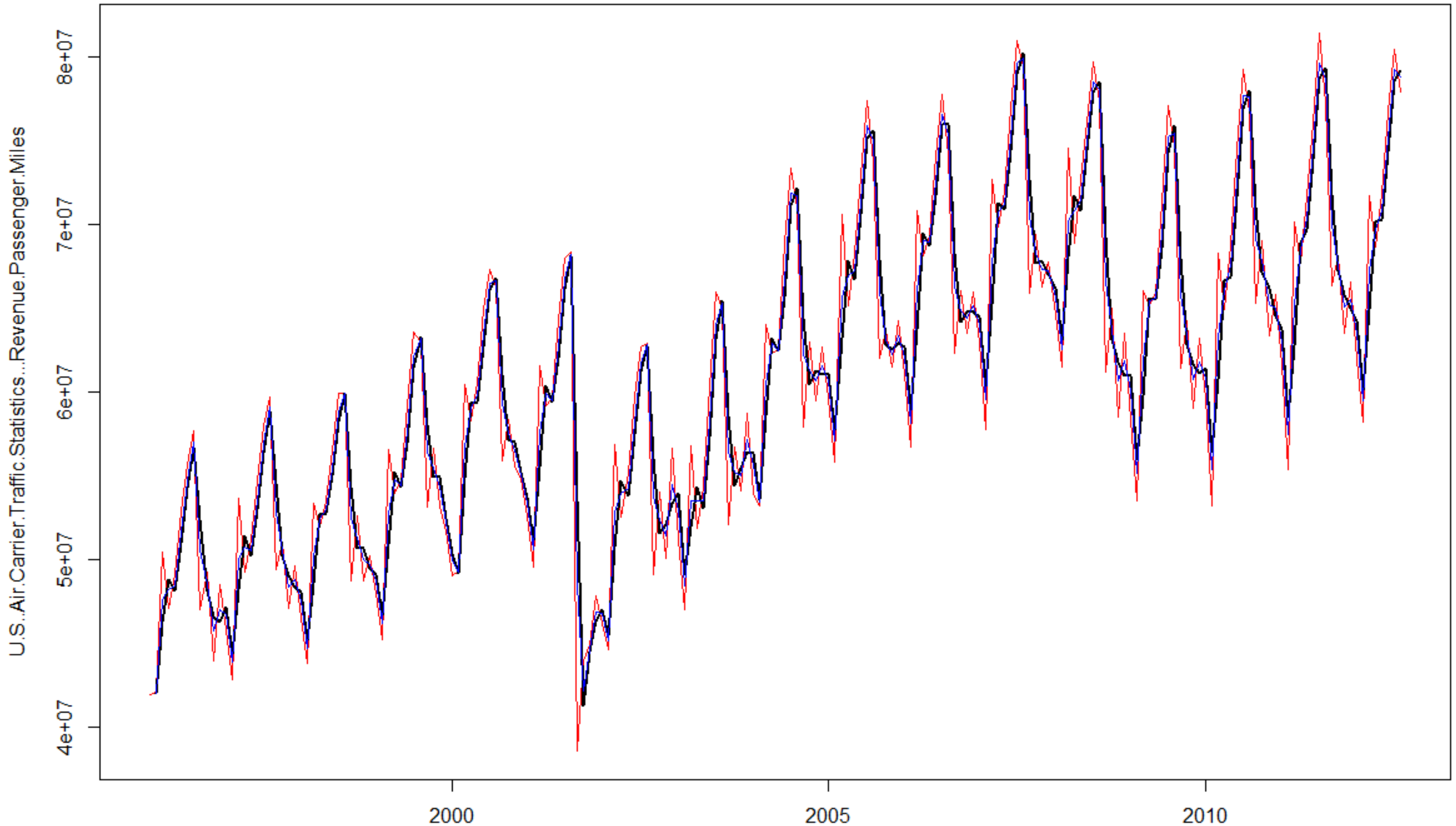
Number of products sold	WMA (K=1)	Error	WMA (K=2)	Error	WMA (K=3)	Error
15						
20	15	5				
16	20	4	18.3333333	2.33333333		
13	16	3	17.3333333	4.33333333	17.1666667	4.16666667
18	13	5	14	4	15.1666667	2.83333333
14	18	4	16.3333333	2.33333333	16	2
15	14	1	15.3333333	0.33333333	15.1666667	0.16666667
17	15	2	14.6666667	2.33333333	15.1666667	1.83333333
20	17	3	16.3333333	3.66666667	15.8333333	4.16666667
20	20	0	19	1	18.1666667	1.83333333
18	20	2	20	2	19.5	1.5
20	18	2	18.6666667	1.33333333	19	1
10	20	10	19.3333333	9.33333333	19.3333333	9.33333333
17	10	7	13.3333333	3.66666667	14.6666667	2.33333333
17	17	0	14.6666667	2.33333333	15.1666667	1.83333333
18	17	1	17	1	15.8333333	2.16666667
18	18	0	17.6666667	0.33333333	17.5	0.5
14	18	4	18	4	17.8333333	3.83333333
14	14	0	15.3333333	1.33333333	16	2
14	14	0	14	0	14.6666667	0.66666667
19	14	5	14	5	14	5
12	19	7	17.3333333	5.33333333	16.5	4.5
13	12	1	14.3333333	1.33333333	14.6666667	1.66666667
14	13	1	12.6666667	1.33333333	13.6666667	0.33333333
		2.91304348		**2.66666667**		**2.55555556**


Number of products sold
WMA (K=1)
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23


Number of products sold
WMA (K=2)
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23


Number of products sold
WMA (K=3)
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23

SMA and WMA – Revenue Passenger Miles




```
> MAPE-SMA 4.093731 > MAPE-WMA 2.729154
```

Stationary Model: Case 3 – Exponential ~~Weighted Moving Averages~~ or Exponential Smoothing

Averaging over long periods dampens fluctuations, removing not only the noise but also trend and seasonality.

Moving averages over short recent periods maintains trend and seasonality but determining an optimum number for periods is tricky, even when using metrics like MAE. If averaged over too few periods, irregularities continue to remain and if averaged over long periods, dampening again becomes a problem.

Exponential smoothing **retains all older periods** while giving a greater weight to more recent periods (hence not a MOVING average).

Caution: It doesn't make any one method superior for all situations.

Stationary Model: Case 3 – Exponential ~~Weighted Moving Averages~~ or Exponential Smoothing

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha\left(Y_t - \hat{Y}_t\right)$$

Above equation indicates that the predicted value for time period t+1 ($\hat{Y}_{t+1}$) is equal to the predicted value for the previous period ($\hat{Y}_t$) plus an adjustment for the error made in predicting the previous period's value ($\alpha\left(Y_t - \hat{Y}_t\right)$).

The parameter α can assume any value between 0 and 1 ($0 \le \alpha \le 1$).

Exponential Smoothing in Other Ways

$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$ can be rewritten variously as

$$= \alpha Y_t + (1-\alpha)\hat{Y}_t$$

$$= Y_t - (1-\alpha)\left(Y_t - \hat{Y}_t\right)$$

$$= \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^n Y_{t-n} + \cdots$$

Exponential Smoothing

- Y at t+1 $$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$
- Y at t+2
- All future predictions are same! This is in accordance with **stationary** assumption.

Exponential Smoothing

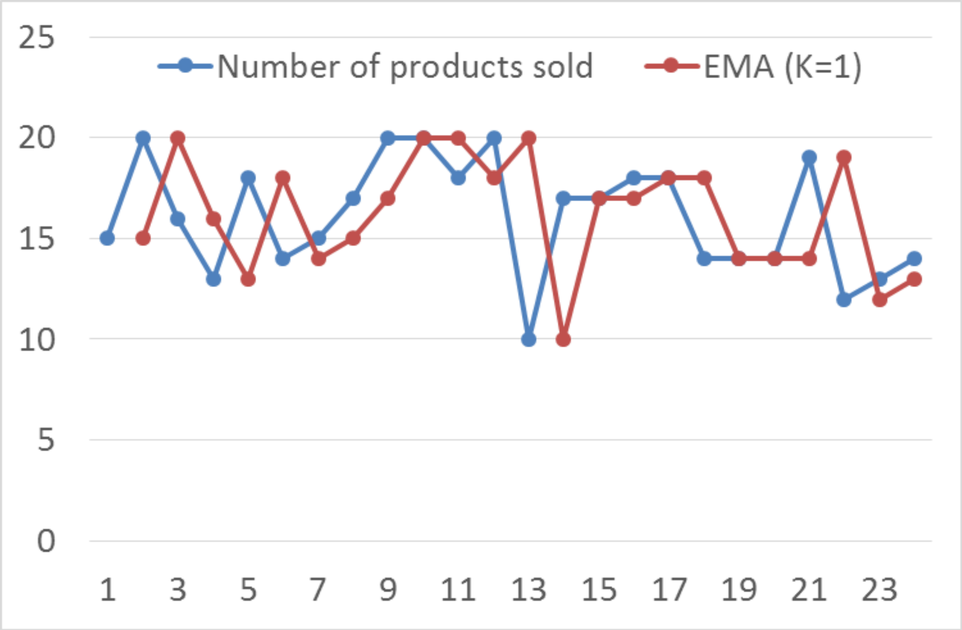
	A	B	C	D	E	F	G
1	Numbe	EMA (K=1)	Error	EMA (K=2)	Error	EMA (K=3)	Error
2	15						
3	20	=A2*1	=ABS(B3-A3)	=15			
4	16	=A3*K2+B3*L2	=ABS(B4-A4)	=A3*K3+D3*L3	=ABS(A4-D4)	=AVERAGE(A2:A3)	
5	13	=A4*K2+B4*L2	=ABS(B5-A5)	=A4*K3+D4*L3	=ABS(A5-D5)	=A4*K4+F4*L4	=ABS(A5-F5)
6	18	=A5*K2+B5*L2	=ABS(B6-A6)	=A5*K3+D5*L3	=ABS(A6-D6)	=A5*K4+F5*L4	=ABS(A6-F6)
7	14	=A6*K2+B6*L2	=ABS(B7-A7)	=A6*K3+D6*L3	=ABS(A7-D7)	=A6*K4+F6*L4	=ABS(A7-F7)
8	15	=A7*K2+B7*L2	=ABS(B8-A8)	=A7*K3+D7*L3	=ABS(A8-D8)	=A7*K4+F7*L4	=ABS(A8-F8)
9	17	=A8*K2+B8*L2	=ABS(B9-A9)	=A8*K3+D8*L3	=ABS(A9-D9)	=A8*K4+F8*L4	=ABS(A9-F9)
10	20	=A9*K2+B9*L2	=ABS(B10-A10)	=A9*K3+D9*L3	=ABS(A10-D10)	=A9*K4+F9*L4	=ABS(A10-F10)
11	20	=A10*K2+B10*L2	=ABS(B11-A11)	=A10*K3+D10*L3	=ABS(A11-D11)	=A10*K4+F10*L4	=ABS(A11-F11)
12	18	=A11*K2+B11*L2	=ABS(B12-A12)	=A11*K3+D11*L3	=ABS(A12-D12)	=A11*K4+F11*L4	=ABS(A12-F12)
13	20	=A12*K2+B12*L2	=ABS(B13-A13)	=A12*K3+D12*L3	=ABS(A13-D13)	=A12*K4+F12*L4	=ABS(A13-F13)
14	10	=A13*K2+B13*L2	=ABS(B14-A14)	=A13*K3+D13*L3	=ABS(A14-D14)	=A13*K4+F13*L4	=ABS(A14-F14)
15	17	=A14*K2+B14*L2	=ABS(B15-A15)	=A14*K3+D14*L3	=ABS(A15-D15)	=A14*K4+F14*L4	=ABS(A15-F15)
16	17	=A15*K2+B15*L2	=ABS(B16-A16)	=A15*K3+D15*L3	=ABS(A16-D16)	=A15*K4+F15*L4	=ABS(A16-F16)
17	18	=A16*K2+B16*L2	=ABS(B17-A17)	=A16*K3+D16*L3	=ABS(A17-D17)	=A16*K4+F16*L4	=ABS(A17-F17)
18	18	=A17*K2+B17*L2	=ABS(B18-A18)	=A17*K3+D17*L3	=ABS(A18-D18)	=A17*K4+F17*L4	=ABS(A18-F18)
19	14	=A18*K2+B18*L2	=ABS(B19-A19)	=A18*K3+D18*L3	=ABS(A19-D19)	=A18*K4+F18*L4	=ABS(A19-F19)
20	14	=A19*K2+B19*L2	=ABS(B20-A20)	=A19*K3+D19*L3	=ABS(A20-D20)	=A19*K4+F19*L4	=ABS(A20-F20)
21	14	=A20*K2+B20*L2	=ABS(B21-A21)	=A20*K3+D20*L3	=ABS(A21-D21)	=A20*K4+F20*L4	=ABS(A21-F21)
22	19	=A21*K2+B21*L2	=ABS(B22-A22)	=A21*K3+D21*L3	=ABS(A22-D22)	=A21*K4+F21*L4	=ABS(A22-F22)
23	12	=A22*K2+B22*L2	=ABS(B23-A23)	=A22*K3+D22*L3	=ABS(A23-D23)	=A22*K4+F22*L4	=ABS(A23-F23)
24	13	=A23*K2+B23*L2	=ABS(B24-A24)	=A23*K3+D23*L3	=ABS(A24-D24)	=A23*K4+F23*L4	=ABS(A24-F24)
25	14	=A24*K2+B24*L2	=ABS(B25-A25)	=A24*K3+D24*L3	=ABS(A25-D25)	=A24*K4+F24*L4	=ABS(A25-F25)
26			**=AVERAGE(C3:C25)**		**=AVERAGE(E3:E25)**		**=AVERAGE(G3:G25)**

