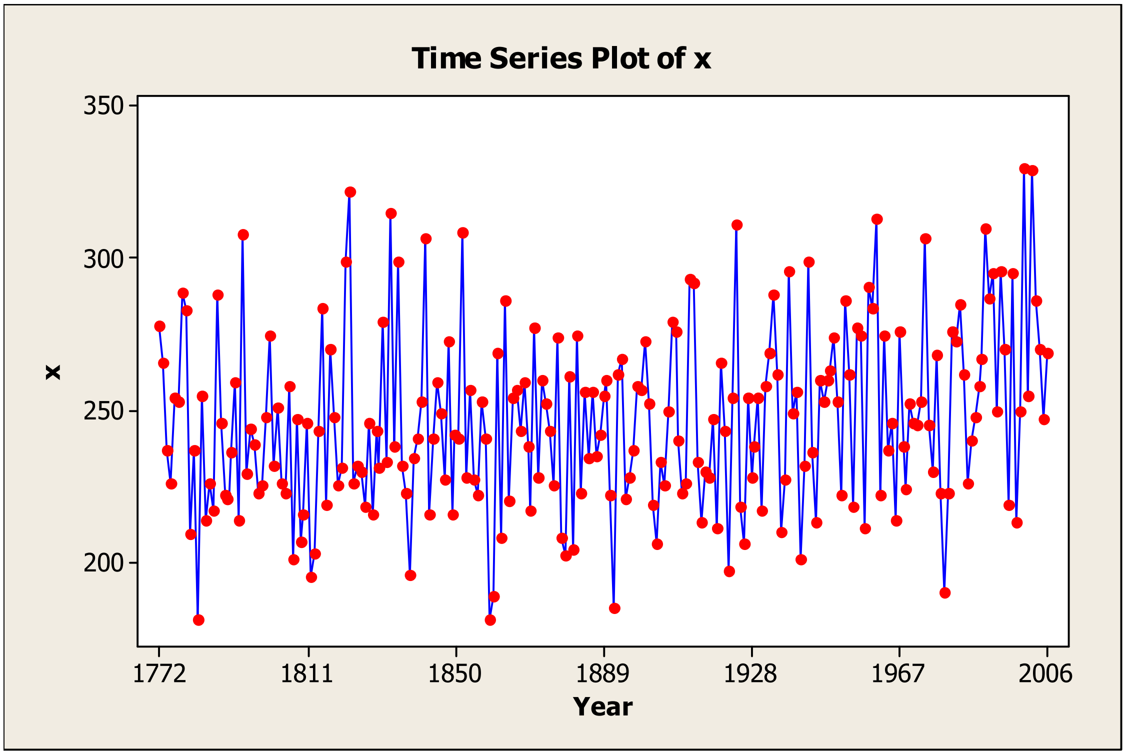
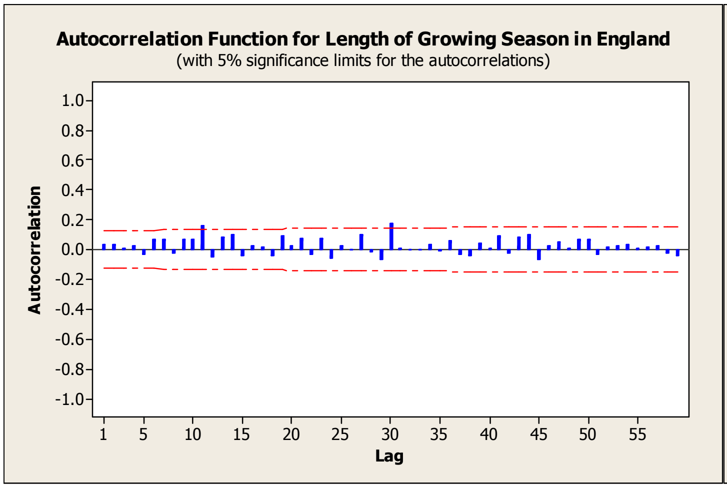
Computer Lab 6

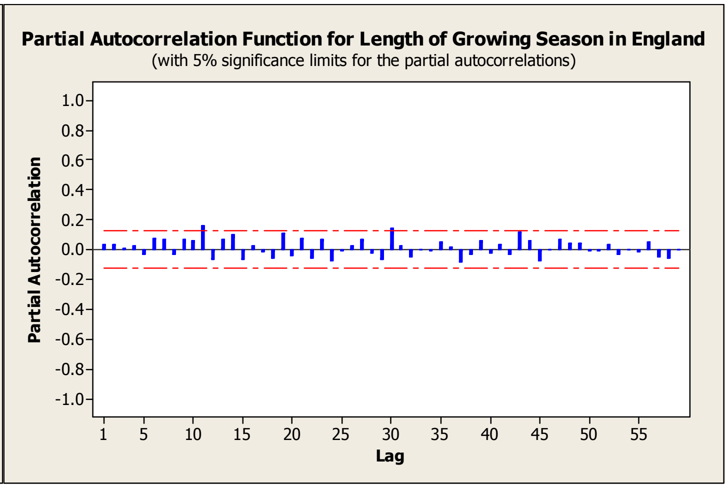
Problem 6.1 Length of Growing Season in England

