Table of Contents

1 ARIMA model fitting

1-1 Stationary

To check the stationarity, we plot the series which is given in Figure 1. As it is clear, the plot shows an upward trend, which indicates that this series is not stationary in the mean. However, it is stationary in the variance and no seasonal effect is seen in the plot. Therefore, a non-seasonal ARIMA model seems appropriate for this series. The autocorrelation (ACF) and partial autocorrelation (PACF) of the series are also given in Figure 2. The ACF also shows that the series is not stationary in the mean.

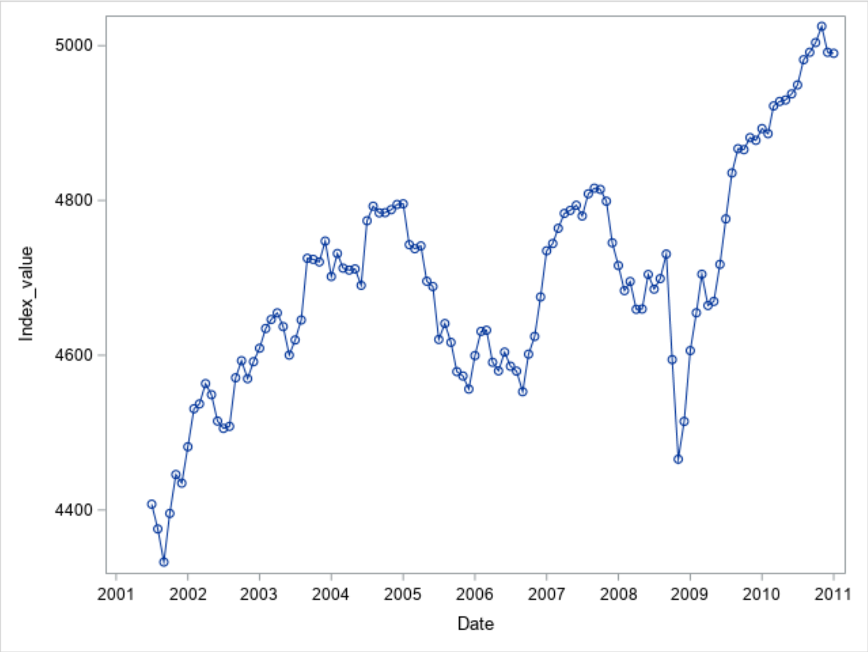

Figure 1: Time series plot

One way to overcome the problem of non-stationarity (in the mean) is to use differencing. For this series, we perform first-order differencing and then plot the differencing series. This plot is given in Figure 3. We observe that the differencing series is stationary.

The ACF and PACF of the differencing series are also given in Figure 3. The ACF and PACF plots are important criteria to identifying approximate order of an ARIMA model. According to the ACF and PACF, it is observed that only one of the correlations is out of bounds. Therefore, an ARMA(1,1) model seems appropriate for the differencing series. Now, following the approach of Box and Jenkins, to fit an ARMA model to the differenced series, we consider different models (with autoregressive and moving average order of at most two) and then select the best model. The criterion for the superiority of one model over another is the smallness of the AIC and SBC of the model.