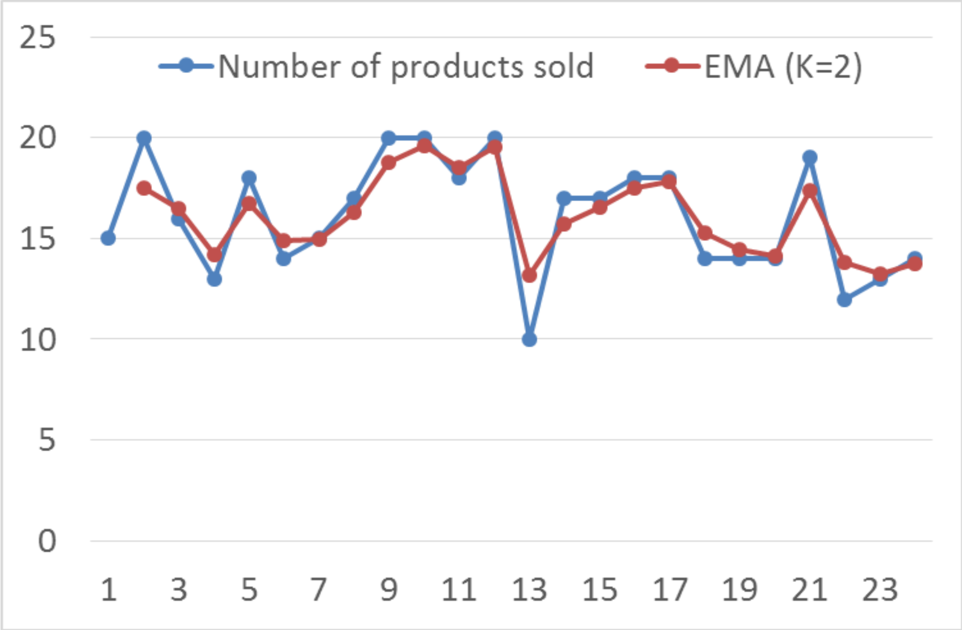
J	K	L
K	2/(K+1)	1-[2/(K+1)]
1	1	=1-K2
2	=2/(J3+1)	=1-K3
3	=2/(J4+1)	=1-K4
4	=2/(J5+1)	=1-K5

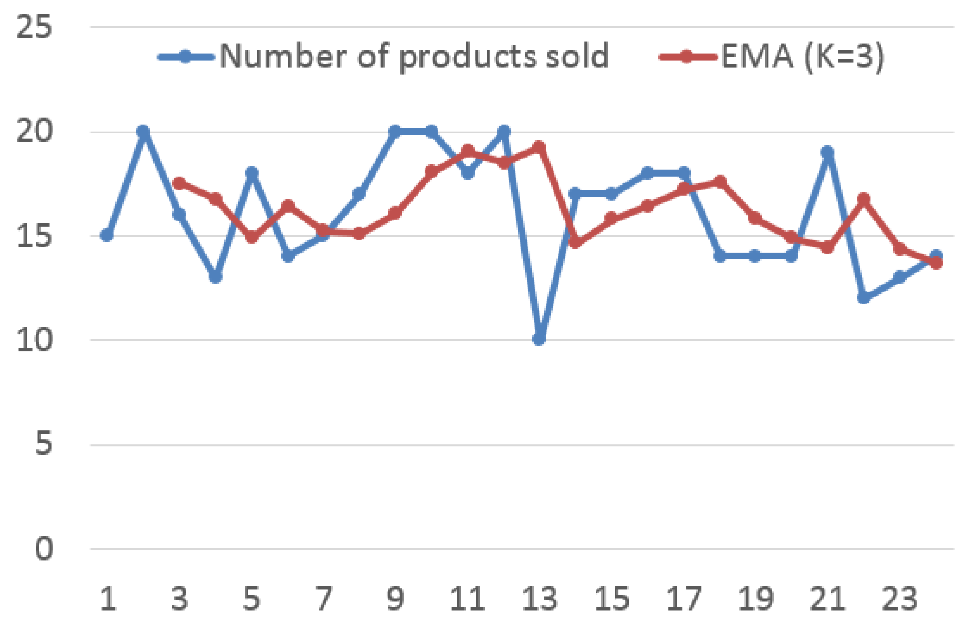
$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$$

Exponential Smoothing

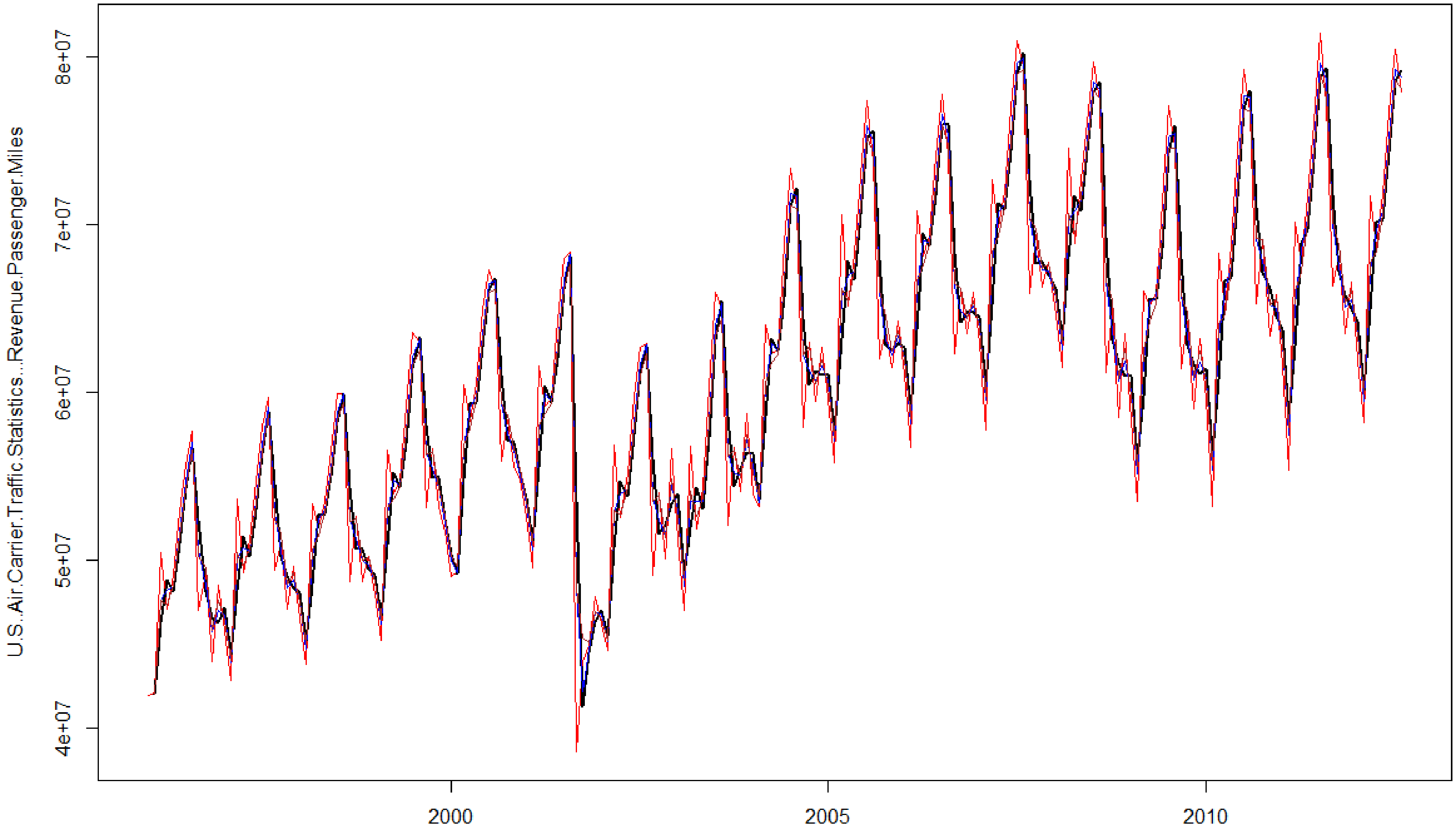
	A	B	C	D	E	F	G	H	I
1	Number of	EMA (K=1)	Error	EMA (K=2)	Error	EMA (K=3)	Error	EMA (K=4)	Error
2	15								
3	20	15	5	15					
4	16	20	4	18.3333333	2.33333	17.5			
5	13	16	3	16.7777778	3.77778	16.75	3.75	17	
6	18	13	5	14.2592593	3.74074	14.875	3.125	15.4	2.6
7	14	18	4	16.7530864	2.75309	16.4375	2.4375	16.44	2.44
8	15	14	1	14.9176955	0.0823	15.21875	0.21875	15.464	0.464
9	17	15	2	14.9725652	2.02743	15.109375	1.890625	15.2784	1.7216
10	20	17	3	16.3241884	3.67581	16.054688	3.945313	15.96704	4.03296
11	20	20	0	18.7747295	1.22527	18.027344	1.972656	17.580224	2.41978
12	18	20	2	19.5915765	1.59158	19.013672	1.013672	18.548134	0.54813
13	20	18	2	18.5305255	1.46947	18.506836	1.493164	18.328881	1.67112
14	10	20	10	19.5101752	9.51018	19.253418	9.253418	18.997328	8.99733
15	17	10	7	13.1700584	3.82994	14.626709	2.373291	15.398397	1.6016
16	17	17	0	15.7233528	1.27665	15.813354	1.186646	16.039038	0.96096
17	18	17	1	16.5744509	1.42555	16.406677	1.593323	16.423423	1.57658
18	18	18	0	17.524817	0.47518	17.203339	0.796661	17.054054	0.94595
19	14	18	4	17.8416057	3.84161	17.601669	3.601669	17.432432	3.43243
20	14	14	0	15.2805352	1.28054	15.800835	1.800835	16.059459	2.05946
21	14	14	0	14.4268451	0.42685	14.900417	0.900417	15.235676	1.23568
22	19	14	5	14.1422817	4.85772	14.450209	4.549791	14.741405	4.25859
23	12	19	7	17.3807606	5.38076	16.725104	4.725104	16.444843	4.44484
24	13	12	1	13.7935869	0.79359	14.362552	1.362552	14.666906	1.66691
25	14	13	1	13.264529	0.73547	13.681276	0.318724	14.000144	0.00014
26			**2.913043**		**2.56867**		**2.49091**		**2.3539**


