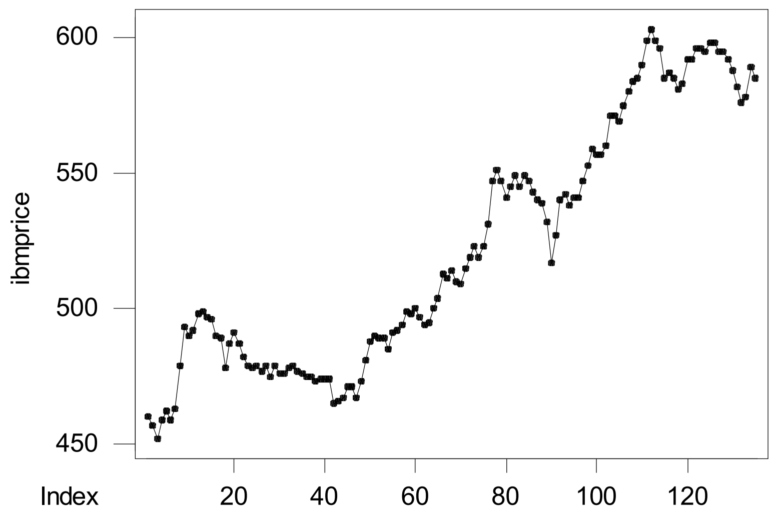
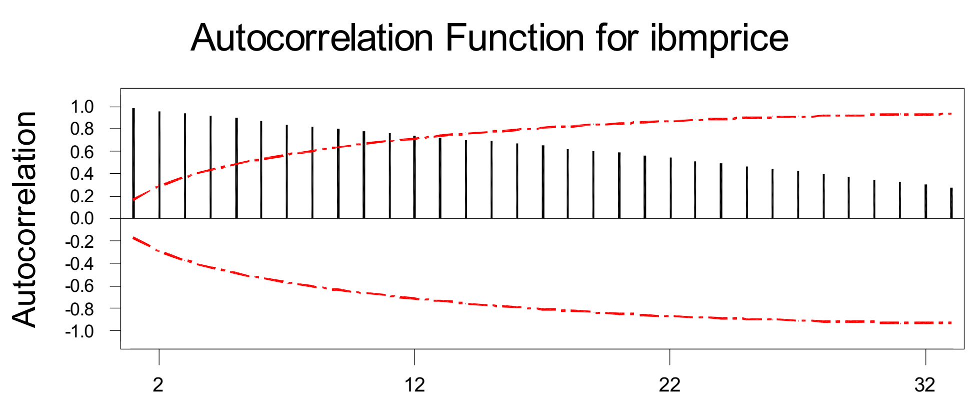
Time Series Analysis and Forecasting Coursework



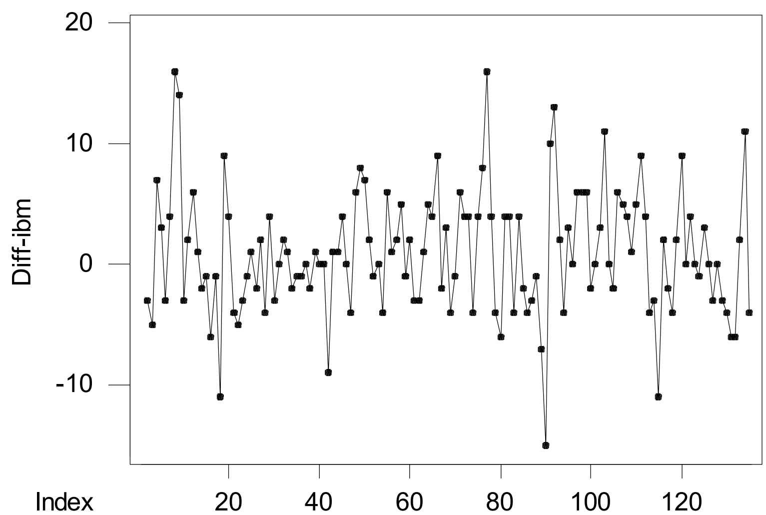
Plot 1. Time series plot of IBM price

The time series plot shows an increasing trend, no seasonal or cyclical components.