Minitab Project Report

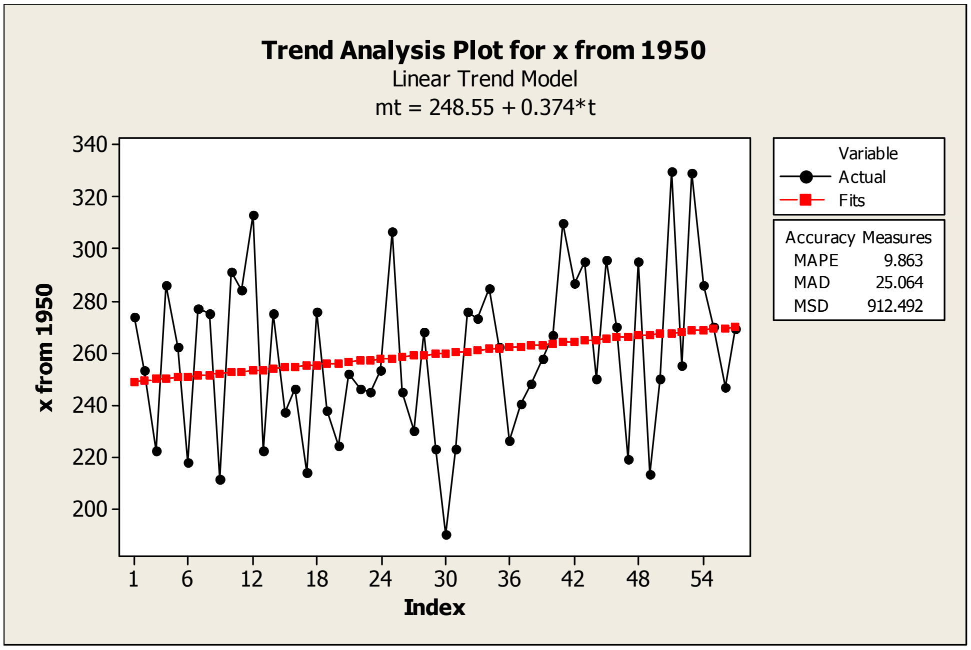


The time series plot indicates a constant trend up to about 1950, then the length of growing season tends to increase. This is not very clear, and the sample ACF and PACF (below) show that the data might actually be a realization of an uncorrelated random variable. It would suggest that the length of growing season in England fluctuates about some constant mean.

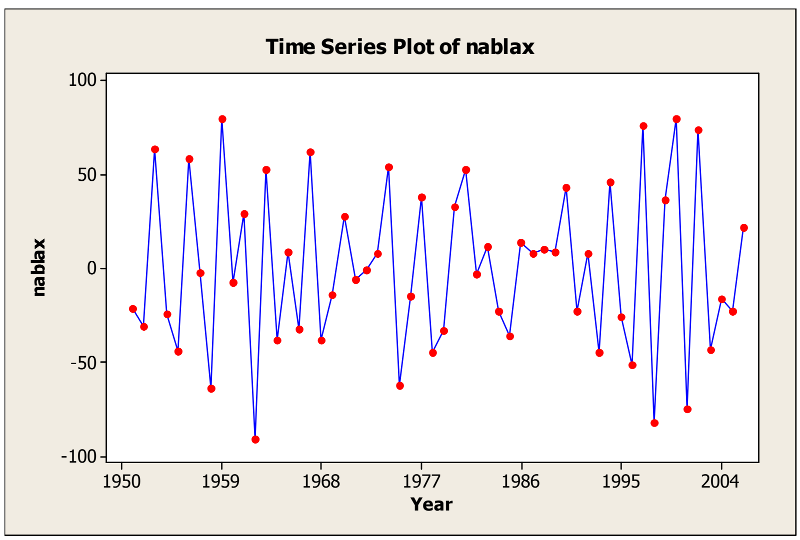


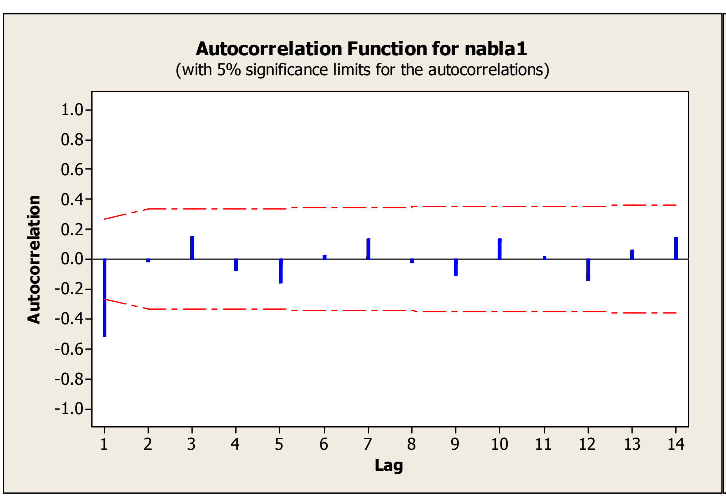


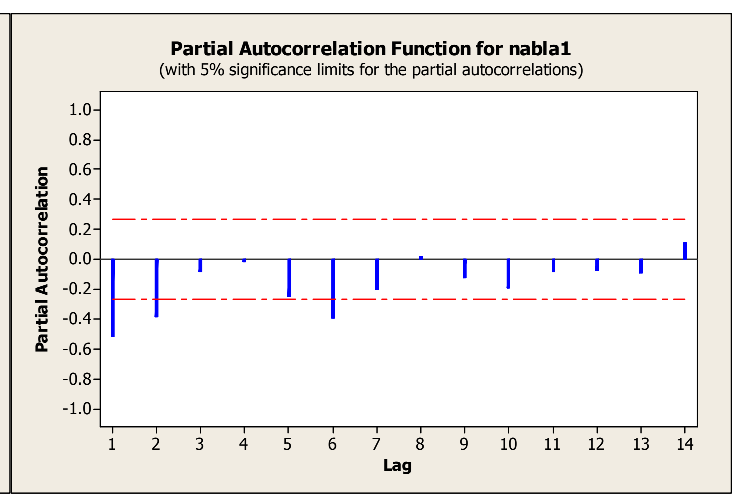
It might be reasonable to look at a part of the data only, for example since 1950, to see if there is any increasing trend. The time series plot below shows the data since 1950 together with a straight line fit.