Number of products sold
EMA (K=1)
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23


Number of products sold
EMA (K=2)
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23


Number of products sold
EMA (K=3)
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23

SMA, WMA and Exponential Smoothing – RPM




```
> MAPESMA 4.093731 > MAPEWMA 2.729154 > MAPEEMA 2.541979
```

ADDING TREND AND SEASONALITY TO MOVING AVERAGE PROCESSES

Holt-Winters Method

- 3 components – Trend, Seasonality and Random/Level/Error/Irregular.

- 3 weights – smoothing parameters – are used to update components at each period.

- Initial values for error and trend components are obtained using linear regression on time.

- Initial values for seasonal component are obtained from a dummy variable regression using de-trended data.

- In the error equation, the series is seasonally adjusted by subtracting the seasonal component.

Holt-Winters Method

Additive Seasonality

$$\hat{Y}_t = E_{t-1} + T_{t-1} + S_{t-p}$$

$$\hat{Y}_{t+n} = E_t + nT_t + S_{t+n-p}$$

Multiplicative Seasonality

$$\hat{Y}_t = (E_{t-1} + T_{t-1})S_{t-p}$$

$$\hat{Y}_{t+n} = (E_t + nT_t)S_{t+n-p}$$

The 3 smoothing equations are:

Additive:

$$E_t = \alpha\left(Y_t - S_{t-p}\right) + (1-\alpha)(E_{t-1} + T_{t-1})$$

$$T_t = \beta(E_t - E_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma(Y_t - E_t) + (1-\gamma)S_{t-p}$$

Multiplicative:

$$E_t = \alpha\frac{Y_t}{S_{t-p}} + (1-\alpha)(E_{t-1} + T_{t-1})$$

$$T_t = \beta(E_t - E_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma\frac{Y_t}{E_t} + (1-\gamma)S_{t-p}$$