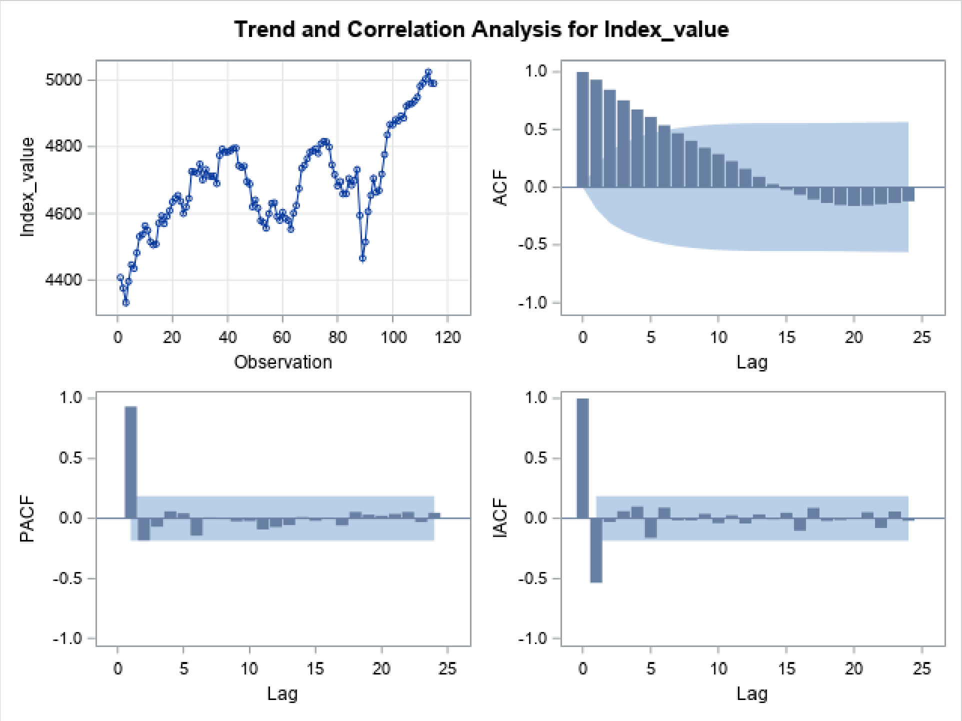

Figure 2: ACF, PACF and IACF of the series

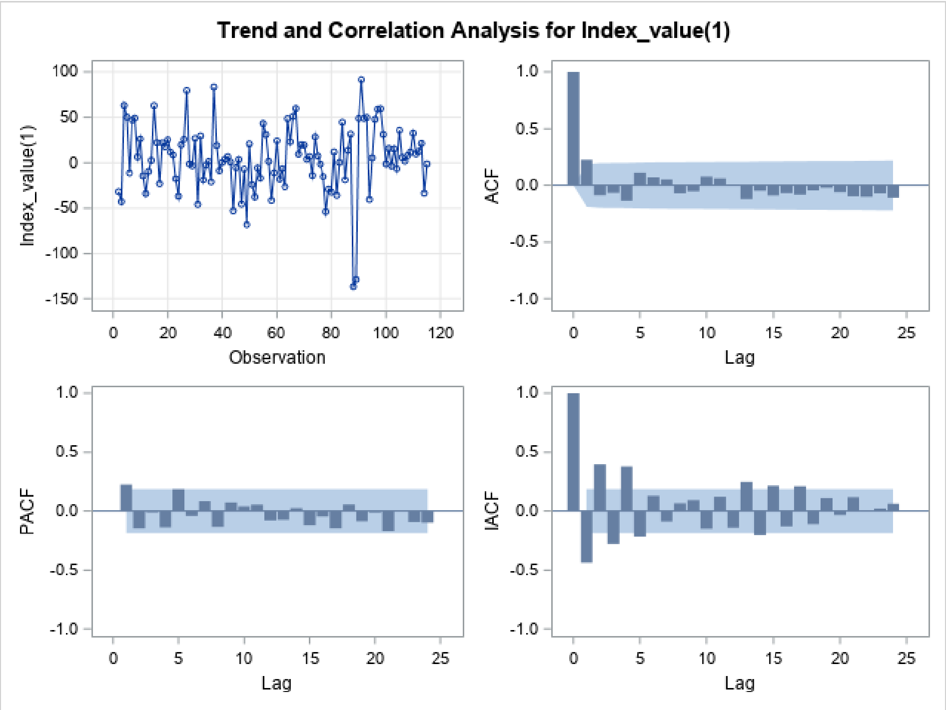

Figure 3: Plot, ACF, PACF and IACF of the differenced series

1-2 Fitting the ARIMA model to the series mean

As mentioned in the previous subsections, a low autoregressive and moving average order ARMA model is suitable for the differenced series. In this subsection, we fit several models to the differenced series and select the best model. Given that only one of the autocorrelations and one of the partial autocorrelations are significant, we consider only the low-order models.

Model	AIC	SBC
AR(1)	1139.046	1144.519
AR(2)	1138.601	1146.81
MA(1)	1137.198	1142.671
MA(2)	1136.912	1145.121
ARMA(1,1)	1134.458	1142.667

Table1 : Fitting different models to the mean of the differencing series (models with intercept)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	5.19331	3.85937	1.35	0.1812	0
MA1,1	-0.81340	0.11298	-7.20	<.0001	1
AR1,1	-0.52048	0.16321	-3.19	0.0019	1

Table2 : Result of fitting the ARMA(1,1) model to the mean of the differenced series (model with intercept)

Table 1 shows the results of fitting different models to the differenced series. Since it is desirable that these two criteria be small, the ARMA(1,1) is the best model. Therefore, we temporarily consider this model for the mean of the differenced series (so the model fitted to the original series will be an ARIMA(1,1,1) model).

The result of fitting this model including parameter estimates and test is presented in Table 2. The intercept is not significant but the other two coefficients are reported to be significant at the 0.05 significance level. According to the estimated parameters, the mathematical form of the model is as follows,

$$Y_t = 5.19331 + \frac{(1 + 0.81340B)}{(1 - B)(1 + 0.52048B)}\varepsilon_t$$

It should be noted that since the intercept is not significant, it can be removed from the model. The result of fitting the model without intercept is given in Table 3. According to this table, the mathematical form of the fitted model is as follows,

$$Y_t = \frac{(1 + 0.80914B)}{(1 - B)(1 + 0.50826B)} \varepsilon_t$$

or

$$(1 - B)(1 + 0.50826B)Y_t = (1 + 0.80914B)\varepsilon_t$$

The above model can be used to predict future values of the series. But before making predictions, we will examine the adequacy of this model in the next subsection.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0.80914	0.11276	-7.18	<.0001	1
AR1,1	-0.50826	0.16270	-3.12	0.0023	1

Table3 : Result of fitting the ARMA(1,1) model to the mean of the differenced series (model without intercept)

1-3 Checking the adequacy of the ARIMA model

In this subsection, we examine the adequacy of the ARIMA(1,1,1) model. To do this, we use the residuals of this model. Since normality of the model residuals is required for constructing prediction intervals, we first examine this hypothesis.

Figure 4 shows the normal probability plot and histogram of the residuals of the ARIMA(1,1,1) model. Except for a few outliers, the rest of the points are well located around the straight line, so the assumption of normality of the residuals can be accepted approximately. The histogram also shows that the residuals are normal. According to the above results, it can be accepted that the distribution of the white noise term in the ARIMA(1,1,1) model is normal.

Figure 5 shows the residuals versus time plot for the ARIMA(1,1,1) model. This plot can be used to examine the constant variance and randomness of the residuals. We observe that the points show a rectangular pattern well, so we can accept the assumption of constant variance of the white noise term in the model. Also, no specific pattern is observed which indicates the randomness (uncorrelatedness) of the residuals of the ARIMA(1,1,1) model.

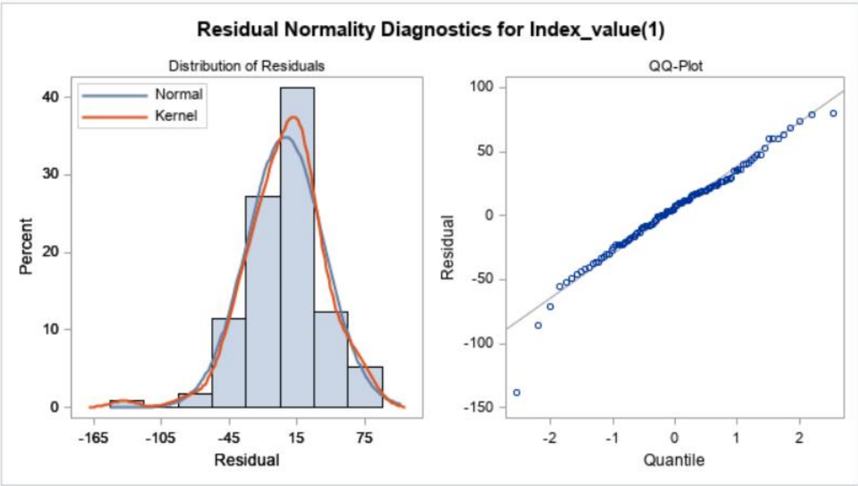

Figure4 : Normal probability plot and histogram of residuals of the ARIMA(1,1,1) model

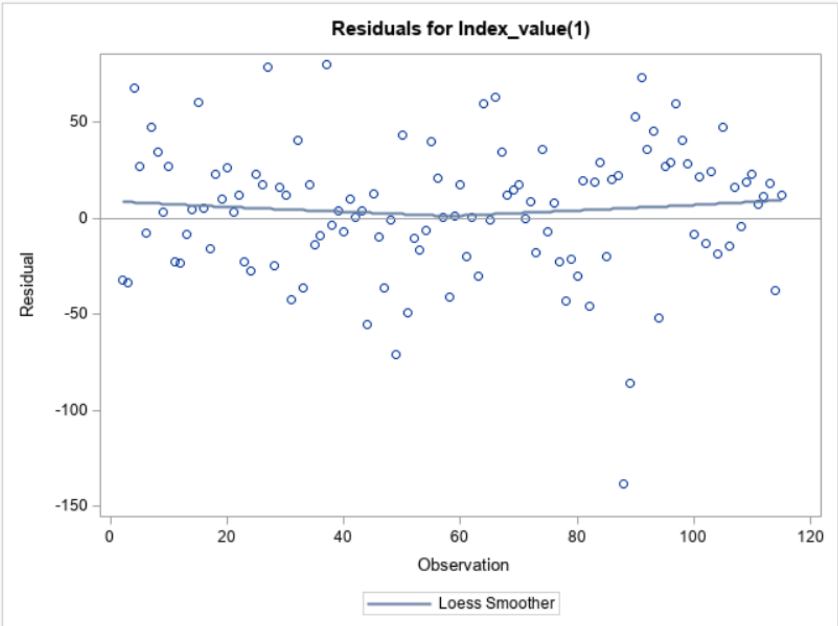

Figure5 : Plot of residuals versus time for the ARIMA(1,1,1) model

The results of the correlation test at some lags for the residuals of the ARIMA(1,1,1) model are given in Table 4. This test can also be used to test the hypothesis that the model errors are uncorrelated. The p-values are all large, so this test also confirms the hypothesis that the white noise terms in the ARIMA(1,1,1) model are uncorrelated.

Figure 6 shows the ACF, PACF, and other plots of the residuals of the ARIMA(1,1,1) model. These plots also confirm the uncorrelatedness of the residuals of the ARIMA(1,1,1) model.

The results show the adequacy of the residuals of the ARIMA(1,1,1) model, so we accept this model as the final model. In the next subsection, we use this model to predict the series values after January 1, 2011 (the day the last data was recorded).

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.95	4	0.4121	0.021	0.041	-0.072	-0.089	0.120	0.056
12	6.88	10	0.7368	0.062	-0.041	-0.051	0.107	0.024	0.055
18	11.81	16	0.7570	-0.136	0.028	-0.106	-0.003	-0.082	-0.006
24	15.51	22	0.8394	-0.017	-0.041	-0.064	-0.090	-0.022	-0.105

Table 4:: Correlation test result for residuals of the ARIMA(1,1,1) model

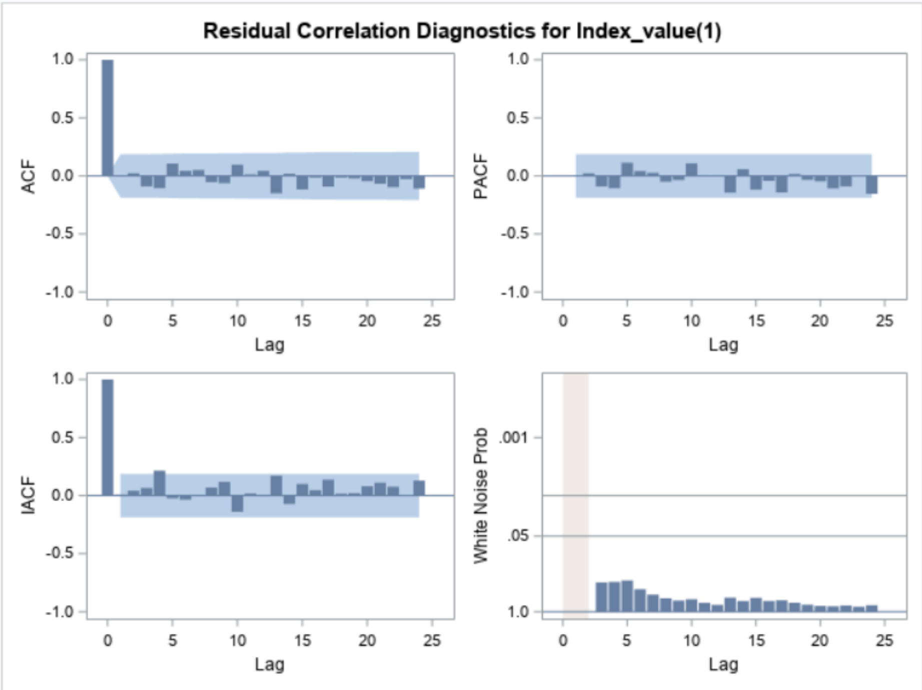

Figure6 : ARIMA(1,1,1) residual plots

1-4 Forecasting with the ARIMA model

In this subsection, we use the model fitted in the previous subsection to predict future values of the series. Table 5 shows the predicted values up to 12 steps ahead along with the prediction confidence interval. Figure 7 also shows the time series graph along with the 12-step-ahead prediction. The prediction confidence interval is also given for the next 12 values.

Forecasts for variable Index_value				
Obs	Forecast	Std Error	95% Confidence Limits	
116	5000.5726	34.7207	4932.5212	5068.6240
117	4995.2448	56.9706	4883.5845	5106.9050
118	4997.9527	69.5291	4861.6781	5134.2273
119	4996.5764	81.5191	4836.8018	5156.3509
120	4997.2759	91.3322	4818.2680	5176.2838
121	4996.9203	100.4773	4799.9884	5193.8523
122	4997.1011	108.7210	4784.0119	5210.1902
123	4997.0092	116.4466	4768.7782	5225.2403
124	4997.0559	123.6597	4754.6873	5239.4245
125	4997.0322	130.4895	4741.2774	5252.7870
126	4997.0442	136.9721	4728.5839	5265.5045
127	4997.0381	143.1649	4716.4401	5277.6361

Table5 : Predicting series values using the fitted ARIMA(1,1,1) model

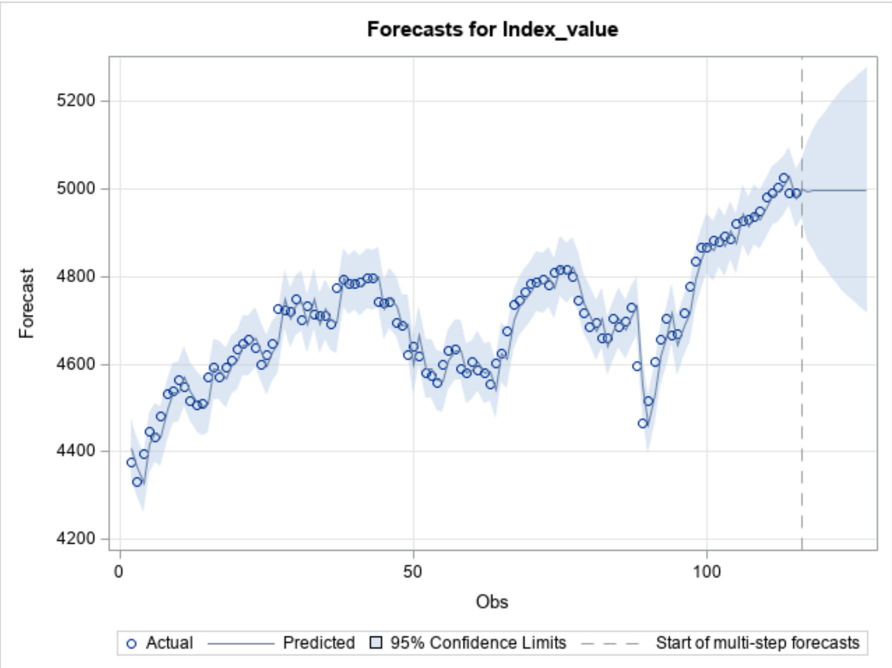

Figure7 : Time series plot with predicted confidence interval up to 12 steps ahead using ARIMA(1,1,1) model

2 Exponential model fitting

Given that the series has a trend but no seasonal effect, in this section a double exponential model (Hult model) is fitted to the data.

2-1 Fitting an exponential model to the series mean

Table 6 shows the results of fitting the double exponential model (Holt's model) to the series. The estimates of the two parameters α and γ are given in this table. According to this table we have

$$\hat{\alpha} = 0.999, \qquad \hat{\gamma} = 0.001$$

Therefore, the mathematical form of the fitted model is as follows,

$$L_t = \hat{\alpha} Y_t + (1-\hat{\alpha})(L_{t-1} + B_{t-1}) = 0.999Y_t + 0.001(L_{t-1} + B_{t-1})$$

$$B_t = \hat{\gamma}(L_t - L_{t-1}) + (1-\hat{\gamma})B_{t-1} = 0.001(L_t - L_{t-1}) + 0.999B_{t-1}$$

According to the above equations, the smoothed series values can be obtained. These values, along with the forecast for the next 12 months, are given in Table 7.

Linear Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|
Level Weight	0.99900	0.06655	15.01	<.0001
Trend Weight	0.0010000	0.0080570	0.12	0.9014

Table 6: Fitting double exponential model (Holt's model) to the series

Obs	Date	Index_value
1	JUL2001	4407.51
2	AUG2001	4375.50
3	SEP2001	4332.68
4	OCT2001	4395.55
5	NOV2001	4445.81
6	DEC2001	4434.68
7	JAN2002	4481.68
8	FEB2002	4530.76

Obs	Date	Index_value
9	MAR2002	4537.09
10	APR2002	4563.29
11	MAY2002	4549.00
12	JUN2002	4514.95
13	JUL2002	4505.45
14	AUG2002	4508.07
15	SEP2002	4570.85
16	OCT2002	4592.89
17	NOV2002	4569.67
18	DEC2002	4591.61
19	JAN2003	4609.03
20	FEB2003	4634.50
21	MAR2003	4646.11
22	APR2003	4654.66
23	MAY2003	4637.02
24	JUN2003	4600.14
25	JUL2003	4619.70
26	AUG2003	4645.54
27	SEP2003	4725.25
28	OCT2003	4723.85
29	NOV2003	4720.57

Obs	Date	Index_value
30	DEC2003	4747.46
31	JAN2004	4701.66
32	FEB2004	4731.39
33	MAR2004	4712.49
34	APR2004	4709.98
35	MAY2004	4711.47
36	JUN2004	4690.26
37	JUL2004	4773.76
38	AUG2004	4792.55
39	SEP2004	4783.95
40	OCT2004	4784.21
41	NOV2004	4788.09
42	DEC2004	4794.80
43	JAN2005	4795.63
44	FEB2005	4742.84
45	MAR2005	4737.62
46	APR2005	4741.21
47	MAY2005	4695.65
48	JUN2005	4688.74
49	JUL2005	4620.36
50	AUG2005	4640.92