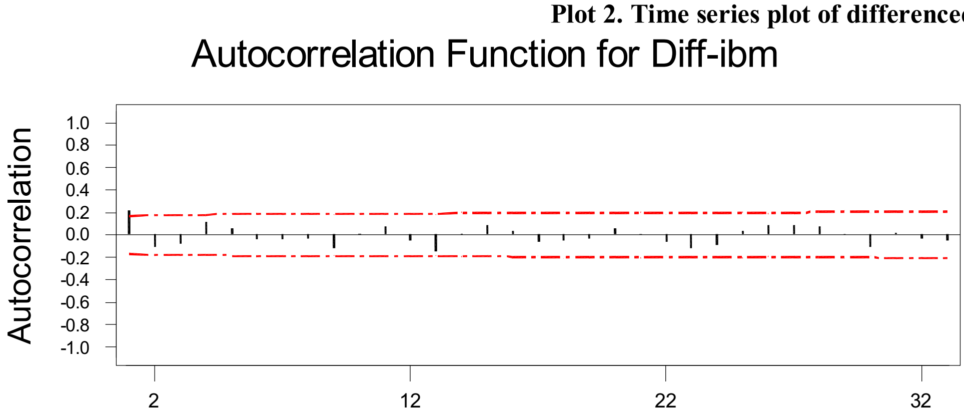


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.98	11.38	132.40	10	0.77	2.31	1095.68	19	0.60	1.43	1737.66	28	0.39	0.85	2099.08
2	0.95	6.49	259.02	11	0.75	2.17	1180.63	20	0.58	1.36	1792.28	29	0.37	0.79	2122.63
3	0.93	4.97	380.56	12	0.74	2.04	1262.00	21	0.56	1.30	1843.56	30	0.34	0.73	2143.32
4	0.91	4.16	497.77	13	0.72	1.93	1339.90	22	0.54	1.22	1890.89	31	0.32	0.68	2161.57
5	0.89	3.62	610.06	14	0.70	1.83	1414.62	23	0.51	1.15	1934.23	32	0.30	0.63	2177.47
6	0.86	3.22	716.72	15	0.68	1.74	1486.26	24	0.49	1.09	1973.72	33	0.27	0.58	2191.16
7	0.84	2.90	817.86	16	0.66	1.66	1554.26	25	0.46	1.02	2009.80				
8	0.81	2.66	914.33	17	0.64	1.58	1618.91	26	0.44	0.97	2042.68				
9	0.79	2.47	1006.93	18	0.62	1.50	1679.93	27	0.42	0.91	2072.50				

The autocorrelation function (acf) has high auto correlations that decrease slowly- giving the shape of a 'thick wedge'- this pattern is indicative of a trend. We cannot use our models because of the presence of the trend; this must be removed by differencing the series with a lag of 1.

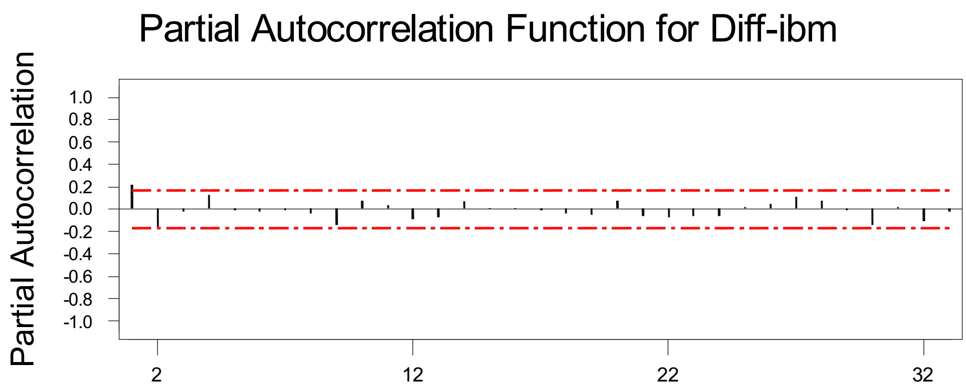


Plot 2. Time series plot of differenced IBM prices



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.21	2.43	6.04	10	0.01	0.10	14.07	19	-0.04	-0.38	21.30	28	0.08	0.76	30.33
2	-0.12	-1.28	7.88	11	0.07	0.77	14.86	20	0.06	0.56	21.80	29	-0.01	-0.08	30.34
3	-0.08	-0.90	8.81	12	-0.05	-0.58	15.31	21	-0.01	-0.09	21.81	30	-0.11	-1.06	32.42
4	0.11	1.23	10.58	13	-0.15	-1.59	18.77	22	-0.07	-0.67	22.52	31	0.01	0.11	32.44
5	0.06	0.63	11.06	14	0.01	0.09	18.78	23	-0.13	-1.29	25.21	32	-0.03	-0.31	32.62
6	-0.04	-0.43	11.29	15	0.09	0.88	19.90	24	-0.09	-0.89	26.52	33	-0.05	-0.48	33.08
7	-0.05	-0.52	11.63	16	0.04	0.36	20.10	25	0.03	0.34	26.71				
8	-0.03	-0.32	11.76	17	-0.06	-0.62	20.68	26	0.09	0.85	27.95				
9	-0.13	-1.34	14.06	18	-0.05	-0.52	21.08	27	0.09	0.87	29.28				

The plot and acf now show no pattern in the differenced series, we now have a stationary series on which to use our models.
The first autocorrelation is significantly different from zero (T=2.43>2) and all the other autocorrelations lie within the confidence limits so this suggests that MA (1) model can be used for the differenced series.