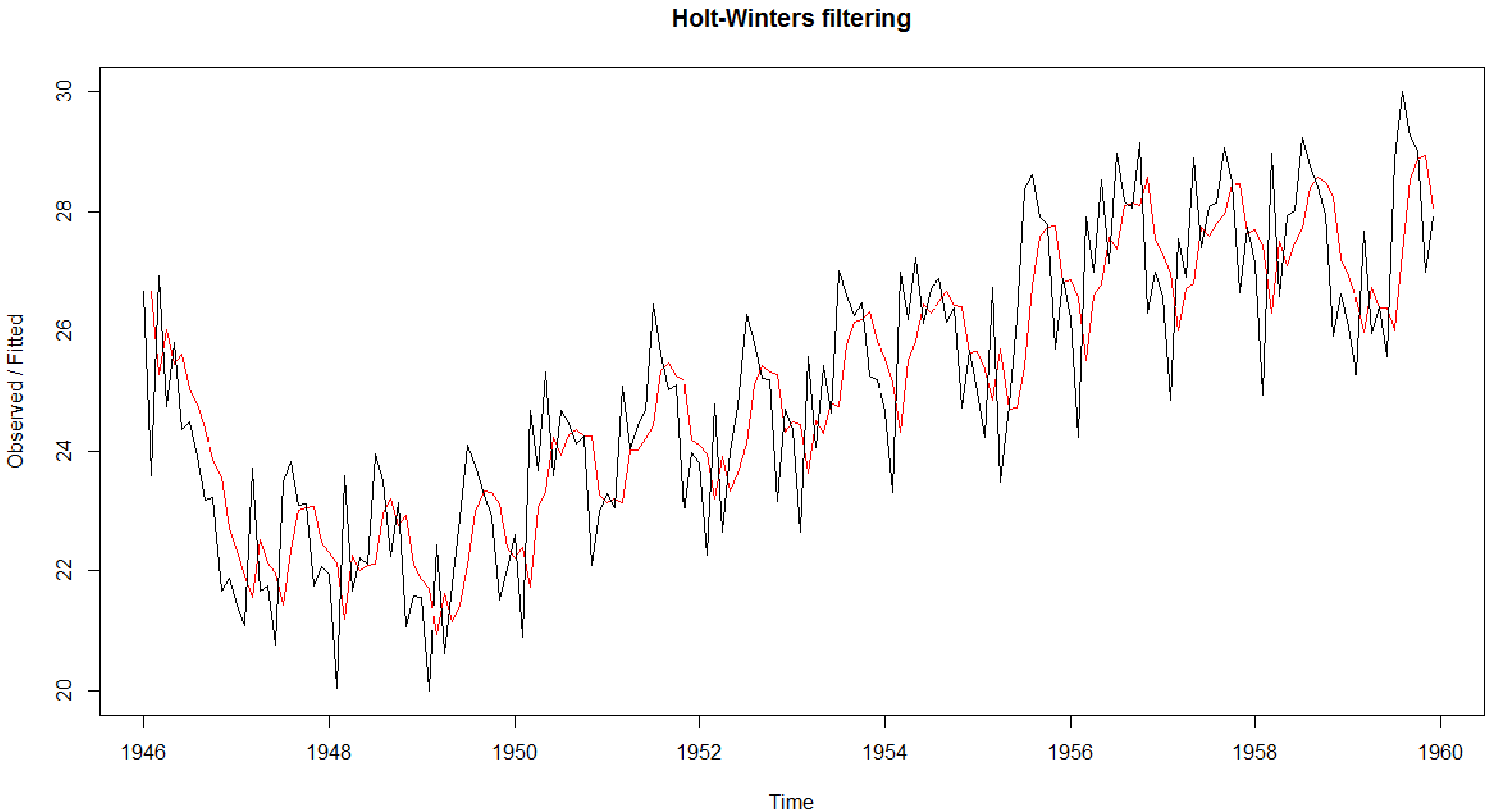
Holt-Winters Additive Seasonal Model

SSE = 103627

Holt-Winters Multiplicative Seasonal Model

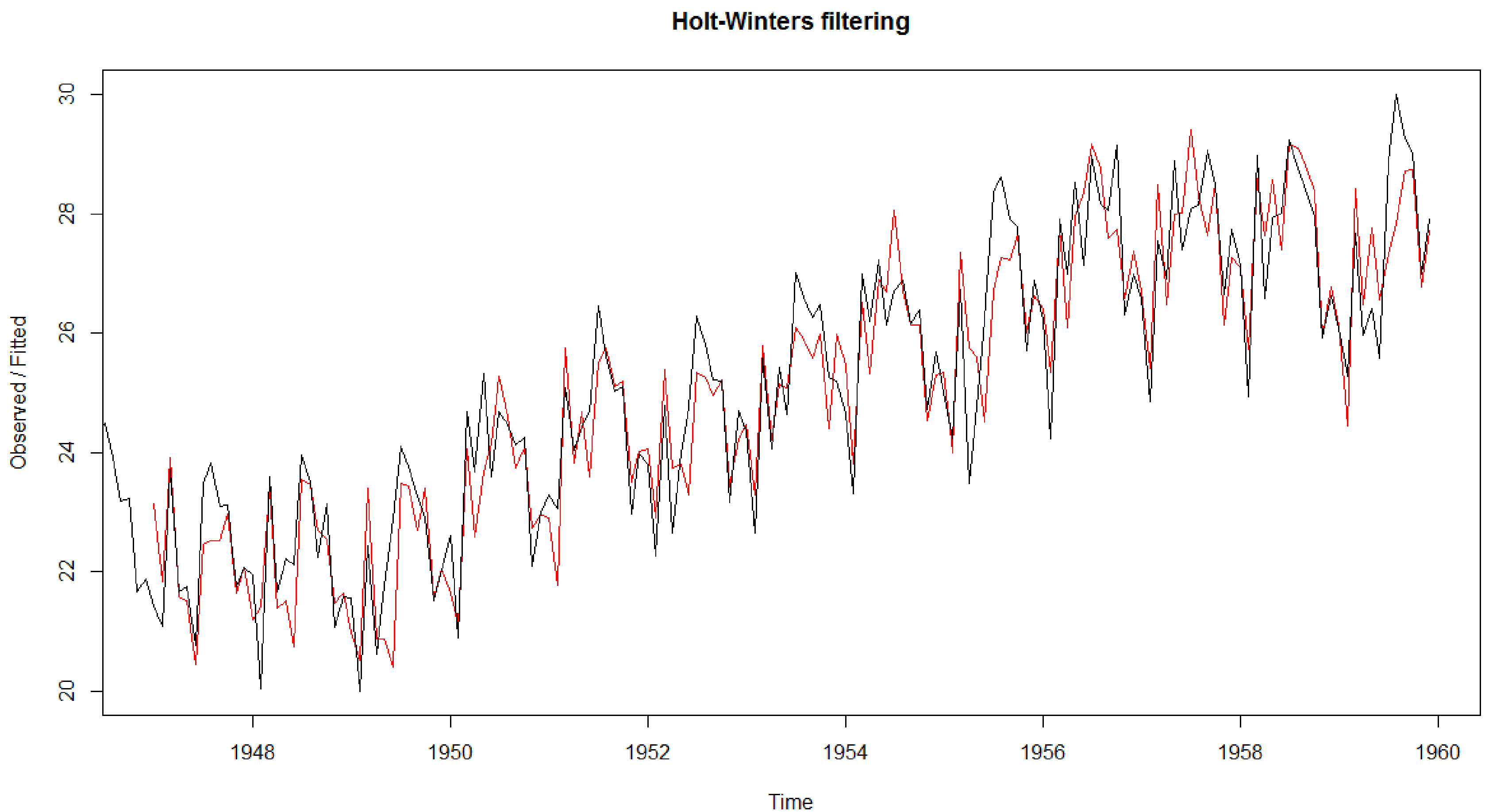
SSE = 49816

Holt-Winters Method: Only Randomness




```
> birthsforecast$SSE
[1] 281.8759
```

Holt-Winters Method: All Components

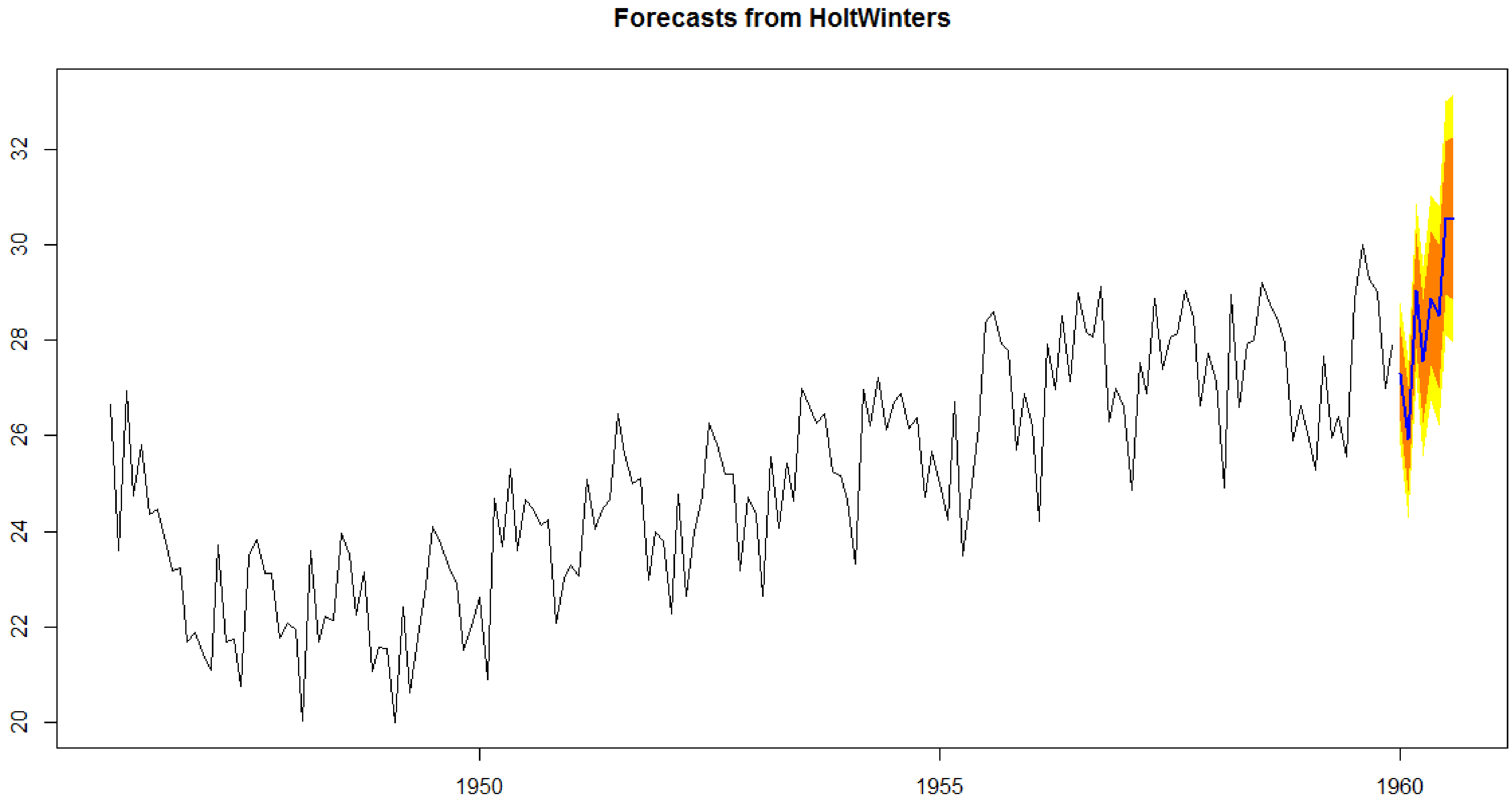



```
> birthsforecast$SSE
[1] 90.94058
```

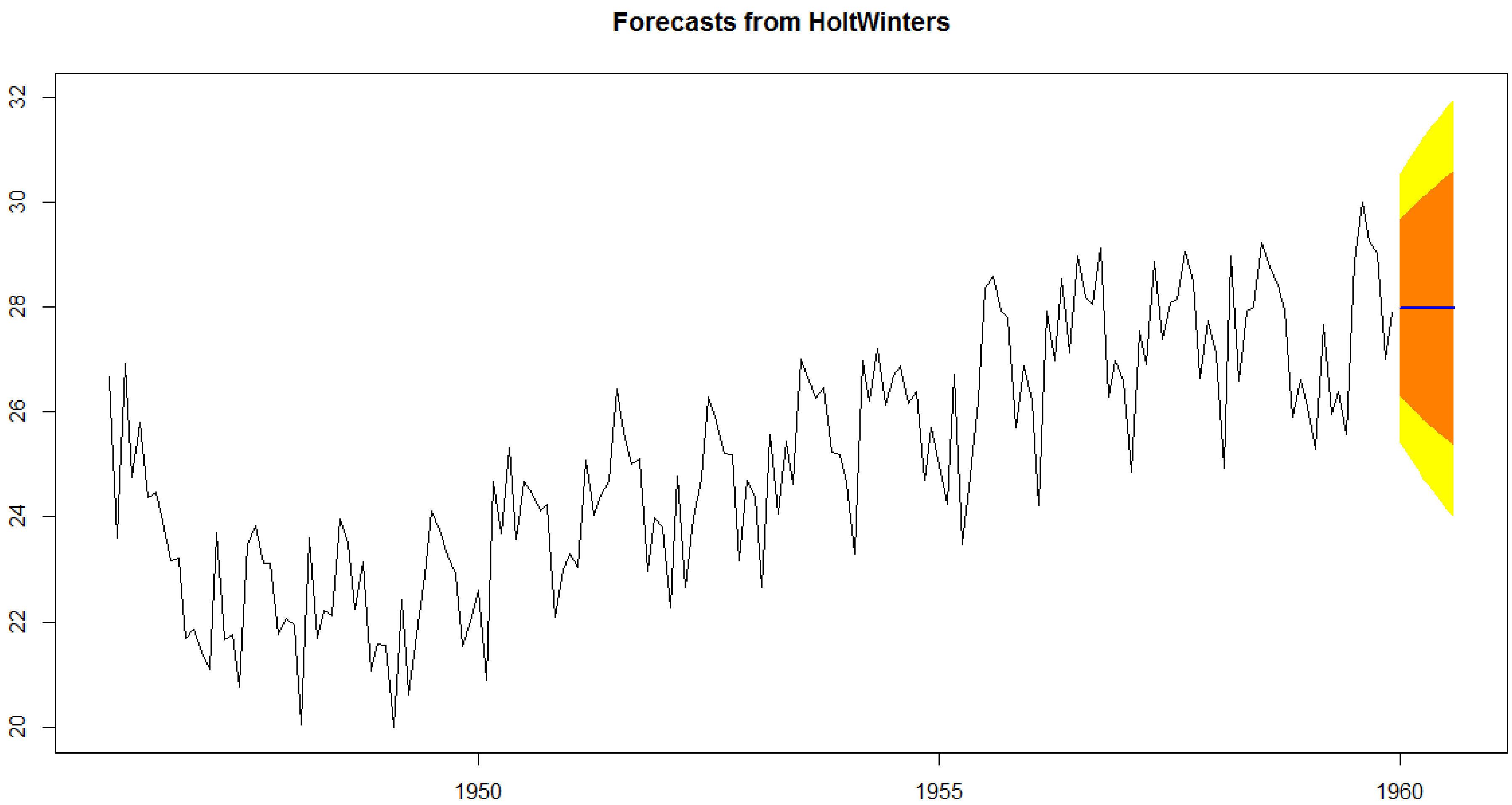
Holt-Winters Method: All Components

```
Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:
HoltWinters(x = birthstimeseries)

Smoothing parameters:
 alpha: 0.4823655
 beta : 0.02988495
 gamma: 0.563186

Coefficients:
           [,1]
a   28.04366357
b    0.04199921
s1  -0.78546221
s2  -2.19944507
s3   0.87813012
s4  -0.65164728
s5   0.63427267
s6   0.21182821
s7   2.23177191
s8   2.17167733
s9   1.52077678
s10  1.16900861
s11 -0.97500043
s12 -0.18636055
> birthsforecast$fitted
             xhat    level         trend      season
Jan 1947 23.13579 23.81055 -0.1567618007 -0.51798958
```

Holt-Winters Method: Forecasting



Holt-Winters Method: Forecasting with No Trend and Seasonality



AR, MA AND ARIMA MODELS

AR(p) models

- Auto-regressive model of order p

$$\hat{y}_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p}$$

- We find the best value of parameters (β_1, β_2,…) that minimize the errors in forecast of $\hat{y}_t$.