A slight increase is indicated by the model fit. A linear model fit suggests that first differencing would detrend the data well. To perform ARIMA(p,d,q) modeling of the data we will first examine the sample ACF and PACF of the differenced data to see what are the possible values of the parameters p and q.







Autocorrelation Function: nablax

Lag	ACF	T	LBQ
1	-0.515664	-3.86	15.70
2	-0.013423	-0.08	15.71
3	0.153468	0.93	17.16
4	-0.076749	-0.46	17.53
5	-0.155160	-0.92	19.06
6	0.022500	0.13	19.09
7	0.136700	0.80	20.33
8	-0.029105	-0.17	20.39
9	-0.104977	-0.61	21.15
10	0.136561	0.78	22.47
11	0.019802	0.11	22.49
12	-0.146204	-0.83	24.07
13	0.061455	0.34	24.36
14	0.145800	0.82	26.00

Partial Autocorrelation Function: nablax

Lag	PACF	T
1	-0.515664	-3.86
2	-0.380516	-2.85
3	-0.083316	-0.62
4	-0.018445	-0.14
5	-0.247768	-1.85
6	-0.390811	-2.92
7	-0.204423	-1.53
8	0.017413	0.13
9	-0.128911	-0.96
10	-0.194051	-1.45
11	-0.080749	-0.60
12	-0.076688	-0.57
13	-0.093185	-0.70
14	0.106985	0.80

The sample ACF and PACF suggest MA(1) for the differenced data, hence ARIMA(0,1,1) could be a good choice of the model for the original data since 1950.

Final Estimates of Parameters

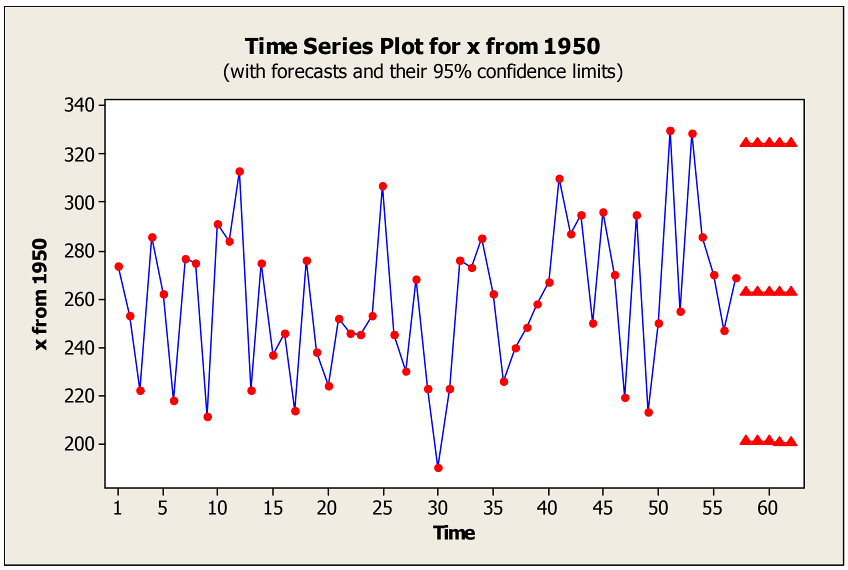
Type		Coef	SE Coef	T	P
MA	1	0.9643	0.0580	16.63	0.000

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	10.1	20.2	24.8	42.7
DF	11	23	35	47
P-Value	0.519	0.629	0.901	0.649

Forecasts from period 57

		95% Limits	
Period	Forecast	Lower	Upper
58	262.642	200.768	324.515
59	262.642	200.729	324.555
60	262.642	200.690	324.594
61	262.642	200.650	324.633
62	262.642	200.611	324.673



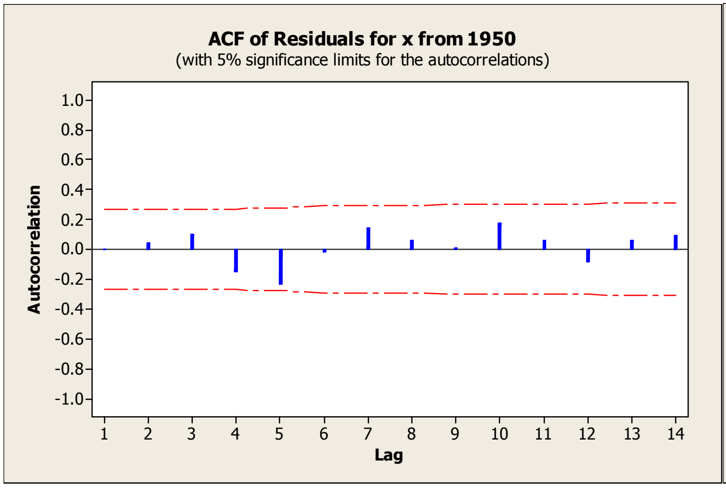
Indeed the MA parameter of the model is highly significant. The forecast of the length of growing season for next five years is constant, equal to about 263 days, with quite large prediction limits of about 201 and 325 days.

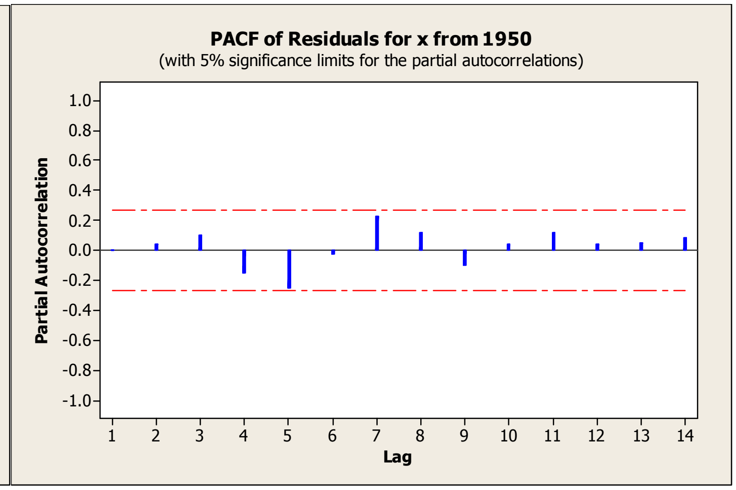
The fitted model can be written as ARIMA(0,1,1) with the MA parameter θ estimated as $\hat{\theta}$ = - 0.96,

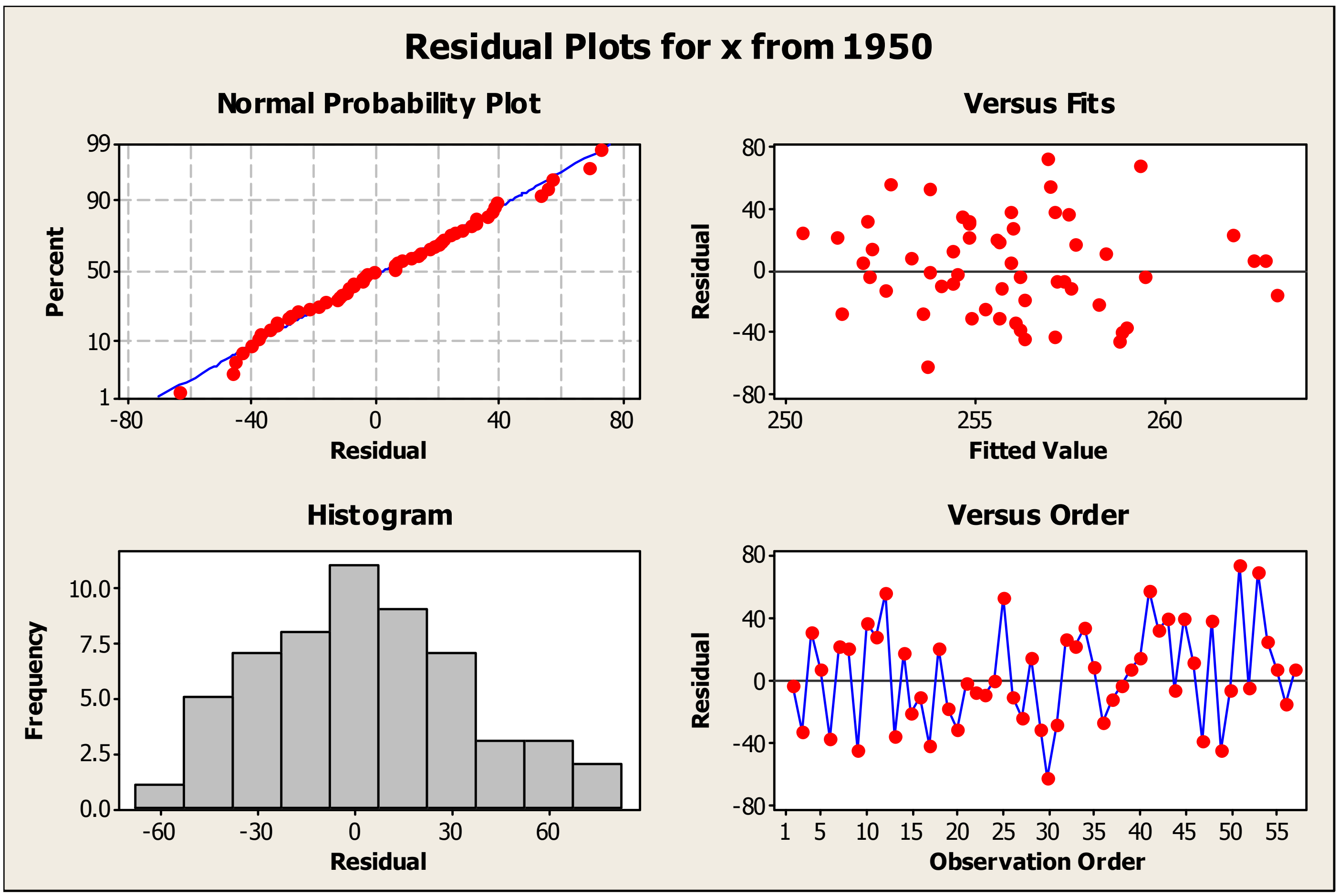
$$\nabla x_t = z_t - 0.96 z_{t-1},$$

where z_t is a realization of White Noise random variable.

Had we considered the data as a realization of an uncorrelated random variable, then the only indication of future values would be the mean of the series, that is, about 247 days.
The residuals indeed show a White Noise characteristics, that is are uncorrelated with zero mean and a constant variance, as can be seen at the diagnostic pictures below.

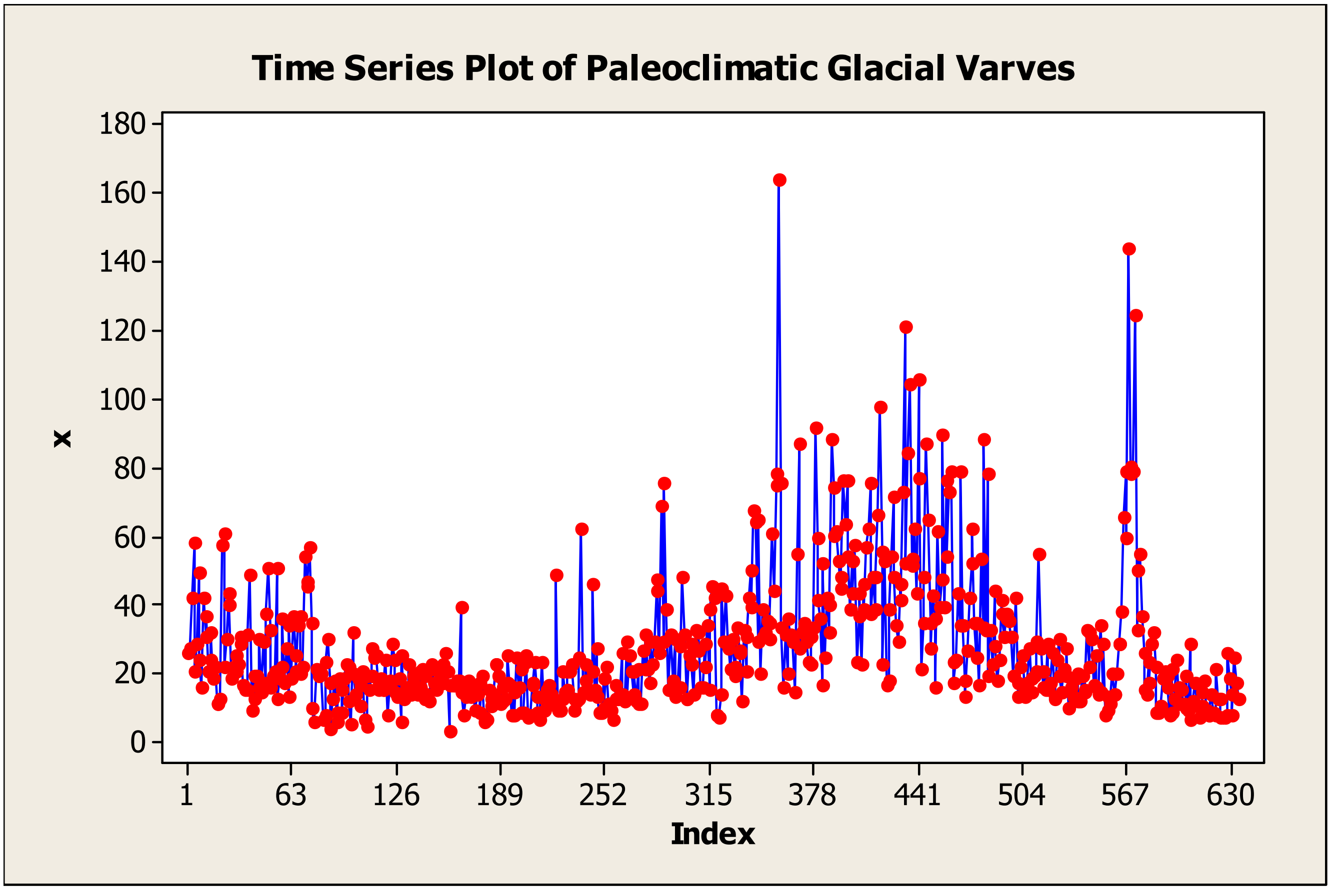

ACF of Residuals for x from 1950
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag


PACF of Residuals for x from 1950
(with 5% significance limits for the partial autocorrelations)
Partial Autocorrelation
Lag

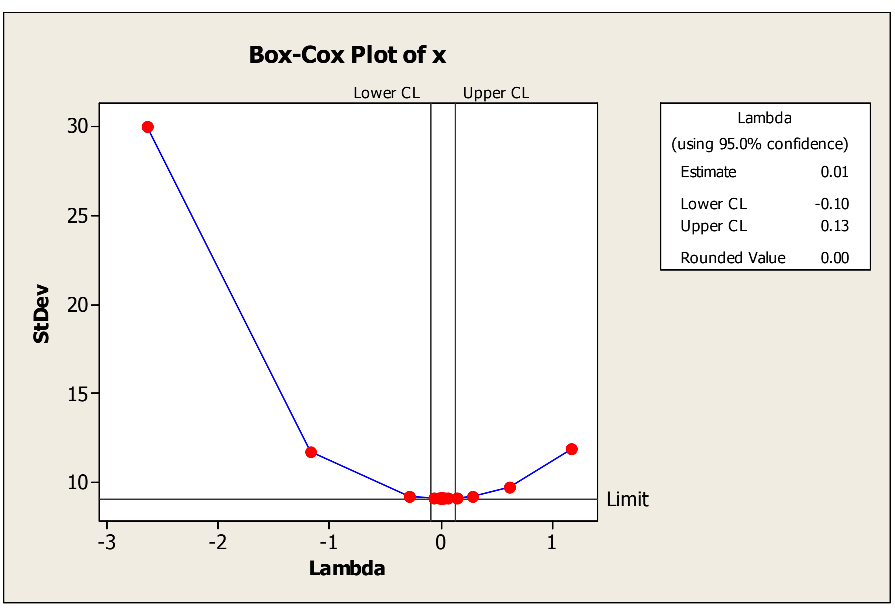

Residual Plots for x from 1950
Normal Probability Plot
Percent
Residual
Versus Fits
Residual
Fitted Value
Histogram
Frequency
Residual
Versus Order
Residual
Observation Order

Problem 6.2 Paleoclimatic Glacial Varves.

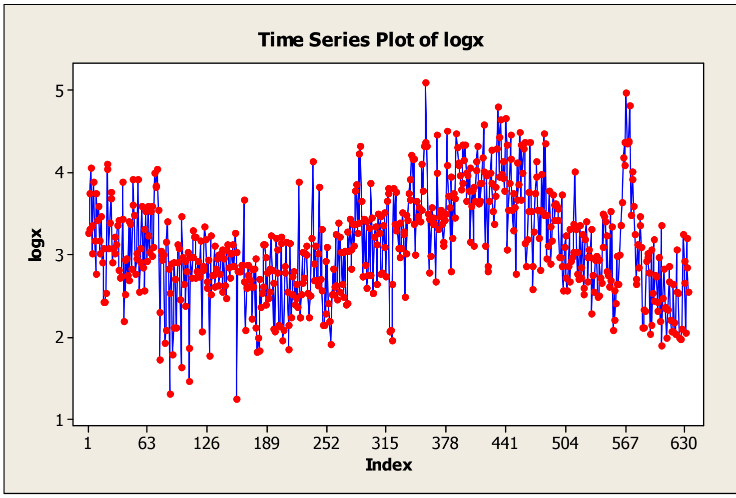
Minitab Project Report

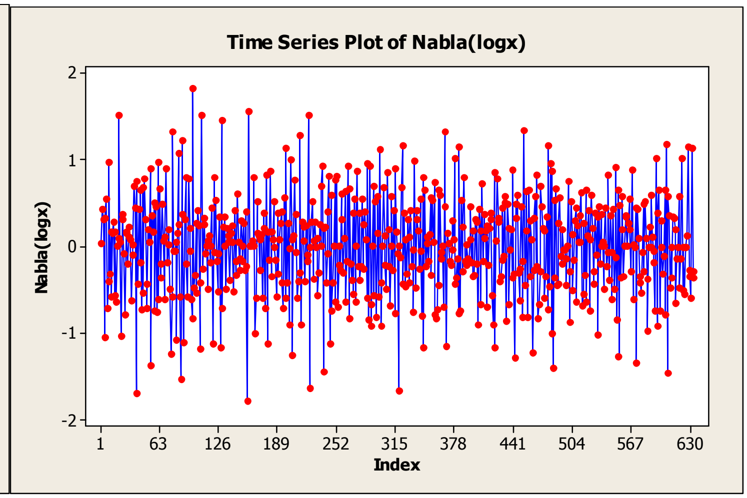


There is no clear increasing or decreasing trend but somewhat wavy pattern can be seen, indicating non-constant mean. There is no seasonality in the data. The time series plot also shows non-constant variance; a transformation is necessary to stabilize it.



The Box-Cox transformation indicates logarithm as the optimal transformation for these data.





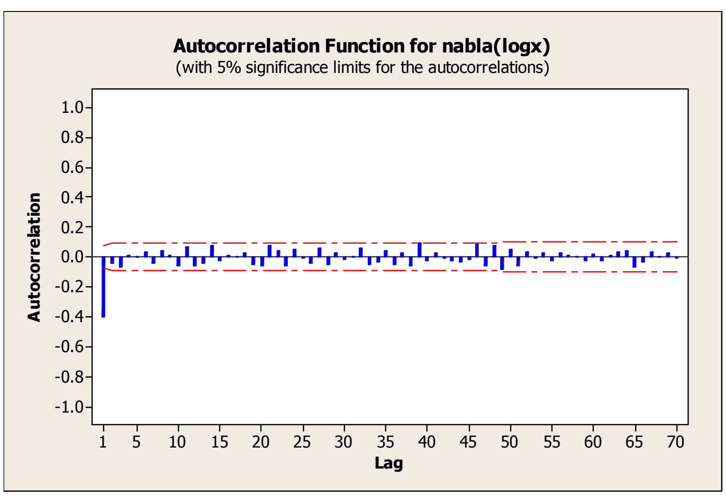
The plot on the left hand side shows the log-transformed series. Its variance is indeed stabilized, the wavy pattern is however maintained. Differencing the transformed data removes the wavy trend as it can be seen in the plot on the right hand side. This suggests d =1 in an ARIMA(p,d,q) model. To find possible values of p and q we will examine the sample ACF and PACF of the transformed and differenced data $\nabla(\log x_t)$.

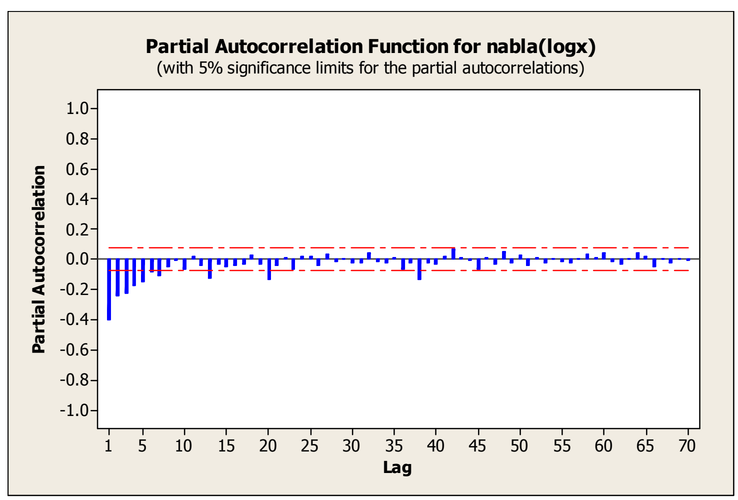
Autocorrelation Function: nabla(logx)

Lag	ACF	T	LBQ
1	-0.397431	-10.00	100.46
2	-0.044481	-0.98	101.72
3	-0.063731	-1.40	104.31
4	0.009204	0.20	104.36
5	-0.002927	-0.06	104.37
6	0.035321	0.77	105.17
7	-0.042932	-0.94	106.35
8	0.040731	0.89	107.42
9	0.009868	0.21	107.48

Partial Autocorrelation Function: nabla(logx)

Lag	PACF	T
1	-0.397431	-10.00
2	-0.240404	-6.05
3	-0.228393	-5.75
4	-0.175778	-4.42
5	-0.148565	-3.74
6	-0.080801	-2.03
7	-0.110836	-2.79
8	-0.047984	-1.21
9	-0.006841	-0.17
10	-0.068347	-1.72





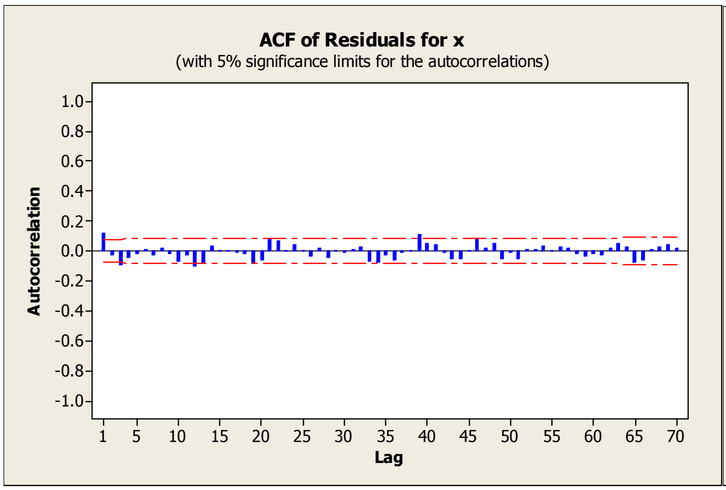
These two functions indicate an MA(1) model with a negative value of θ. Fitting ARIMA(0,1,1) to the transformed data, logx, we obtain the following output.

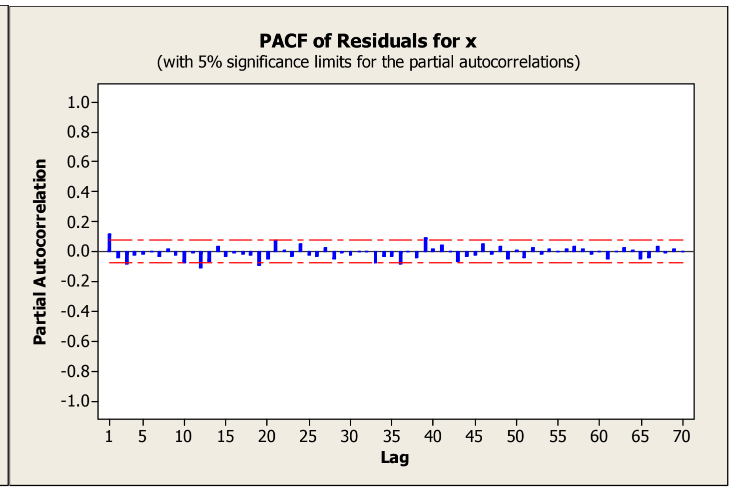
ARIMA Model: logx

```
Final Estimates of Parameters
Type       Coef  SE Coef      T      P
MA   1  0.7727   0.0252  30.61  0.000

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag            12     24     36     48
Chi-Square   27.6   46.4   58.6   79.9
DF             11     23     35     47
P-Value     0.004  0.003  0.007  0.002
```

The MA parameter θ is indeed significant, however, the Ljung-Box statistics show that the ARIMA(0,1,1) does not completely account for the correlations in the data (the p-values are very small and we should reject the hypothesis that correlations of the indicated groups of noise values are non-significant). Indeed, the plots of the sample ACF and PACF below show significant correlation at lag 1.





Hence, it is worth trying another ARIMA model. Adding AR part to the model, that is, fitting ARIMA(1,1,1) improves the residuals and gives both parameters, φ and θ, significant. The new fitted model can be written as

$$(1-0.2348B)\nabla x_t = (1-0.8884B)z_t \qquad \text{or} \quad (1-0.2348B)(1-B)x_t = (1-0.8884B)z_t$$

$$\text{or} \quad x_t - 1.2348x_{t-1} + 0.2348x_{t-2} = z_t - 0.8884z_{t-1}$$

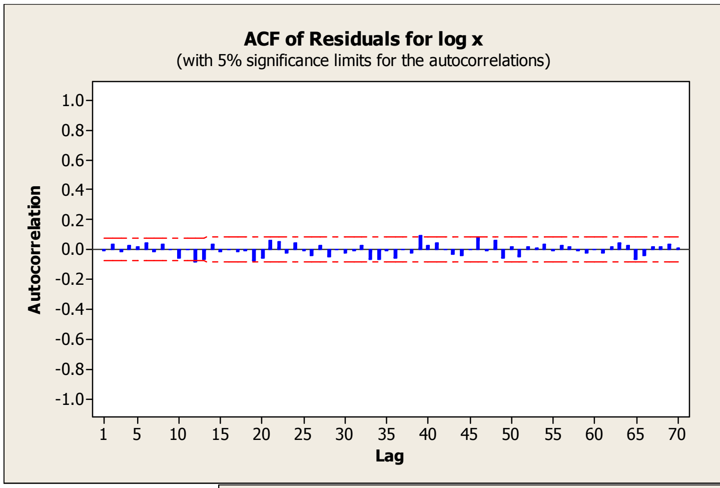
Now, z_t meets the requirements of a white noise variable.

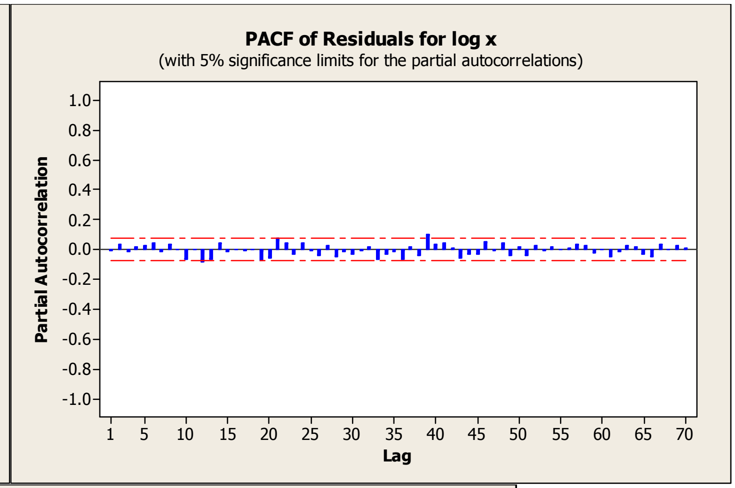
ARIMA Model: logx

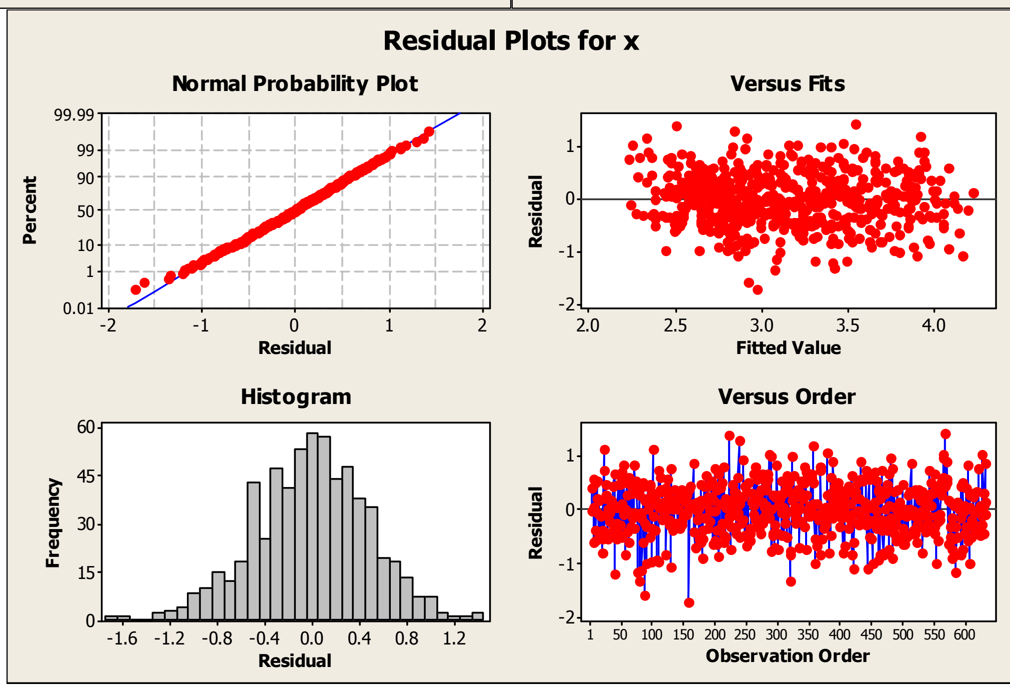
```
Type       Coef  SE Coef      T      P
AR   1  0.2348   0.0462   5.09  0.000
MA   1  0.8884   0.0213  41.62  0.000

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag            12     24     36     48
Chi-Square   10.8   25.9   38.1   55.0
DF             10     22     34     46
P-Value     0.373  0.256  0.290  0.172

Forecasts from period 634
                          95% Limits
Period  Forecast    Lower    Upper  Actual
   635   2.55896  1.62054  3.49737
   636   2.55955  1.56641  3.55268
   637   2.55968  1.55018  3.56919
   638   2.55972  1.53953  3.57990
   639   2.55972  1.53007  3.58938
```







The forecasted values for next five observations slightly increase; the prediction intervals are quite large though.