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.21	2.43	10	0.08	0.88	19	-0.05	-0.58	28	0.08	0.89
2	-0.17	-1.93	11	0.03	0.37	20	0.08	0.88	29	-0.02	-0.20
3	-0.02	-0.23	12	-0.10	-1.14	21	-0.07	-0.78	30	-0.14	-1.67
4	0.13	1.47	13	-0.07	-0.87	22	-0.07	-0.85	31	0.02	0.20
5	-0.01	-0.17	14	0.06	0.73	23	-0.07	-0.77	32	-0.12	-1.36
6	-0.03	-0.31	15	0.00	0.04	24	-0.06	-0.75	33	-0.03	-0.29
7	-0.01	-0.12	16	0.01	0.09	25	0.02	0.19			
8	-0.04	-0.46	17	-0.02	-0.21	26	0.05	0.55			
9	-0.14	-1.62	18	-0.05	-0.54	27	0.10	1.14			

Only the first partial autocorrelation (pac) is significantly different from zero (T=2.43>2) suggesting the use of an AR (1) model. However, the second pac is close to being significant with a T value of -1.93. This suggests it might be worth considering an AR (2) model.

Overall conclusion

It would be best to go for MA (1) as there is no confusion like in the AR model.

Now we will look at the ARIMA command to identify the most suitable model to apply to the time series.

ARIMA Model: Diff-ibm (MA (1) model with constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type           Coef     SE Coef        T       P
MA   1      -0.2685      0.0851    -3.16   0.002
Constant     0.9140      0.5501     1.66   0.099
Mean         0.9140      0.5501

Number of observations:  134
Residuals:     SS =  3327.03  (backforecasts excluded)
               MS =   25.20  DF = 132

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square       8.9      17.6      24.8      32.5
```

DF	10	22	34	46
P-Value	0.540	0.727	0.875	0.933

Firstly we examine the parameter estimates to see if they are significantly different from zero. If P>0.05, the parameter is not significantly different from zero and may therefore be excluded from the model.
Examining the MA (1) model we see that the constant term is not significantly different from zero (T=1.66, p=0.099>0.05). We can therefore remove the constant term.

ARIMA Model: Diff-ibm (MA (1) model without constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
MA   1     -0.2853      0.0842     -3.39    0.001

Number of observations:  134
Residuals:    SS =  3395.98  (backforecasts excluded)
              MS =    25.53  DF = 133

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square       8.9      17.4      24.6      32.5
DF                11        23        35        47
P-Value        0.634     0.787     0.905     0.946
```

Examining the MA (1) model without a constant, we see that the parameter is significantly different from zero (T=-3.39, p=0.001). The Box-Pierce statistic is not significant (chi-square (χ^2) =17.4, p=0.787>0.05).
So the model provides an adequate fit. Now we see what happens when a second MA term is added.

ARIMA Model: Diff-ibm (MA(2) without constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
MA   1     -0.2691      0.0875     -3.07    0.003
MA   2      0.0325      0.0885      0.37    0.714

Number of observations:  134
Residuals:    SS =  3392.54  (backforecasts excluded)
              MS =    25.70  DF = 132

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square       8.1      16.6      23.8      31.9
DF                10        22        34        46
P-Value        0.618     0.784     0.904     0.943
```

The second MA parameter is not significantly different from zero (T=0.37, p=0.714>0.05) and so the parameter is not required in the model. We see that the Box-Pierce statistic remains fairly similar at 16.6 so there is no improvement in fit obtained by using an MA(2) model.

Conclusion
We will therefore use an MA(1) model without a constant. This conclusion is consistent with that reached by looking at the acf.

Now we will look at the AR model:
ARIMA Model: Diff-ibm (AR (1) model with constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1      0.2124      0.0854      2.49    0.014
Constant    0.7207      0.4369      1.65    0.101
Mean        0.9151      0.5547

Number of observations:  134
Residuals:    SS =  3376.16  (backforecasts excluded)
              MS =   25.58  DF = 132

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag              12        24        36        48
Chi-Square      11.9      21.7      29.0      35.7
DF               10        22        34        46
P-Value       0.291     0.477     0.712     0.864
```

Examining the AR (1) model we see that the constant term is not significantly different from zero (T=1.65, p=0.101>0.05). We can therefore remove the constant term.

ARIMA Model: Diff-ibm (AR(1) without constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1      0.2388      0.0844      2.83    0.005

Number of observations:  134
Residuals:    SS =  3443.69  (backforecasts excluded)
              MS =   25.89  DF = 133

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag              12        24        36        48
Chi-Square      12.1      21.8      29.0      35.8
DF               11        23        35        47
P-Value       0.355     0.533     0.752     0.882
```

Examining the AR (1) model without a constant, we see that the parameter is significantly different from zero (T=2.83, p=0.005). The Box-Pierce statistic is not significant (χ^2 =21.8, p=0.533>0.05)
So the model provides an adequate fit. Now we see what happens when a second AR term is added.

ARIMA Model: Diff-ibm (AR(2) without constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1      0.2769      0.0866      3.20    0.002
AR   2     -0.1543      0.0881     -1.75    0.082

Number of observations:  134
Residuals:    SS =  3366.41  (backforecasts excluded)
              MS =    25.50  DF = 132

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square       6.3      13.8      21.0      29.8
DF                10        22        34        46
P-Value        0.787     0.909     0.960     0.969
```

The second AR parameter is not significantly different from zero (T=-1.75, p=0.082>0.05) although this is close. In addition the Box-Pierce statistic shows that χ^2 has dropped to 13.8 (p=0.909) suggesting an improved fit. We could therefore consider the AR(2) model. This result is consistent to that reached with the pacf.
Now we see what happens when a third AR term is added.

ARIMA Model: Diff-ibm (AR(3) without constant)

```
ARIMA model for Diff-ibm


Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1      0.2785      0.0880      3.17    0.002
AR   2     -0.1576      0.0922     -1.71    0.090
AR   3      0.0112      0.0895      0.13    0.900

Number of observations:  134
Residuals:    SS =  3366.05  (backforecasts excluded)
              MS =    25.70  DF = 131

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square       6.3      13.8      21.1      30.0
DF                 9        21        33        45
P-Value        0.706     0.876     0.945     0.958
```

There is no point in adding 3rd term as the 3rd parameter is not significant and the χ^2 value is the same. Therefore we could use AR(1) or AR(2).

Overall conclusion

MA(1) model preferred on grounds of clarity and parsimony but we may want to consider AR(2) as it gives a better fit. Overall go for MA(1) so the fit is not that much better.

ARIMA Model: Diff-ibm (MA (1) model without constant)

```
ARIMA model for Diff-ibm

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
MA   1     -0.2853      0.0842     -3.39    0.001

Number of observations:  134
Residuals:    SS =  3395.98  (backforecasts excluded)
              MS =   25.53  DF = 133

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12         24         36         48
Chi-Square       8.9       17.4       24.6       32.5
DF                11         23         35         47
P-Value        0.634      0.787      0.905      0.946

Forecasts from period 135
                             95 Percent Limits
Period      Forecast         Lower        Upper       Actual
 136         -1.9599      -11.8659       7.9462
 137          0.0000      -10.3013      10.3013
```

<u>We need to forecast X_{136} and X_{137}.</u>

Let Xt = IBM price on day t

Yt = Xt – Xt-1 (series of first differences)

From ARIMA command for MA(1) model without constant this model is estimated by

Yt = Zt + 0.2853Zt-1

We need to forecast X_{136} and X_{137}.

From forecasts above $\underline{\boldsymbol{Y_{136}}}$ **= -1.96** and $\underline{\boldsymbol{Y_{137}}}$ **= 0**

To complete the forecasting of next 2 observations we must connect the X's and Y's.

Yt = Xt – Xt-1
Xt = Yt + Xt-1

$X_{136} = Y_{136} + X_{135}$

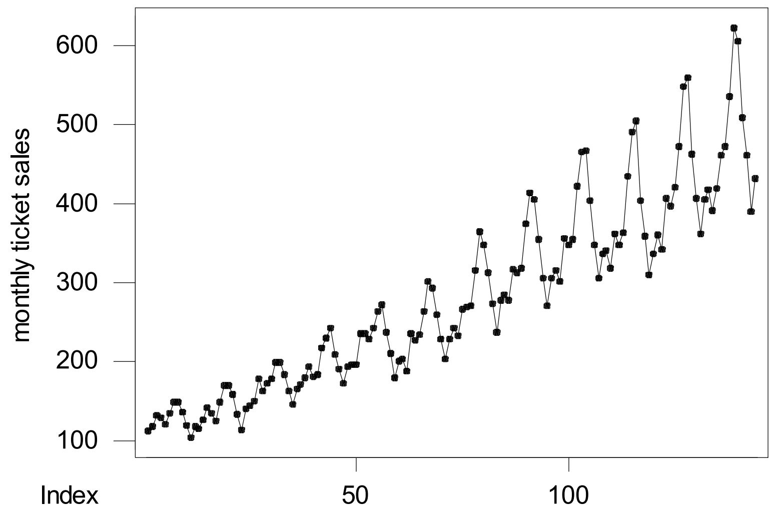
$\underline{\boldsymbol{X_{136}}} = \underline{\boldsymbol{Y_{136}}} + X_{135}$ (we know $\underline{\boldsymbol{Y_{136}}}$ from forecast results and X_{135} from the data)
= -1.96 + 585
= **583.04**

Finally, $X_{137} = \underline{\boldsymbol{Y_{137}}} + \underline{\boldsymbol{X_{136}}}$ (we know $\underline{\boldsymbol{Y_{137}}}$ from forecast results and $\underline{\boldsymbol{X_{136}}}$ was just forecasted)

$\underline{\boldsymbol{X_{137}}} = 0 + 583.04$

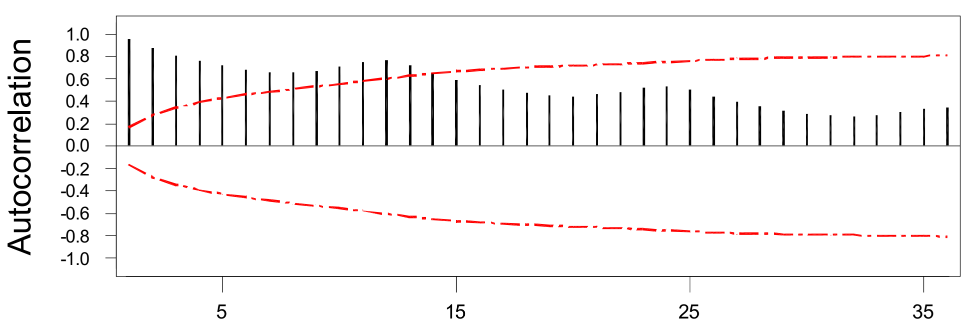
= **583.04**

2a



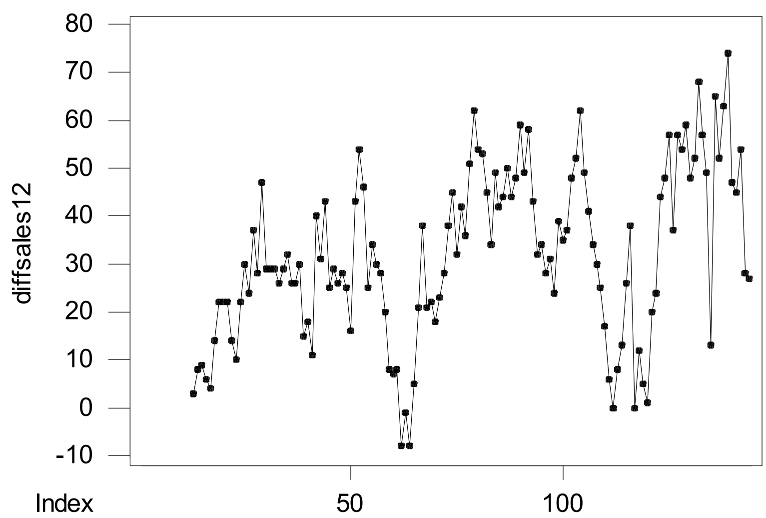
Plot 3.Time series plot of monthly ticket sales

The plot shows an increasing trend and peaks at regular intervals indicating a seasonal or cyclical component in the data. Increase in variation.



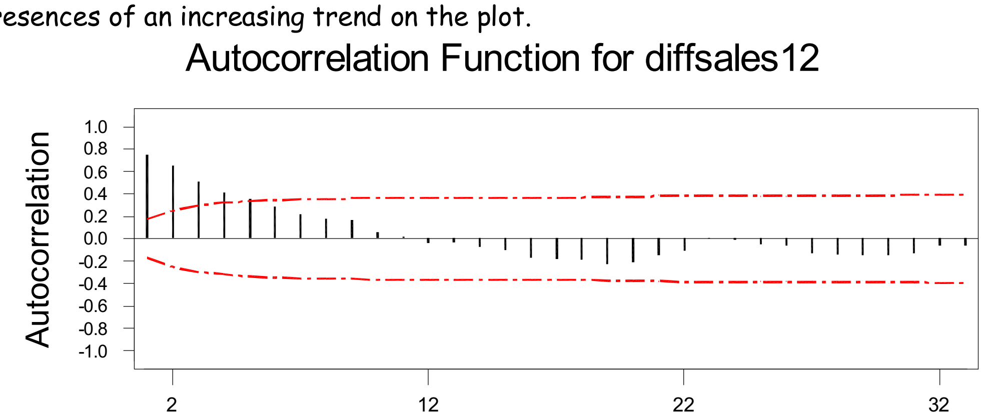
Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.95	11.38	132.13	10	0.70	2.51	852.90	19	0.45	1.24	1394.93	28	0.35	0.88	1725.14
2	0.88	6.28	245.49	11	0.74	2.55	940.27	20	0.44	1.21	1427.54	29	0.31	0.79	1743.09
3	0.80	4.64	342.03	12	0.76	2.49	1032.31	21	0.46	1.24	1462.95	30	0.29	0.72	1758.16
4	0.75	3.80	426.51	13	0.71	2.24	1113.85	22	0.48	1.30	1503.06	31	0.27	0.67	1771.46
5	0.71	3.28	502.92	14	0.65	1.97	1181.41	23	0.52	1.38	1549.61	32	0.26	0.65	1784.20
6	0.68	2.92	573.07	15	0.58	1.73	1237.05	24	0.53	1.40	1599.22	33	0.27	0.68	1798.52
7	0.66	2.69	639.90	16	0.54	1.56	1284.21	25	0.49	1.28	1642.38	34	0.30	0.74	1815.52
8	0.65	2.53	705.77	17	0.50	1.42	1325.17	26	0.44	1.12	1676.54	35	0.33	0.80	1836.06
9	0.67	2.49	775.48	18	0.47	1.31	1361.37	27	0.39	0.98	1703.38	36	0.34	0.83	1858.19

Here we see the observations 12 months apart are correlated, this indicates a seasonal component in the series. Because of the presence of the seasonal component we must difference the series with a lag of 12 to remove the seasonal component.



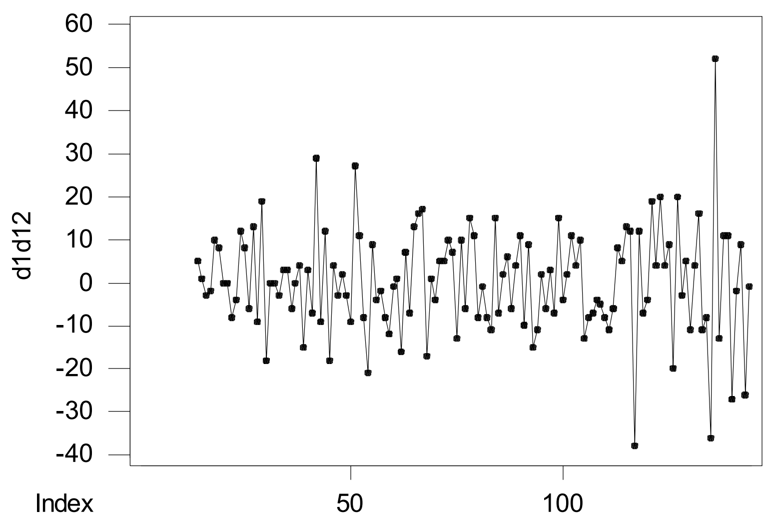
Plot 4. Time series plot of differenced ticket sales

Having removed the seasonal component the difference series still shows signs of the presences of an increasing trend on the plot.

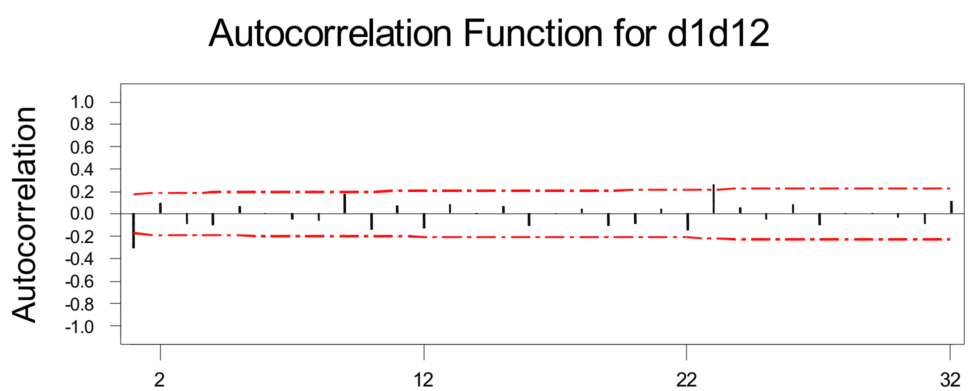


Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.75	8.58	75.24	10	0.06	0.31	234.16	19	-0.23	-1.25	261.44	28	-0.14	-0.73	282.59
2	0.65	5.11	132.21	11	0.02	0.10	234.21	20	-0.22	-1.14	268.85	29	-0.16	-0.80	286.74
3	0.50	3.38	167.16	12	-0.04	-0.24	234.49	21	-0.16	-0.82	272.80	30	-0.15	-0.76	290.64
4	0.41	2.51	189.99	13	-0.03	-0.19	234.67	22	-0.12	-0.61	275.00	31	-0.13	-0.68	293.74
5	0.35	2.09	207.52	14	-0.07	-0.41	235.51	23	-0.00	-0.00	275.01	32	-0.06	-0.32	294.44
6	0.28	1.62	218.79	15	-0.11	-0.58	237.23	24	-0.01	-0.07	275.04	33	-0.06	-0.32	295.15
7	0.22	1.21	225.41	16	-0.18	-0.96	241.95	25	-0.05	-0.28	275.54				
8	0.18	0.97	229.78	17	-0.18	-0.97	246.91	26	-0.07	-0.35	276.31				
9	0.16	0.91	233.68	18	-0.20	-1.05	252.82	27	-0.13	-0.67	279.16				

Also the acf of the differenced series exhibits a 'thick wedge' pattern, indicative of a trend. Now we must remove the trend from the differenced series, by differencing that series with a lag of 1.

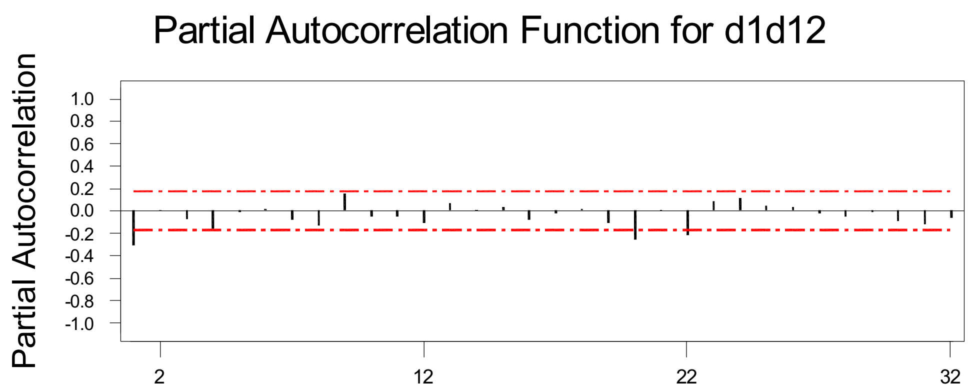


Plot 5. Time series plot of doubly differenced ticket sales



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	-0.31	-3.55	12.86	10	-0.14	-1.39	25.45	19	-0.11	-1.08	34.66	28	0.00	0.01	54.63
2	0.10	1.00	14.09	11	0.07	0.68	26.16	20	-0.09	-0.85	35.97	29	0.00	0.00	54.63
3	-0.10	-1.01	15.37	12	-0.13	-1.30	28.77	21	0.04	0.39	36.25	30	-0.04	-0.34	54.90
4	-0.10	-1.02	16.71	13	0.09	0.84	29.90	22	-0.16	-1.46	40.20	31	-0.09	-0.81	56.41
5	0.06	0.62	17.23	14	0.00	0.02	29.90	23	0.26	2.36	50.91	32	0.12	1.00	58.77
6	-0.00	-0.00	17.23	15	0.07	0.62	30.54	24	0.05	0.46	51.36				
7	-0.06	-0.57	17.67	16	-0.11	-1.04	32.34	25	-0.05	-0.45	51.79				
8	-0.06	-0.62	18.20	17	-0.00	-0.00	32.34	26	0.08	0.71	52.89				
9	0.18	1.79	22.62	18	0.04	0.42	32.64	27	-0.10	-0.89	54.63				

We see from the plot and acf of the final series that both the trend and seasonal components have been removed. We now have a stationary series on which to use our models.
The first autocorrelation is significantly different from zero (T=-3.55<-2) suggesting the use of an MA (1) model. However, the 23rd acf is also significant. This suggests that an MA model may not be suitable but we might get away with using a MA (1) model.



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	-0.31	-3.55	10	-0.06	-0.67	19	-0.11	-1.30	28	-0.05	-0.63
2	-0.00	-0.01	11	-0.05	-0.60	20	-0.26	-2.94	29	-0.01	-0.14
3	-0.07	-0.86	12	-0.12	-1.32	21	0.01	0.08	30	-0.09	-1.03
4	-0.17	-1.91	13	0.06	0.69	22	-0.22	-2.57	31	-0.12	-1.40
5	-0.02	-0.17	14	0.00	0.05	23	0.08	0.97	32	-0.07	-0.78
6	0.02	0.21	15	0.04	0.42	24	0.12	1.34			
7	-0.09	-1.01	16	-0.09	-0.99	25	0.04	0.46			
8	-0.13	-1.53	17	-0.03	-0.32	26	0.03	0.34			
9	0.16	1.79	18	0.02	0.22	27	-0.02	-0.23			

Only the first partial autocorrelation (pac) is significantly different from zero (T=-3.55<-2) suggesting the use of an AR (1) model. However, the 20th and 22nd pac are also significant with a T values of –2.94 and –2.57 respectively. This suggests that an AR model may not be suitable.

Overall conclusion

It would be best to go for MA (1) as there is more confusion in the AR model.

Now we will look at the ARIMA command to identify the most suitable model to apply to the time series.

ARIMA Model: d1d12 (MA(1) with constant)

```
ARIMA model for d1d12

Final Estimates of Parameters
Type           Coef      SE Coef         T        P
MA   1       0.3223       0.0840      3.84    0.000
Constant     0.1937       0.6991      0.28    0.782
Mean         0.1937       0.6991
```

```
Number of observations:  131
Residuals:    SS =  17968.6  (backforecasts excluded)
              MS =    139.3  DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square      11.9      38.7      46.1      62.2
DF                10        22        34        46
P-Value        0.289     0.015     0.080     0.056
```

Examining the MA (1) model we see that the constant term is not significantly different from zero (T=0.28, p=0.782>0.05). We can therefore remove the constant term.

ARIMA Model: d1d12 (MA(1) without constant)

```
ARIMA model for d1d12

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
MA   1      0.3212      0.0837      3.84    0.000

Number of observations:  131
Residuals:    SS =  17978.9  (backforecasts excluded)
              MS =    138.3  DF = 130

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square      12.0      38.6      46.1      62.1
DF                11        23        35        47
P-Value        0.367     0.022     0.099     0.069
```

Examining the MA (1) model without a constant, we see that the parameter is significantly different from zero (T=3.84, p<0.0005). The Box-Pierce statistic is significant (χ^2=38.6, p=0.022<0.05).
So the model provides an adequate fit. Now we see what happens when a second MA term is added.

ARIMA Model: d1d12 (MA(2) without constant)

```
ARIMA model for d1d12

Unable to reduce sum of squares any further

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
MA   1      0.3205      0.0882      3.63    0.000
MA   2     -0.0094      0.0895     -0.11    0.916

Number of observations:  131
Residuals:    SS =  17977.7  (backforecasts excluded)
              MS =    139.4  DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square      11.9      38.6      46.0      61.9
DF                10        22        34        46
P-Value        0.290     0.016     0.082     0.059
```

The second MA parameter is not significantly different from zero (T=-0.11, p=0.916>0.05) and so the parameter is not required in the model. We see that the Box-Pierce statistic remains the same at 38.6 so there is no improvement in fit obtained by using an MA(2) model.