- The order of the model p is determined based on the number beyond which PACF terms are zero.

Moving Average or MA(q) models

- Model attempts to predict future values using past error in predictions $\varepsilon_1 = \hat{y}_1 - y_1$
- So MA(2) model is

$$\hat{y}_t = \mu + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2}$$

Where, μ is the average value of the time series

- Again, the parameters (ϕ_1 , ϕ_2) are determined so that prediction error is minimized.
- q is the maximum lag beyond which the ACF is 0

ARMA(p,q) model

- Combines both AR(p) and MA(q) models

ARIMA(p,d,q) Model

- *p* is the number of autoregressive terms (a linear regression of the current value of the series against one or more prior values of the series)
  - Maximum lag beyond which PACF is 0
- *d* is the number of non-seasonal differences (order of the differencing) used to make the time series stationary
- *q* is the number of past prediction error terms used for the future forecasts.

Box–Jenkins Methodology

- Model identification and model selection
  - Make sure variables are stationary. Difference as necessary to get a constant mean and transformations to get constant variance.
  - Check for seasonality: Decays and spikes at regular intervals in ACF plots.
- Parameter estimation
  - Compute coefficients that best fit the selected model.
- Model checking
  - Check if residuals are independent of each other and constant in mean and variance over time (white noise).

http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf

Model Selection

- Check ACF, PACF
- Identify important lag periods
- Create a data frame (table) with these past lag values as independent variables and value to be predicted as dependent variable
- Perform autoregression (AR models)
- To incorporate randomness, use MA

SHAPE	INDICATED MODEL
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

Model Selection in Practice

- There are techniques that automate model selection

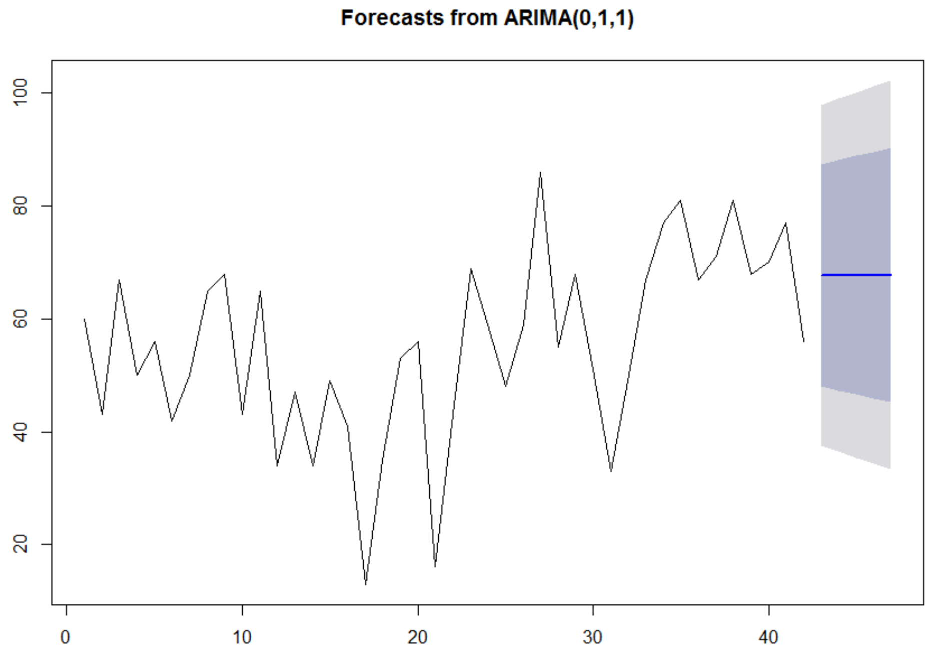
- *auto.Arima* command in R picks the best p,d & q parameters for ARIMA(p,d,q)

Kings Life Expectancy

Source: http://robjhyndman.com/tsdldata/misc/kings.dat

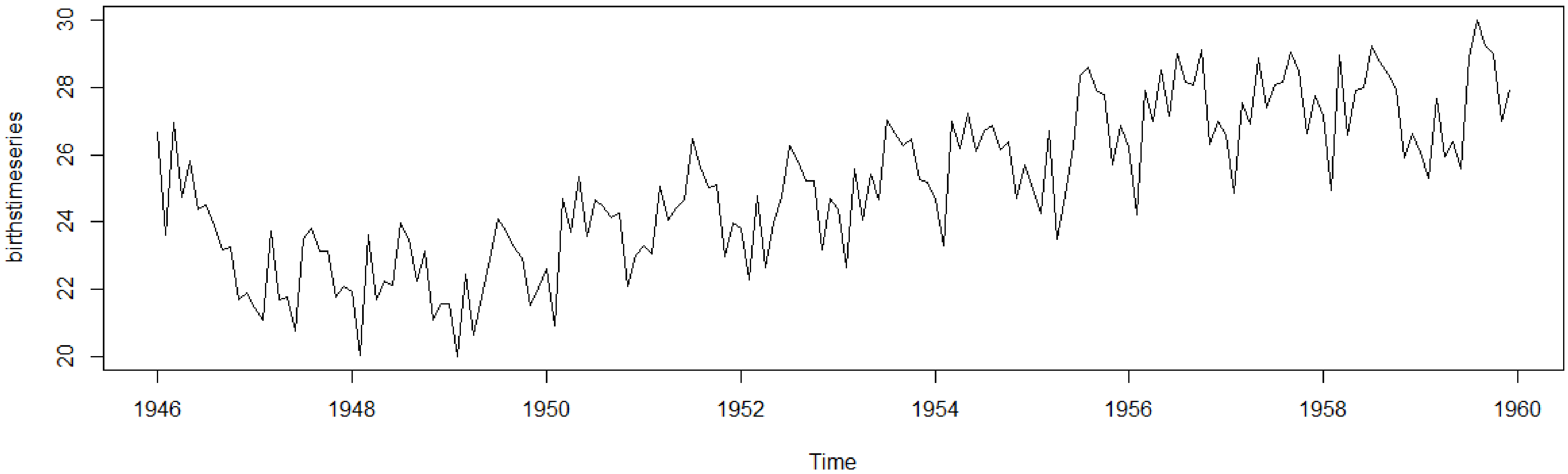
Kings Life expectancy: ARIMA forecast

```
> kingsArima <- auto.arima(kingstimeseries)
> kingsArima
Series: kingstimeseries
ARIMA(0,1,1)

Coefficients:
          ma1
      -0.7218
s.e.   0.1208

sigma^2 estimated as 236.2:  log likelihood=-170.06
AIC=344.13   AICc=344.44   BIC=347.56
> # Forecast the life of next 5 kings
> kingsforecast <- forecast.Arima(kingsArima, h=5)
>
> plot.forecast(kingsforecast)
>
```

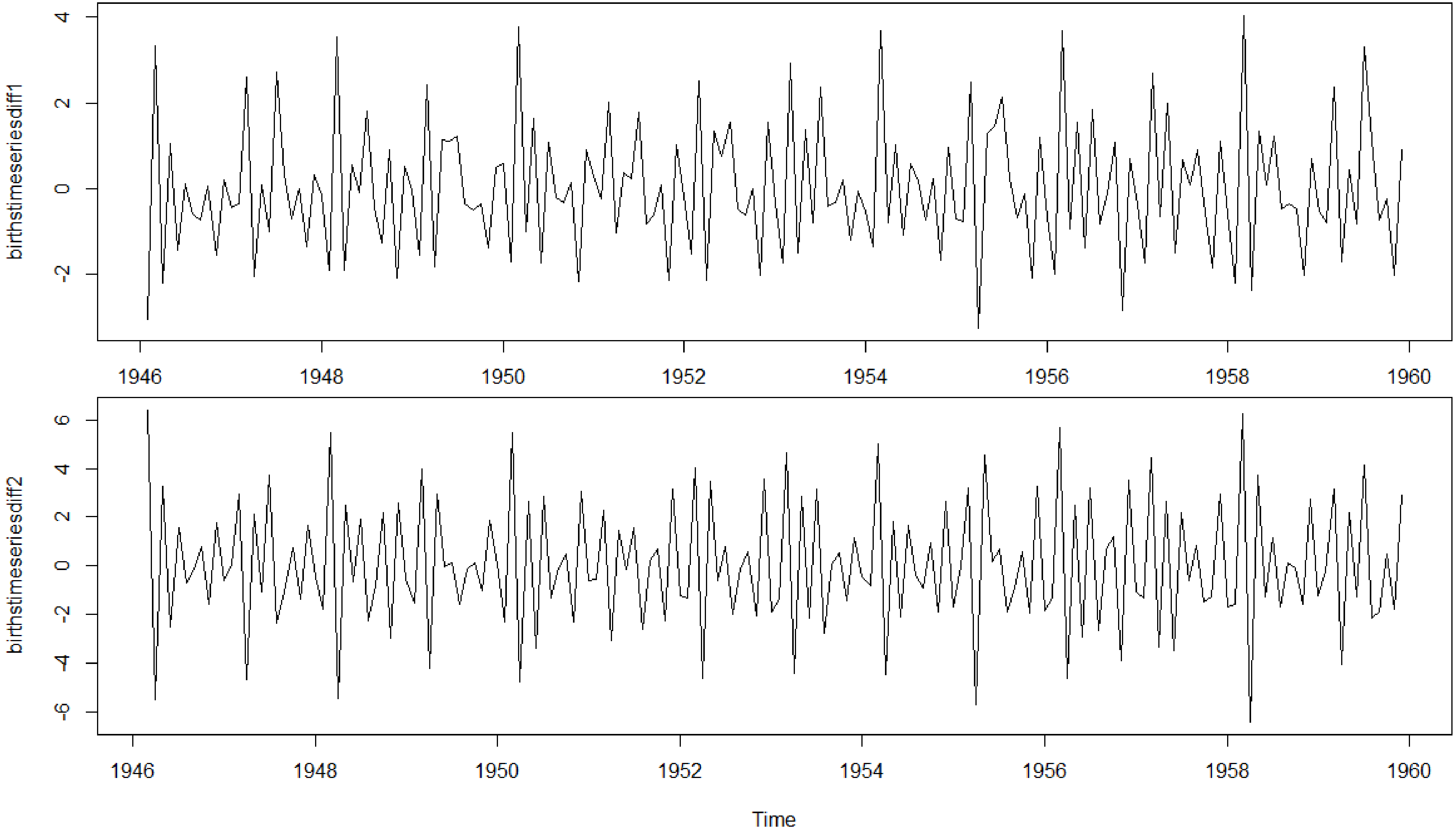


- Non-seasonal ARIMA models are denoted ARIMA(p,d,q)
- Seasonal ARIMA (SARIMA) models are denoted ARIMA(p,d,q)(P,D,Q)$_m$, where m refers to the number of periods in each season and (P,D,Q) refer to the autoregressive, differencing and moving average terms of the seasonal part of the ARIMA model.