Obs	Date	Index_value
51	SEP2005	4616.57
52	OCT2005	4578.78
53	NOV2005	4573.15
54	DEC2005	4556.22
55	JAN2006	4599.50
56	FEB2006	4630.75
57	MAR2006	4632.27
58	APR2006	4590.81
59	MAY2006	4579.77
60	JUN2006	4603.99
61	JUL2006	4585.77
62	AUG2006	4579.39
63	SEP2006	4552.82
64	OCT2006	4601.35
65	NOV2006	4624.28
66	DEC2006	4675.31
67	JAN2007	4734.94
68	FEB2007	4744.32
69	MAR2007	4764.00
70	APR2007	4783.24
71	MAY2007	4787.09

Obs	Date	Index_value
72	JUN2007	4793.79
73	JUL2007	4779.93
74	AUG2007	4808.51
75	SEP2007	4815.83
76	OCT2007	4814.23
77	NOV2007	4799.02
78	DEC2007	4745.26
79	JAN2008	4715.90
80	FEB2008	4683.47
81	MAR2008	4695.25
82	APR2008	4659.32
83	MAY2008	4659.68
84	JUN2008	4704.33
85	JUL2008	4685.39
86	AUG2008	4698.97
87	SEP2008	4730.72
88	OCT2008	4594.32
89	NOV2008	4465.65
90	DEC2008	4514.46
91	JAN2009	4605.96
92	FEB2009	4654.81

Obs	Date	Index_value
93	MAR2009	4704.67
94	APR2009	4664.21
95	MAY2009	4669.53
96	JUN2009	4717.34
97	JUL2009	4776.14
98	AUG2009	4835.52
99	SEP2009	4866.89
100	OCT2009	4865.60
101	NOV2009	4881.17
102	DEC2009	4877.67
103	JAN2010	4892.83
104	FEB2010	4886.24
105	MAR2010	4922.08
106	APR2010	4927.94
107	MAY2010	4929.81
108	JUN2010	4937.69
109	JUL2010	4949.24
110	AUG2010	4981.86
111	SEP2010	4991.46
112	OCT2010	5003.93
113	NOV2010	5025.03

Obs	Date	Index_value
114	DEC2010	4991.41
115	JAN2011	4990.09
116	FEB2011	4993.58
117	MAR2011	4997.06
118	APR2011	5000.54
119	MAY2011	5004.03
120	JUN2011	5007.51
121	JUL2011	5010.99
122	AUG2011	5014.48
123	SEP2011	5017.96
124	OCT2011	5021.44
125	NOV2011	5024.93
126	DEC2011	5028.41
127	JAN2012	5031.89

Table7 : Smoothed series values with prediction of the next 12 values by exponential model

Statistics	Value
MAPE	0.56599
MAE	26.3194
RMSE	35.9868
AIC	828.125
SBC	833.615

Table 8: The values of some goodness-of-fit statistics for the exponential model

The values of some goodness-of-fit statistics are given in Table 8.

2-2 Forecasting with an exponential model

In this subsection, we use the model fitted in the previous subsection to predict future values of the series. Table 9 shows the predicted values up to 12 steps ahead. Figure 8 also shows the time series plot with the 12-step ahead forecast. The prediction confidence interval for the next 12 values of the series is also given in the figure.

Forecast Summary												
Variable	FEB2011	MAR2011	APR2011	MAY2011	JUN2011	JUL2011	AUG2011	SEP2011	OCT2011	NOV2011	DEC2011	JAN2012
Index_value	4993.578	4997.061	5000.544	5004.027	5007.511	5010.994	5014.477	5017.960	5021.443	5024.926	5028.409	5031.893