Conclusion

We will therefore use an MA(1) model without a constant. This conclusion is consistent with that reached by looking at the acf.

ARIMA Model: d1d12 (AR(1) with constant)

```
ARIMA model for d1d12

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1     -0.3102      0.0837     -3.71    0.000
Constant     0.231       1.030      0.22    0.823
Mean        0.1766      0.7864

Number of observations:  131
Residuals:     SS =  17939.9  (backforecasts excluded)
               MS =    139.1  DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12         24         36         48
Chi-Square      11.7       38.3       45.0       60.1
DF                10         22         34         46
P-Value        0.304      0.017      0.098      0.080
```

Examining the AR (1) model we see that the constant term is not significantly different from zero (T=0.22, p=0.823>0.05). We can therefore remove the constant term.

ARIMA Model: d1d12 (AR(1) without constant)

```
ARIMA model for d1d12

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1     -0.3099      0.0834     -3.72    0.000

Number of observations:  131
Residuals:     SS =  17946.8  (backforecasts excluded)
               MS =    138.1  DF = 130

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12         24         36         48
Chi-Square      11.7       38.3       45.0       60.1
DF                11         23         35         47
P-Value        0.385      0.024      0.120      0.096
```

Examining the AR (1) model without a constant, we see that the parameter is significantly different from zero (T=-3.72, p<0.0005). The Box-Pierce statistic is significant (χ^2=38.3, p=0.024<0.05).

So the model provides an adequate fit. Now we see what happens when a second AR term is added.

ARIMA Model: d1d12 (AR(2) without constant)

```
ARIMA model for d1d12

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
AR   1     -0.3102      0.0882     -3.52    0.001
AR   2     -0.0008      0.0898     -0.01    0.993

Number of observations:  131
Residuals:    SS =  17946.8  (backforecasts excluded)
              MS =    139.1  DF = 129

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square      11.7      38.3      45.0      60.1
DF                10        22        34        46
P-Value        0.304     0.017     0.098     0.080
```

The second AR parameter is not significantly different from zero (T=-0.01, p=0.993>0.05) and so the parameter is not required in the model. We see that the Box-Pierce statistic remains the same at 38.3 so there is no improvement in fit obtained by using an AR(2) model.

Conclusion

We will therefore use an AR(1) model without a constant. This conclusion is consistent with that reached by looking at the pacf.

Overall conclusion

We could argue that either one of these models (MA (1) or AR (1)) would be suitable to apply to the time series.

ARIMA Model: d1d12 (MA(1) without constant)

```
ARIMA model for d1d12

Final Estimates of Parameters
Type          Coef     SE Coef         T        P
MA   1      0.3212      0.0837      3.84    0.000

Number of observations:  131
Residuals:    SS =  17978.9  (backforecasts excluded)
              MS =    138.3  DF = 130

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag               12        24        36        48
Chi-Square      12.0      38.6      46.1      62.1
DF                11        23        35        47
P-Value        0.367     0.022     0.099     0.069

Forecasts from period 144
                             95 Percent Limits
Period      Forecast         Lower        Upper       Actual
 145          0.3043      -24.9497      25.5582
 146         -0.0926      -26.4896      26.3044
```

<u>We need to forecast X_{145} and X_{146}.</u>

Let Xt = monthly ticket sales (in thousands) in month t

$$\mathbf{Wt = Yt - Yt\text{-}12} \quad \text{(series of the 12}^{th}\text{ differences of the Yt)}$$

From ARIMA command for AR(1) model without constant this model is estimated by

$$\mathbf{Wt = Z_t - 0.3212Z_{t-1}}$$

We need to forecast X_{145} and X_{146}.

From forecasts above $\underline{\mathbf{W_{145}}}$ **= 0.304** and $\underline{\mathbf{W_{146}}}$ **= -0.093**

To complete the forecasting of next 2 observations we must connect the X's and W's.

$$\begin{aligned} W_t &= Y_t - Y_{t-12} \\ &= (X_t - X_{t-1}) - (X_{t-12} - X_{t-13}) \\ &= X_t - X_{t-1} - X_{t-12} + X_{t-13} \end{aligned}$$

Therefore $X_t = W_t + X_{t-1} + X_{t-12} - X_{t-13}$

We use this expression to help forecast X_{145}:

$$X_{145} = W_{145} + X_{144} + X_{133} - X_{132}$$

On the right hand side we have a forecast for W_{145} and we know the other X's.

Therefore $\underline{\mathbf{X_{145}}} = \underline{\mathbf{W_{145}}} + X_{144} + X_{133} - X_{132}$

$$\begin{aligned} &= 0.304 + 432 + 417 - 405 \\ &= \mathbf{444.3} \end{aligned}$$

To forecast X_{146} we not that :

$$\begin{aligned} \underline{\mathbf{X_{146}}} &= \underline{\mathbf{W_{146}}} + \underline{\mathbf{X_{145}}} + X_{134} - X_{133} \\ &= -0.093 + 444.3 + 391 - 417 \\ &= \mathbf{418.2} \end{aligned}$$