Birth Timeseries: Stationary?



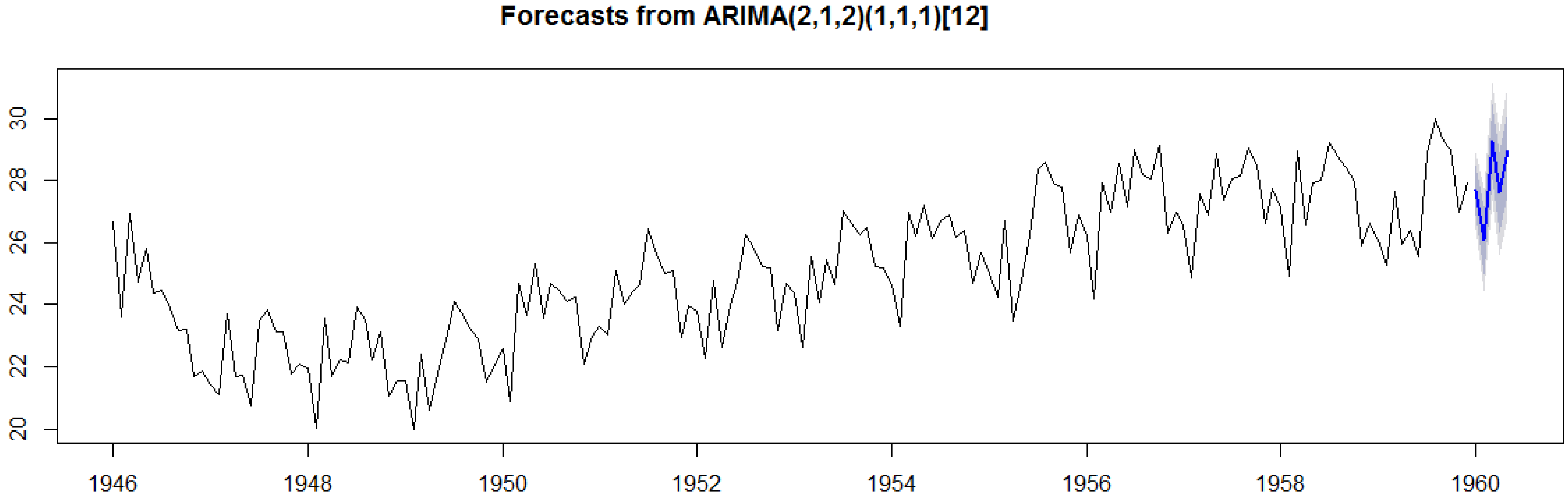
ARIMA(p,1,q) and ARIMA(p,2,q)



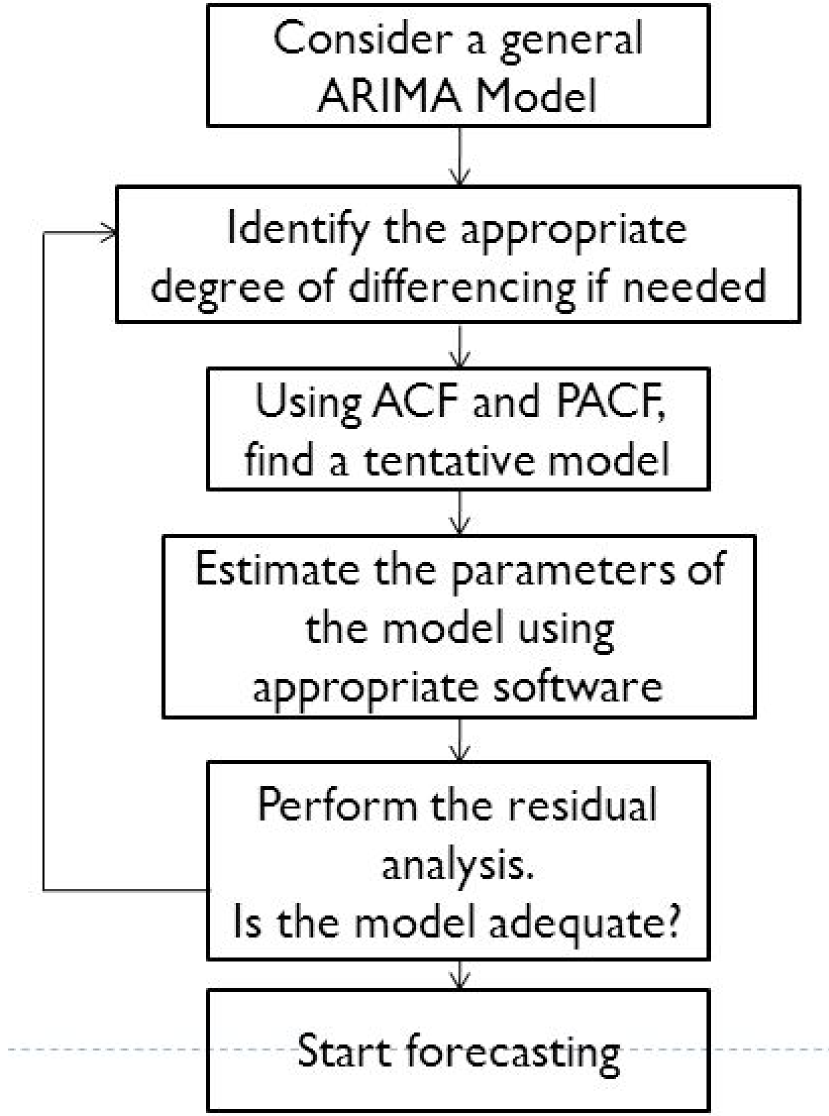
Seasonal ARIMA Model

```
Series: birthstimeseries
ARIMA(2,1,2)(1,1,1)[12]

Coefficients:
         ar1      ar2      ma1     ma2     sar1     sma1
      0.6539  -0.4540  -0.7255  0.2532  -0.2427  -0.8451
s.e.  0.3004   0.2429   0.3228  0.2879   0.0985   0.0995

sigma^2 estimated as 0.3918:  log likelihood=-157.45
AIC=328.91   AICc=329.67   BIC=350.21
>
```