Table 9: Forecasting series values using the exponential model

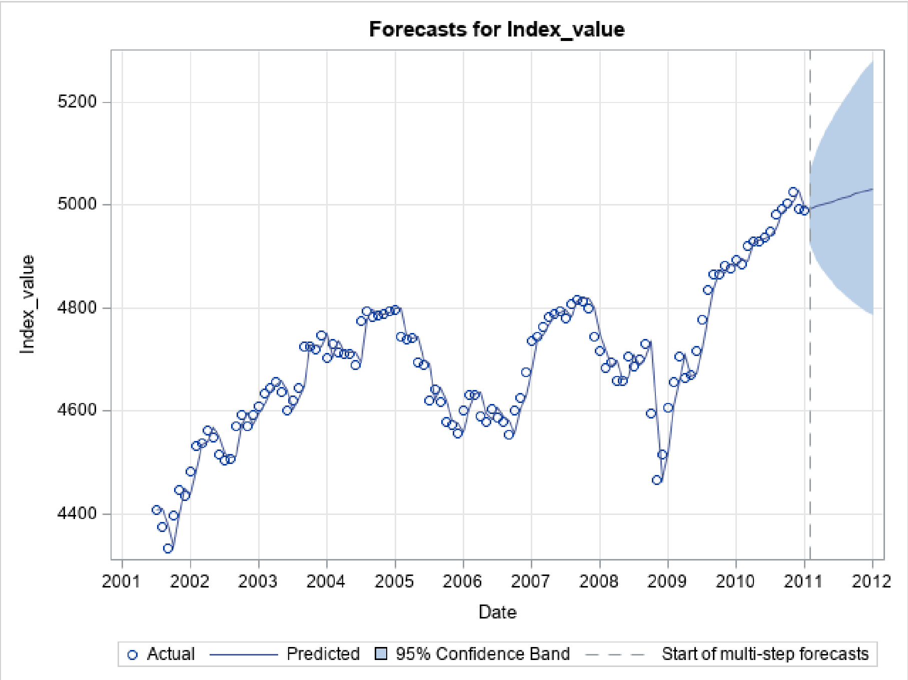

Figure8 : Series plot with forecast and confidence interval forecast up to 12 steps ahead using exponential model

3 Comparison of models

To compare different statistical models, statistics such as AIC or SBC can be used. The values of these statistics for the ARIMA and the exponential models are given in Table 10. Given that it is desirable for these criteria to be small, the exponential model has a better fit to the data.

Model	AIC	SBC
ARIMA(1,1,1)	1134.293	1139.765
Exponential	828.125	833.615

Table10 : AIC and SBC values for ARIMA and exponential models

4 Conclusion

In this project, a time series data was examined. The series plot showed that the series has a trend but does not have seasonal component, so a double exponential model (Holt's model) was considered for this data. A non-seasonal ARIMA model was also fitted to the data. Using these models, future values of the series were predicted. Finally, the performance of these models was compared using the AIC and SBC criteria. It was found that for the double exponential model (Holt's model), both AIC and SBC statistics are smaller, so this model has a better fit to the time series data.

5 Appendix: Computer codes

Arima model codes:

```
data TS1;
infile "/home/u64149066/sasuser.v94/Forecasting-project/data/ts_23845421
(1).csv" delimiter="," firstobs=2;
informat Date mmddyy10.;
format Date mmddyy10.;
*format Date MONYY7.;
input t Date Index_value;
run;
proc sgplot data=TS1;
   series x=t y=Index_value /markers;
run;


proc arima data=TS1
plots=(residual(smooth) forecast(forecasts));
identify var=Index_value;
identify var=Index_value(1);
estimate p=1 q=0;
estimate p=2 q=0;
estimate p=0 q=1;
estimate p=1 q=1 noconstant outest=table_out outstat=metrics
outmodel=res_out;
forecast lead=12 out=pred_values;
quit;
run;
```

Double Exponential Model Codes (Holt’s model):

```
data TS1;
infile "/home/u64149066/sasuser.v94/Forecasting-project/data/ts_23845421
(1).csv" delimiter="," firstobs=2;
informat Date ddmmyy10.;
format Date ddmmyy10.;
input t Date Index_value;
run;
proc esm data=TS1 out=q_fcast lead=12 print=summary print=estimates
plot=forecasts outstat=statistics;
id Date interval=month;
forecast Index_value /model=LINEAR;
run;
proc print data=q_fcast;
run;
proc print data=statistics;
run;
```