Seasonal ARIMA Model - Forecast



Time Series Model Building Using ARIMA

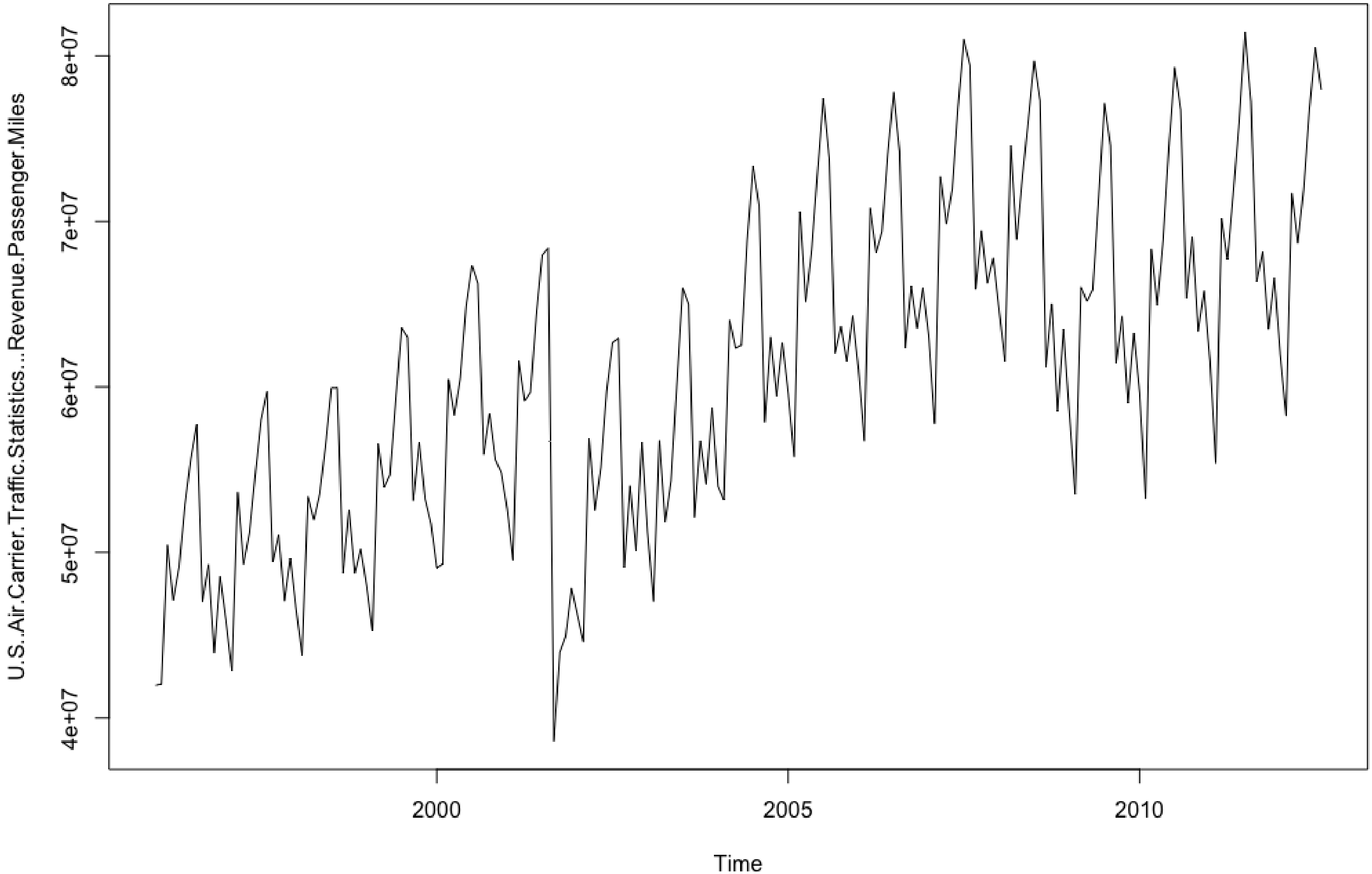


Time Series Model Building Using ARIMA

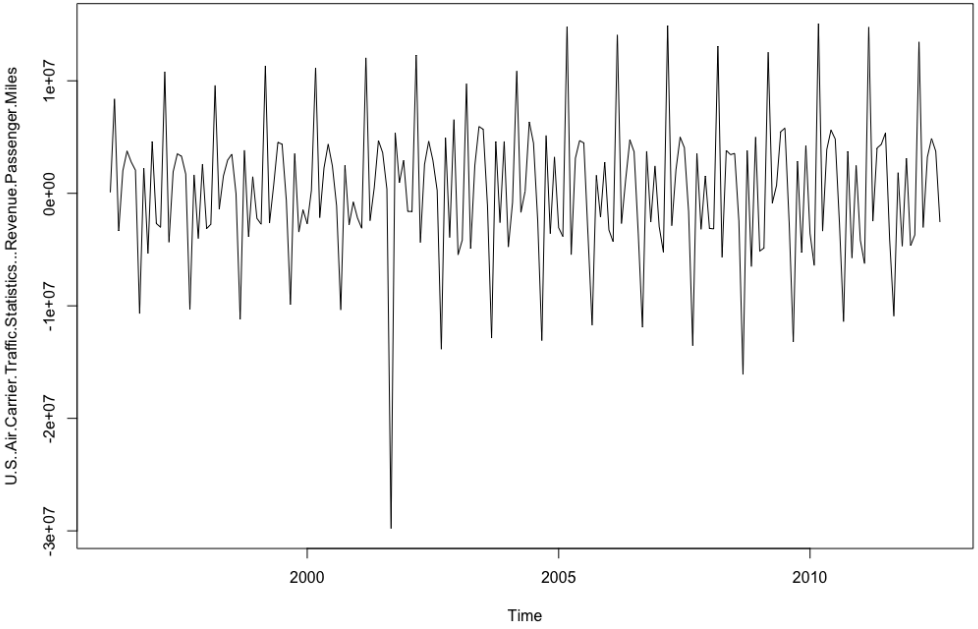
A nice summary of rules for identifying ARIMA models

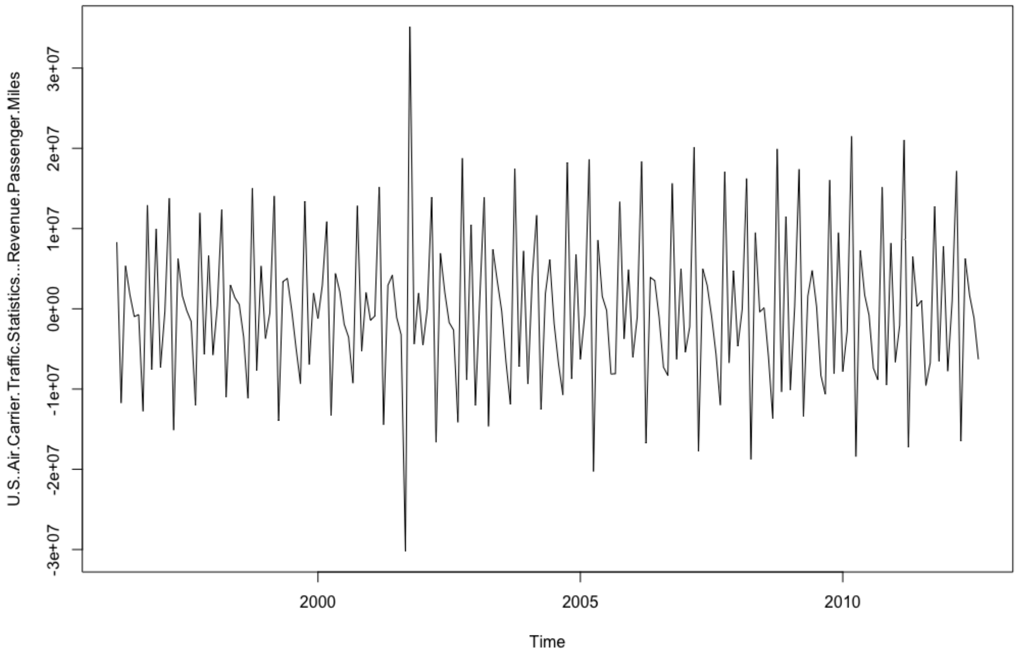
http://people.duke.edu/~rnau/arimrule.htm

Time Series Model Building Using ARIMA - RPM



Time Series Model Building Using ARIMA - RPM Differencing





Time Series Model Building Using ARIMA - RPM Auto ARIMA

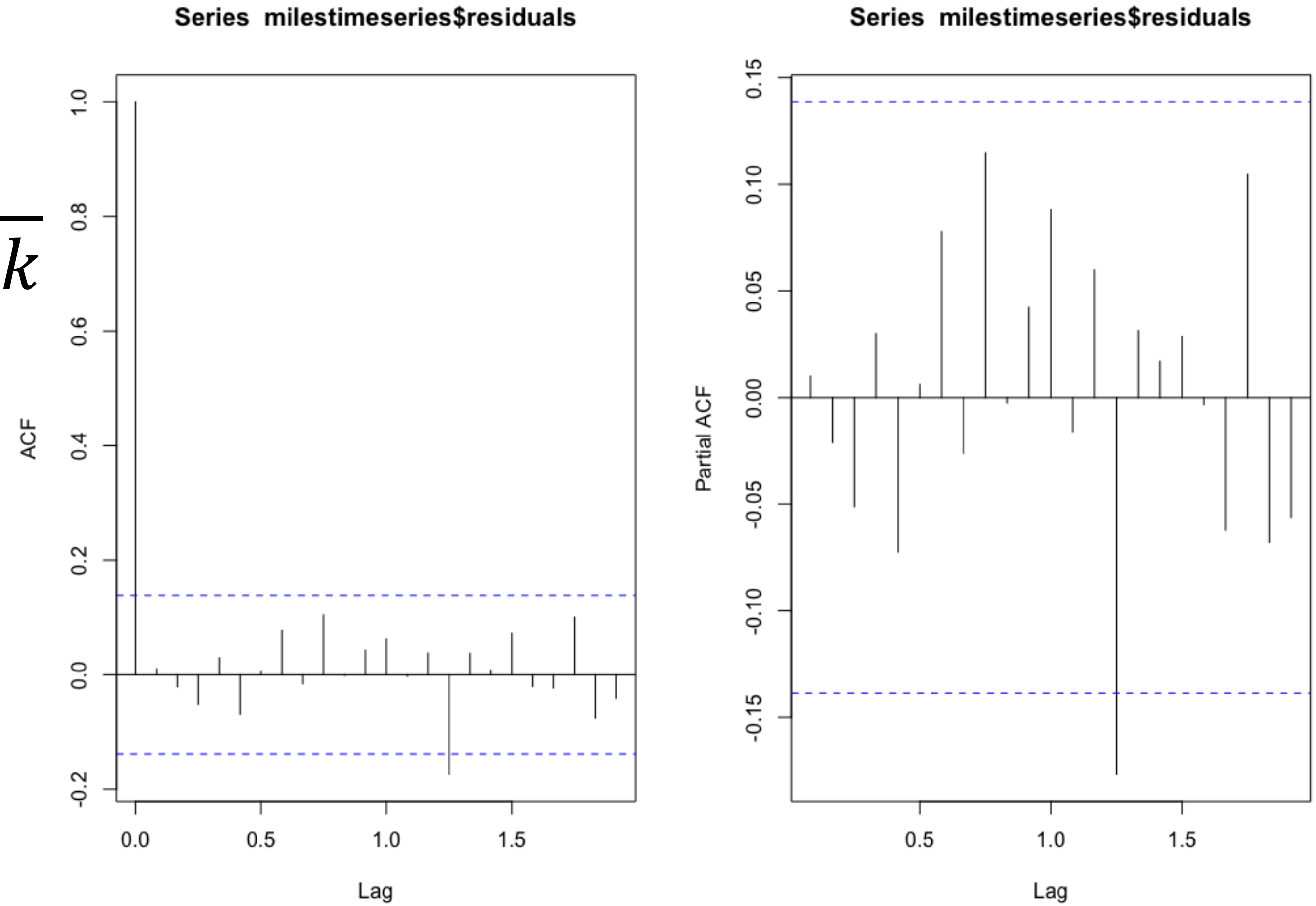
```
Series: milestimeseries
ARIMA(1,0,1)(0,1,1)[12] with drift

Coefficients:
         ar1      ma1     sma1      drift
      0.9078  -0.2093  -0.7266  110280.44
s.e.  0.0364   0.0885   0.0682   31856.26

sigma^2 estimated as 3.901e+12:  log likelihood=-2994.93
AIC=5999.86   AICc=6000.19   BIC=6016.04
```

Time Series Model Building Using ARIMA - RPM Residuals

$$Q^* = n(n+2)\sum_{k=1}^{h}\frac{{r_k}^2}{n-k}$$

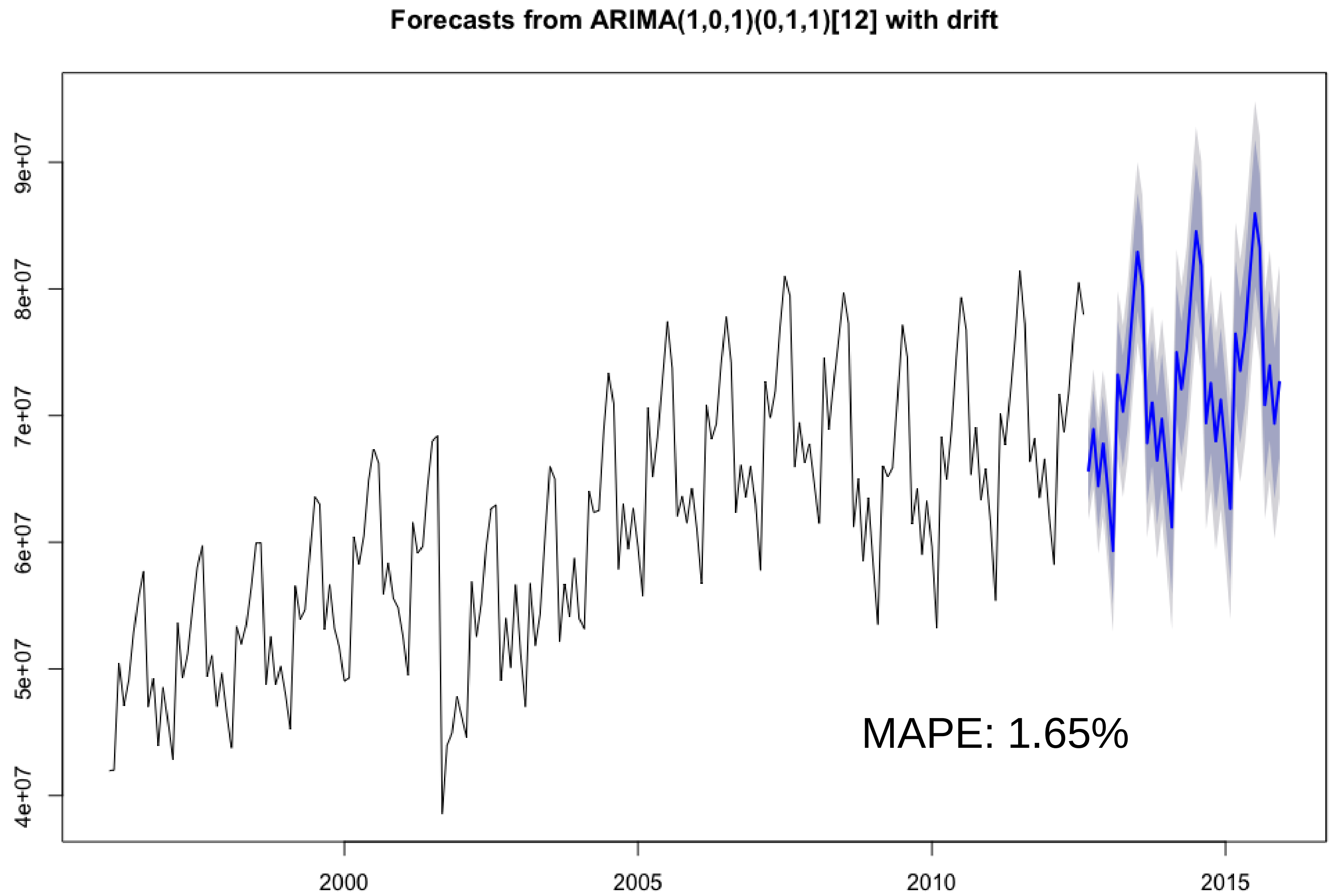
h is the maximum lag being considered
n is the # of observations
r_k is the autocorrelation

If residuals are white noise (purely random), then Q^* has a χ^2 distribution

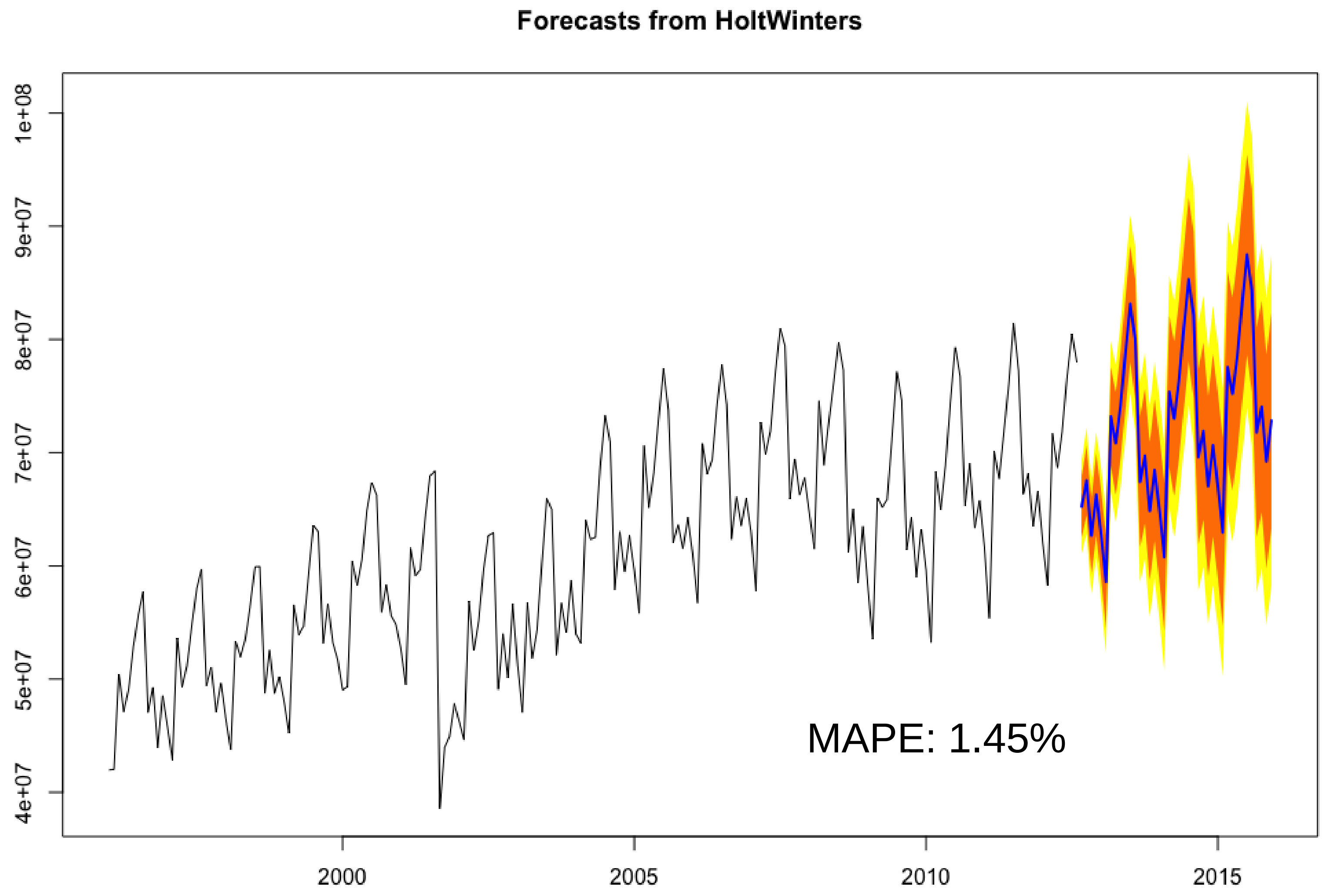



```
        Box-Ljung test

data:  milestimeseries$residuals
X-squared = 15.288, df = 20, p-value = 0.7597
```

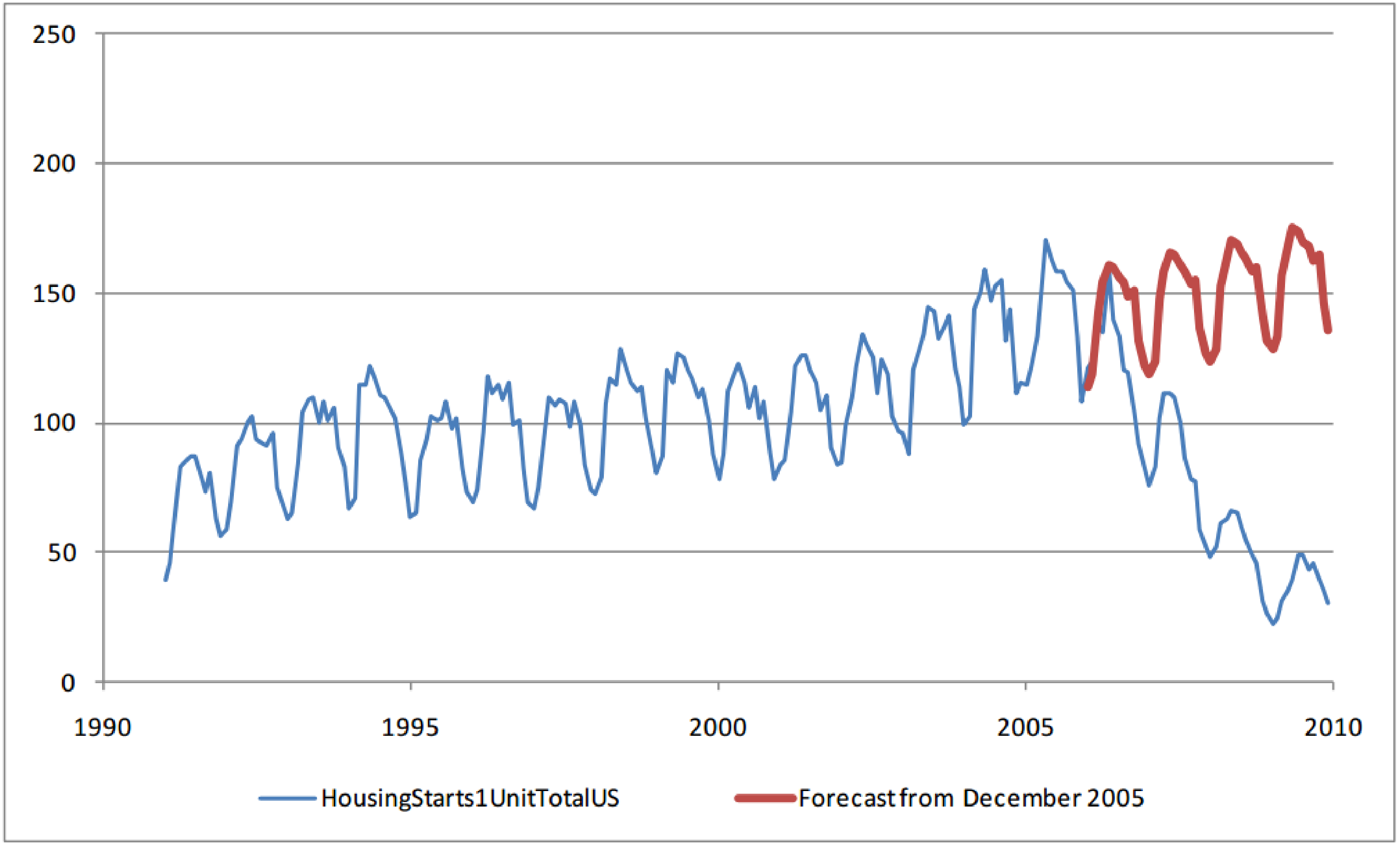
Time Series Model Building Using ARIMA - RPM Forecast



Forecast using Holt-Winters - RPM



Caution: Forecasting is Risky!



"Prediction is very difficult, especially if it's about the future."
--Niels Bohr, Nobel laureate in Physics

Resources

- https://www.otexts.org/fpp

An good open online book on Forecasting methods and practices

- http://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html A short condensed summary oon time-series
- https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/ A short tutorial on using ARIMA models

International School of Engineering

Plot 63/A, Floors 1&2, Road # 13, Film Nagar, Jubilee Hills, Hyderabad - 500 033

For Individuals:	+91-9502334561/63 or 040-65743991
For Corporates:	+91-9618483483
Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering