MM704 – Forecasting and Credit Risk Analysis

Assessment – 1

Overview

- Introduction
- Identification Stage
- Estimation Stage
- Forecasting Stage
- Exponential Smoothing Method
- Conclusion
- References
- Appendix

Introduction

To predict future results, we will examine previous sales data in this report. Also, we will compare various models to see which one fits the data the best. Finally, forecasting will be done using the ARIMA and Exponential Smoothing Methodologies.

Identification stage

- In this dataset we are going to forecast the sales data.

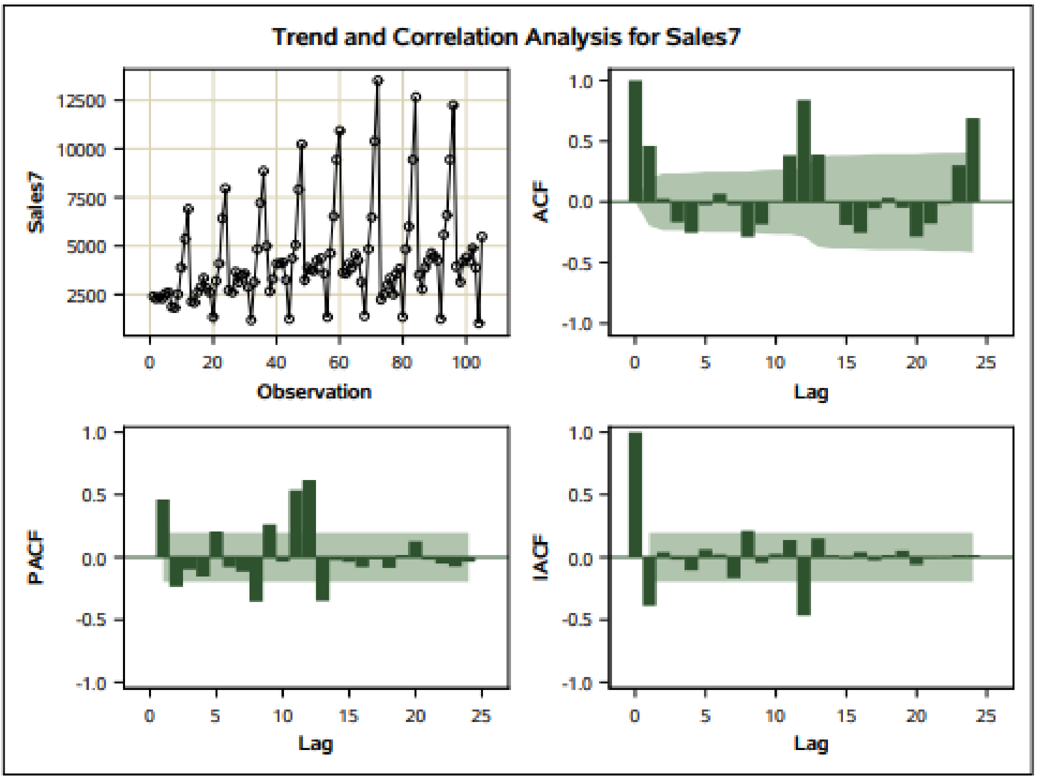
a) Here, we are dealing with annual sales data – (12 months).
b) The sales data is continuous and there are no missing values.

Now, we are going to check for stationarity in the data.

The ARIMA Procedure

Name of Variable = Sales7	
Mean of Working Series	4361.152
Standard Deviation	2541.314
Number of Observations	105

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	34.10	6	<.0001	0.463	0.028	-0.167	-0.255	-0.026	0.067
12	150.93	12	<.0001	-0.025	-0.289	-0.185	-0.000	0.383	0.841
18	182.83	18	<.0001	0.390	-0.009	-0.190	-0.256	-0.054	0.035
24	276.52	24	<.0001	-0.048	-0.285	-0.178	-0.017	0.301	0.690



From the graph it is clear that the data is nonstationary, so we need to transform the data.

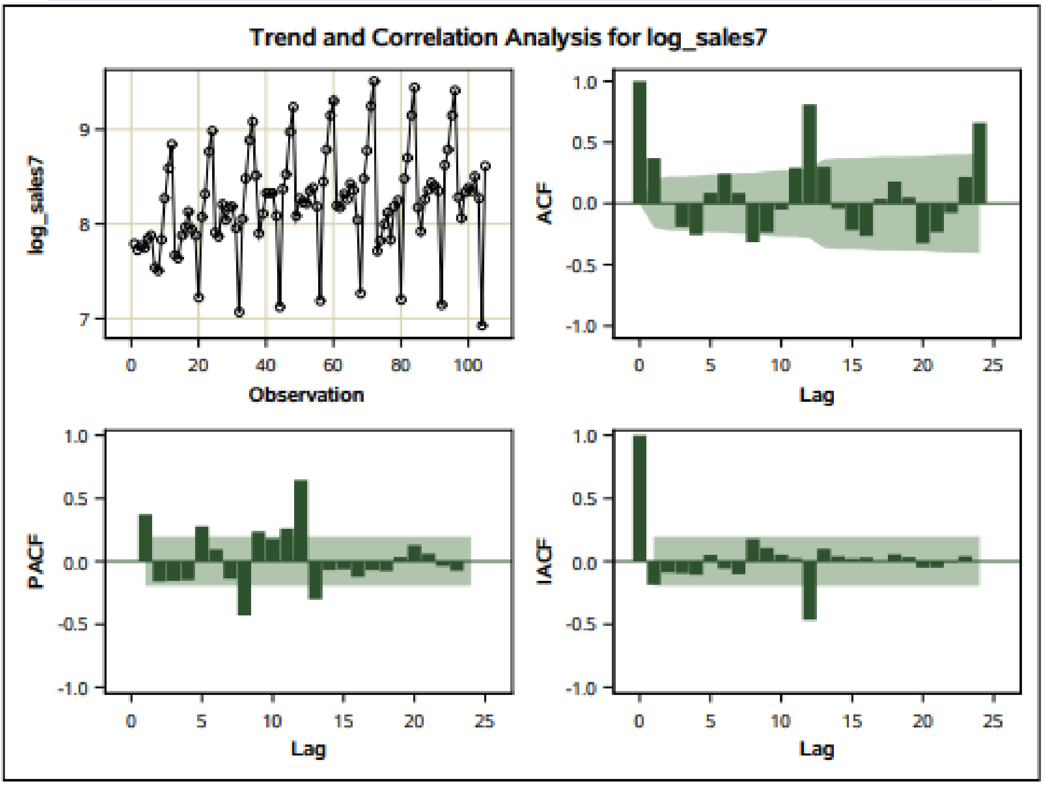
Step - 2

Next, we are going to apply log function on the graph to smoothen the curve.

The ARIMA Procedure

Name of Variable = log_sales7	
Mean of Working Series	8.235817
Standard Deviation	0.534608
Number of Observations	105

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	33.95	6	<.0001	0.373	0.001	-0.189	-0.257	0.089	0.244
12	142.84	12	<.0001	0.088	-0.315	-0.236	-0.050	0.293	0.811
18	173.14	18	<.0001	0.300	-0.041	-0.217	-0.265	0.036	0.180
24	262.71	24	<.0001	0.050	-0.324	-0.236	-0.076	0.217	0.662

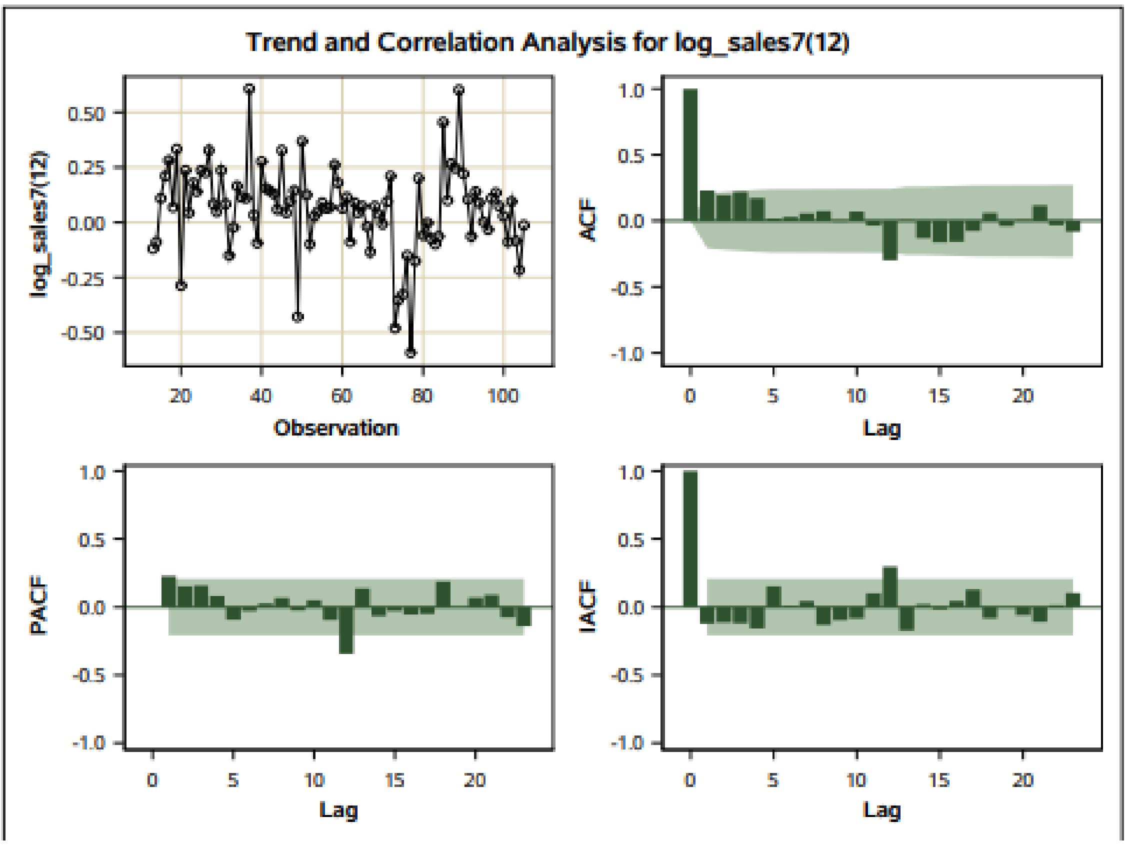


Also, we are going to difference our series with the period of S=12. And then check if there is any seasonal pattern left.

The series were differenced just one hence D=1.

Name of Variable = log_sales7	
Period(s) of Differencing	12
Mean of Working Series	0.061777
Standard Deviation	0.195593
Number of Observations	93
Observation(s) eliminated by differencing	12

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	16.50	6	0.0113	0.229	0.196	0.219	0.171	0.016	0.030
12	27.78	12	0.0060	0.056	0.073	-0.004	0.069	-0.037	-0.299
18	36.61	18	0.0059	0.005	-0.131	-0.162	-0.158	-0.076	0.059



From the graph we can observe that graph is non-stationary

- The graph does not fluctuate with constant mean and constant variance,

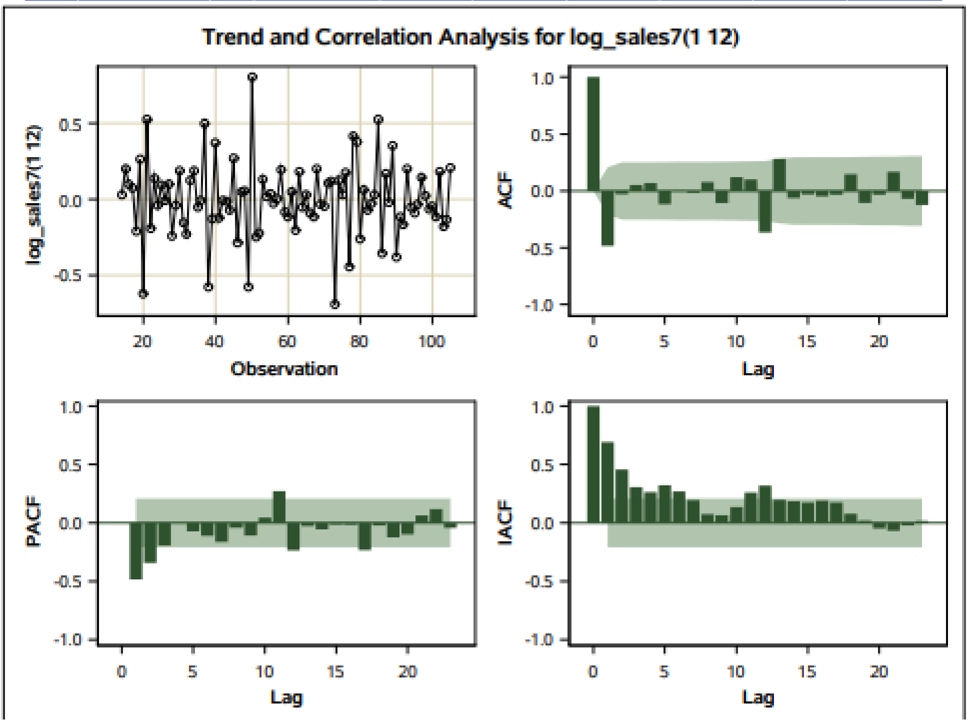
Step - 3

To get the stationary graph we need to difference at lag 1.

Differencing at lag 1 would be d=1 (Inácio, S. (2023). 'Introduction ARIMA').

Name of Variable = log_sales7	
Period(s) of Differencing	1,12
Mean of Working Series	0.001163
Standard Deviation	0.243377
Number of Observations	92
Observation(s) eliminated by differencing	13

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	24.43	6	0.0004	-0.484	-0.028	0.052	0.070	-0.116	0.002
12	43.32	12	<.0001	-0.014	0.078	-0.106	0.122	0.101	-0.365
18	55.57	18	<.0001	0.284	-0.062	-0.030	-0.047	-0.033	0.148



After first differencing, now graph fluctuate with constant mean and constant variation also there is no seasonal component present.

From the ACP & PACF plots above: The autocorrelations and partial autocorrelations are borderline significant at lags 1 and 2. So could fit 1 MA or 1, 2 AR parameters for non-seasonal part of the model.

Estimation Stage:

Step - 4

Now, from the above graph, I have decided to asses' model.

- ARIMA (1,1,0) (1,1,0) _s12 we have S=12, p=1, q = 0, d= 1, P=1, D=1 and Q=0
- ARIMA (0,1,1) (0,1,1)_s12 we have S=12, p=0, q = 1, d= 1, P=0, D=1 and Q=1
- ARIMA (0,1,1) (1,1,0) _s12 we have S=12, p=0, q = 1, d= 1, P=1, D=1 and Q=0

Assessing the model ARIMA (1,1,0) (1,1,0) s12

- we have S=12, p=1, q = 0, d= 1, P=1, D=1 and Q=0

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0001510	0.01083	0.01	0.9889	0
AR1,1	-0.47868	0.09455	-5.06	<.0001	1
AR2,1	-0.36163	0.10167	-3.56	0.0006	12

So, now we need to remove the mean (Mu) and retest it.

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	24.43	6	0.0004	-0.484	-0.028	0.052	0.070	-0.116	0.002
12	43.32	12	<.0001	-0.014	0.078	-0.106	0.122	0.101	-0.365
18	55.57	18	<.0001	0.284	-0.062	-0.030	-0.047	-0.033	0.148

In autocorrelation check for white noise, p-value less than 0.05 suggest that Autocorrelation at the lag 6,12,18 is not white noise **(**Inácio, S. (2023) 'Estimation Residuals Forecasting').

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
AR1,1	-0.47869	0.09403	-5.09	<.0001	1
AR2,1	-0.36163	0.10109	-3.58	0.0006	12

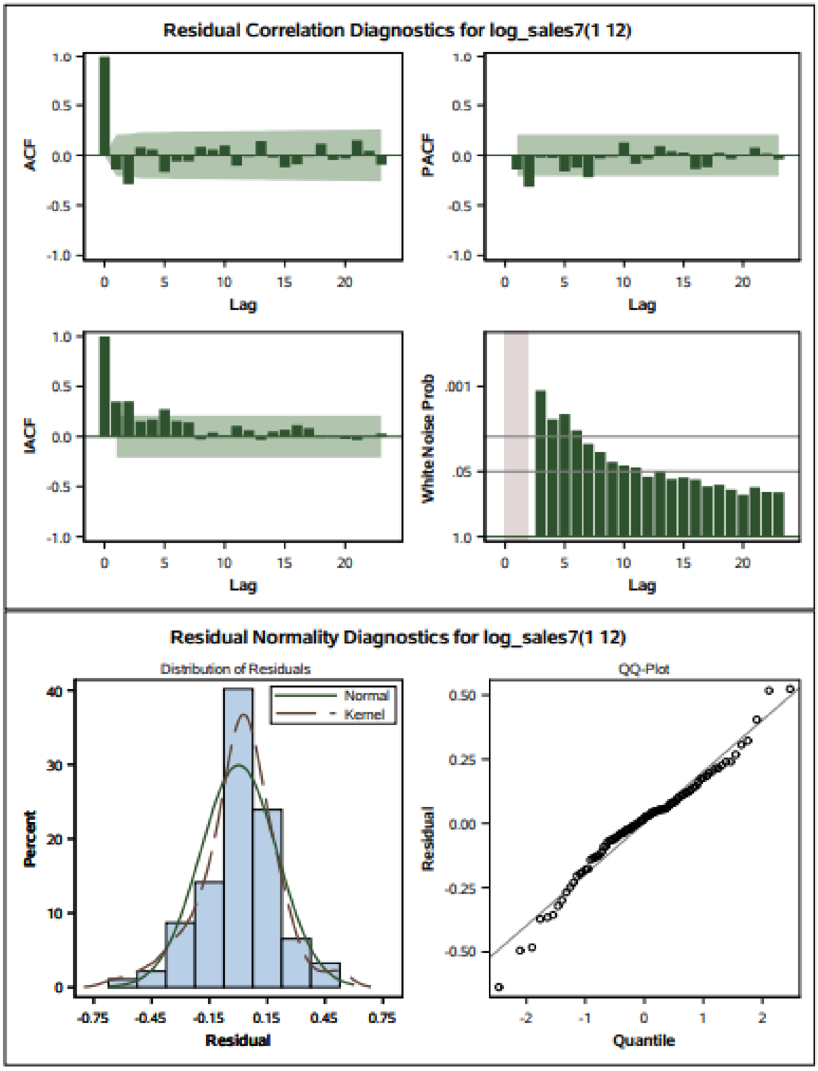
Variance Estimate	0.040449
Std Error Estimate	0.201121
AIC	-32.0459
SBC	-27.0023
Number of Residuals	92

** AIC and SBC do not include log determinant.*

Correlations of Parameter Estimates		
Parameter	AR1,1	AR2,1
AR1,1	1.000	0.102
AR2,1	0.102	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	13.91	4	0.0076	-0.141	-0.288	0.078	0.057	-0.166	-0.061
12	17.56	10	0.0629	-0.059	0.083	0.057	0.099	-0.106	-0.014
18	23.90	16	0.0917	0.141	-0.021	-0.120	-0.088	-0.012	0.117
24	29.60	22	0.1286	-0.046	-0.034	0.150	0.044	-0.094	-0.098

p-value is less than 0.05 in lag2 suggest that there is sufficient evidence to reject H0. Hence, we should retain both the parameters in the model **(**Inácio, S. (2023) ‘Estimation Residuals Forecasting’).



Model for variable log_sales7	
Period(s) of Differencing	1,12

Satu

The ARIMA Procedure

No mean term in this model.

Autoregressive Factors	
Factor 1:	1 + 0.47869 B**(1)
Factor 2:	1 + 0.36163 B**(12)

Overall, model does not appear to be adequate.

Although, From Q-Q plot and histogram indicate that the residuals follow a normal distribution and forecasts from this method could be good.

The plots ACF and PACF shows significant autocorrelations which suggest that the forecasts are not good **(**Inácio, S. (2023) 'Estimation Residuals Forecasting').

<u>Assessing the model ARIMA (0,1,1) (0,1,1) s12</u>

- we have S=12, p=0, q = 1, d= 1, P=0, D=1 and Q=1

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0001857	0.0033439	0.06	0.9558	0
MA1,1	0.74401	0.07172	10.37	<.0001	1
MA2,1	0.39662	0.10167	3.90	0.0002	12

So, now we need to remove the mean (MU) and retest it.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.74375	0.07113	10.46	<.0001	1
MA2,1	0.39640	0.10079	3.93	0.0002	12

Variance Estimate	0.034224
Std Error Estimate	0.184998
AIC	-47.421
SBC	-42.3774
Number of Residuals	92

** AIC and SBC do not include log determinant.*

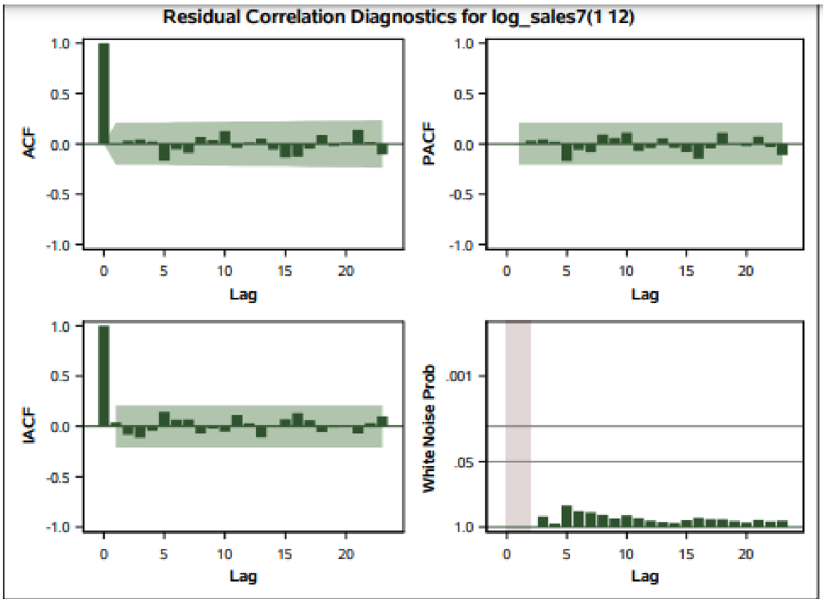
Correlations of Parameter Estimates		
Parameter	MA1,1	MA2,1
MA1,1	1.000	-0.046
MA2,1	-0.046	1.000

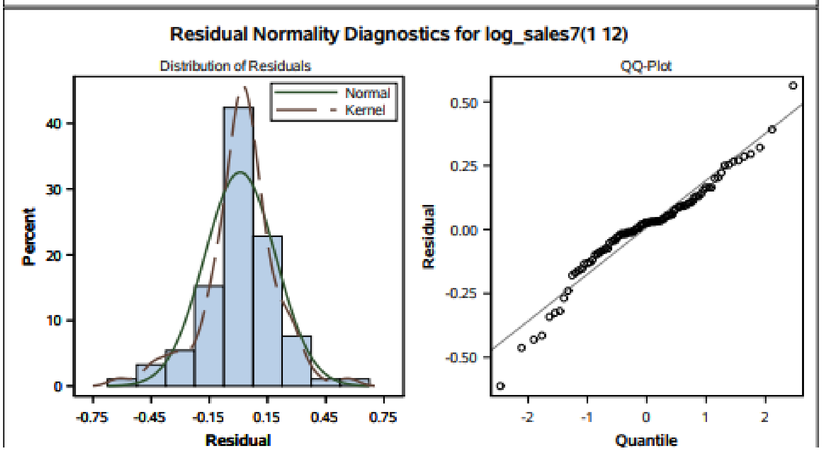
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.41	4	0.4918	0.004	0.033	0.044	0.024	-0.167	-0.055
12	6.71	10	0.7529	-0.088	0.071	0.036	0.126	-0.040	0.012
18	12.45	16	0.7126	0.052	-0.057	-0.136	-0.128	-0.050	0.085
24	16.26	22	0.8026	-0.021	0.011	0.138	0.016	-0.105	0.011

p-value is less than 0.05 in lag2 suggest that there is sufficient evidence to reject H0. Hence, we should retain both the parameters in the model **(**Inácio, S. (2023) 'Estimation Residuals Forecasting').

In autocorrelation check for white noise, p-value less than 0.05 suggest that Autocorrelation at the lag 6,12,18 is not white noise **(**Inácio, S. (2023) 'Estimation Residuals Forecasting').

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	24.43	6	0.0004	-0.484	-0.028	0.052	0.070	-0.116	0.002
12	43.32	12	<.0001	-0.014	0.078	-0.106	0.122	0.101	-0.365
18	55.57	18	<.0001	0.284	-0.062	-0.030	-0.047	-0.033	0.148





Overall, model appears to be adequate.

The plots ACF and PACF shows no significant autocorrelations which suggest that the forecasts are good.

From Q-Q plot and histogram indicate that the residuals follow a normal distribution and forecasts from this method will be good (Inácio, S. (2023) 'Estimation Residuals Forecasting').

In the scatterplot, the residuals appear to be small which suggest that the model is accurate.

In Autocorrelation Check for Residuals section, p value of chi square test is greater than 0.05 suggest that there is no significant autocorrelation of records up to lag 24 (Inácio, S. (2023) 'Estimation Residuals Forecasting').

So now, model will be decided based on AIC value (Inácio, S. (2023) 'Estimation Residuals Forecasting').

The ARIMA Procedure

No mean term in this model.

Moving Average Factors	
Factor 1:	1 - 0.74375 B**(1)
Factor 2:	1 - 0.3964 B**(12)

Assessing the model ARIMA (0,1,1) (1,1,0) s12

- we have S=12, p=0, q = 1, d= 1, P=1, D=1 and Q=0

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.0002337	0.0040517	0.06	0.9541	0
MA1,1	0.73793	0.07267	10.15	<.0001	1
AR1,1	-0.33782	0.10246	-3.30	0.0014	12

So, now we need to remove the mean (MU) and retest it.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.73767	0.07206	10.24	<.0001	1
AR1,1	-0.33766	0.10176	-3.32	0.0013	12

Variance Estimate	0.034693
Std Error Estimate	0.186261
AIC	-46.1685
SBC	-41.125
Number of Residuals	92

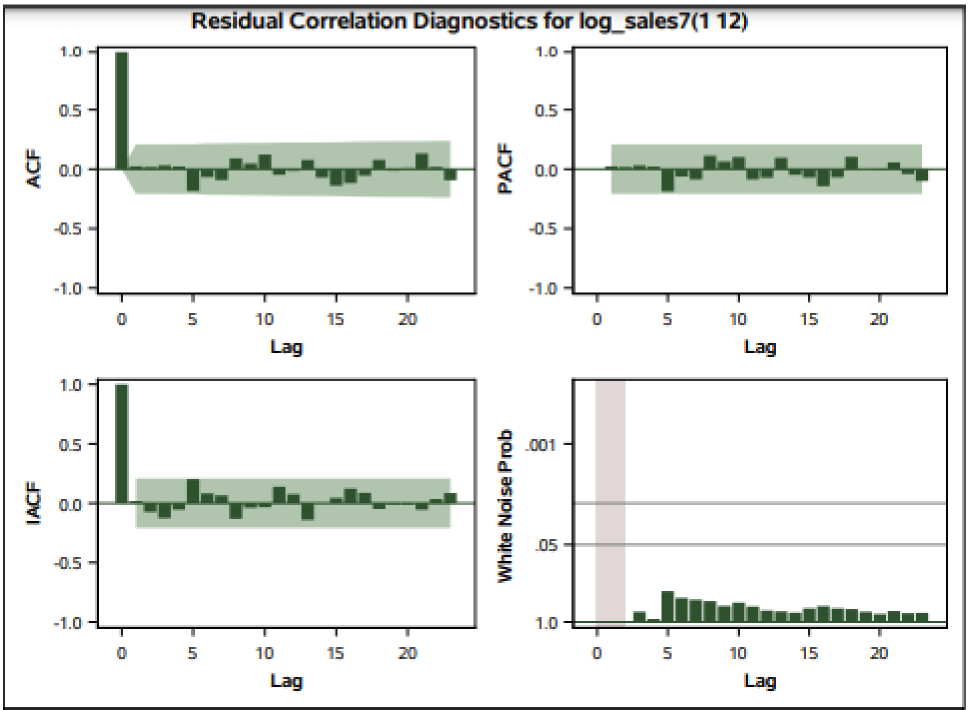
** AIC and SBC do not include log determinant.*

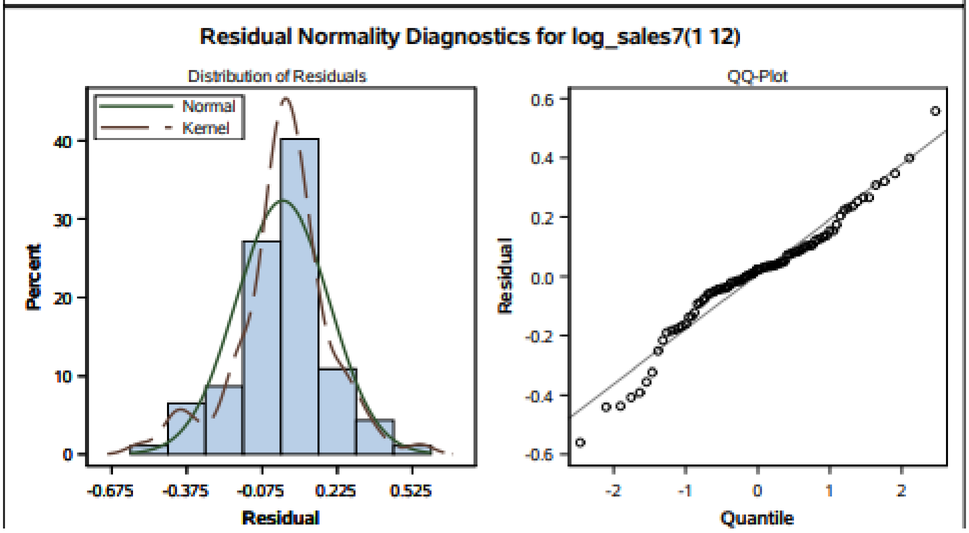
Correlations of Parameter Estimates		
Parameter	MA1,1	AR1,1
MA1,1	1.000	0.054
AR1,1	0.054	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.08	4	0.3957	0.025	0.019	0.033	0.023	-0.184	-0.067
12	7.85	10	0.6432	-0.090	0.091	0.047	0.123	-0.047	-0.015
18	13.82	16	0.6123	0.077	-0.070	-0.139	-0.117	-0.055	0.077
24	18.16	22	0.6965	-0.013	0.008	0.131	0.016	-0.094	-0.092

p-value is less than 0.05 in lag2 suggest that there is sufficient evidence to reject H0. Hence, we should retain both the parameters in the model.

In autocorrelation check for white noise, p-value less than 0.05 suggest that Autocorrelation at the lag 6,12,18 is not white noise **(**Inácio, S. (2023) ‘Estimation Residuals Forecasting’).





Autoregressive Factors	
Factor 1:	1 + 0.33766 B**(12)

Moving Average Factors	
Factor 1:	1 - 0.73767 B**(1)

Overall, model appears to be adequate.

The plots ACF and PACF shows no significant autocorrelations which suggest that the forecasts are good.

In the scatterplot, the residuals appear to be small which suggest that the model is accurate **(**Inácio, S. (2023) 'Estimation Residuals Forecasting').

From Q-Q plot and histogram indicate that the residuals follow a normal distribution and forecasts from this method will be good.

In Autocorrelation Check for Residuals section, p value of chi square test is greater than 0.05 suggest that there is no significant autocorrelation of records up to lag 24.

So now, model will be decided based on AIC value.

Model Selection

Among those three models only two of them seems to be adequate but comparing the AIC value between them ARIMA (0, 1, 1) $(0, 1, 1)_{12}$ seems to have the lowest AIC & SBC value compared to another model.

Testing outliers in this model

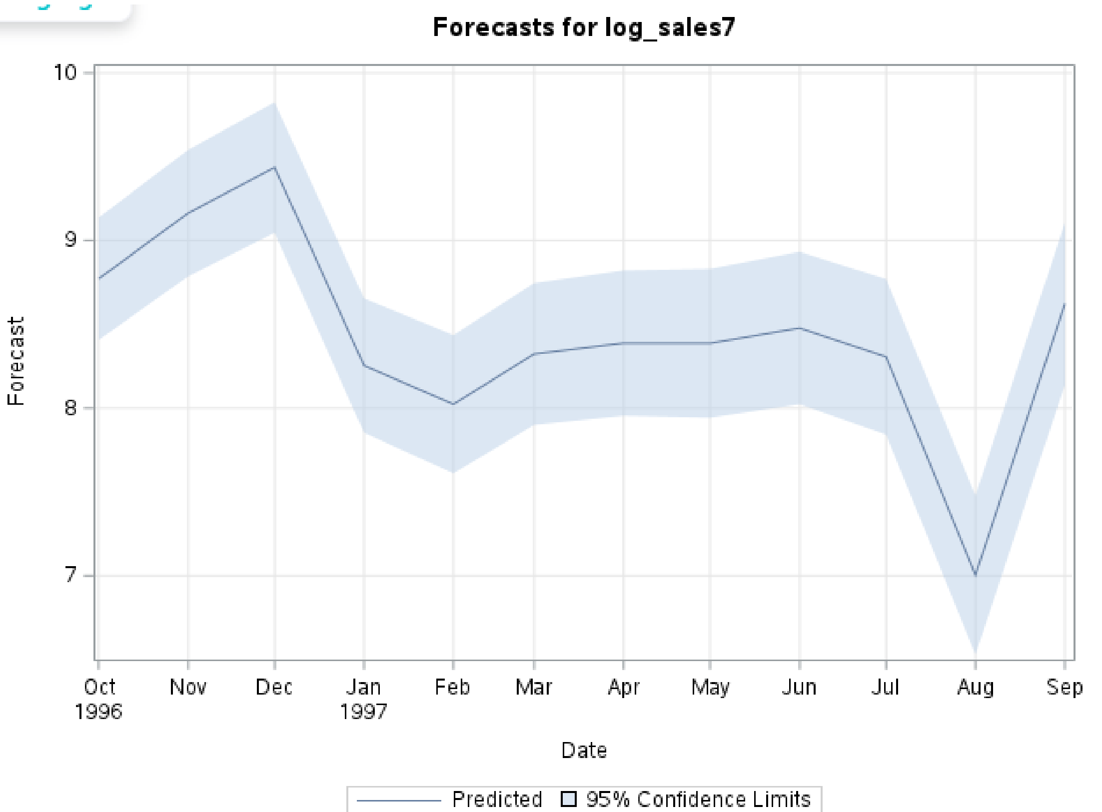
Outlier Details				
Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
37	Additive	0.55273	32.28	<.0001
73	Shift	-0.38813	34.78	<.0001
8	Additive	0.57233	32.29	<.0001

Since the model diagnostic tests show that all the parameter estimates are significant, and the residual series is white noise.

Best Fitted Model**: ARIMA (0, 1, 1) $(0, 1, 1)_{12}$**

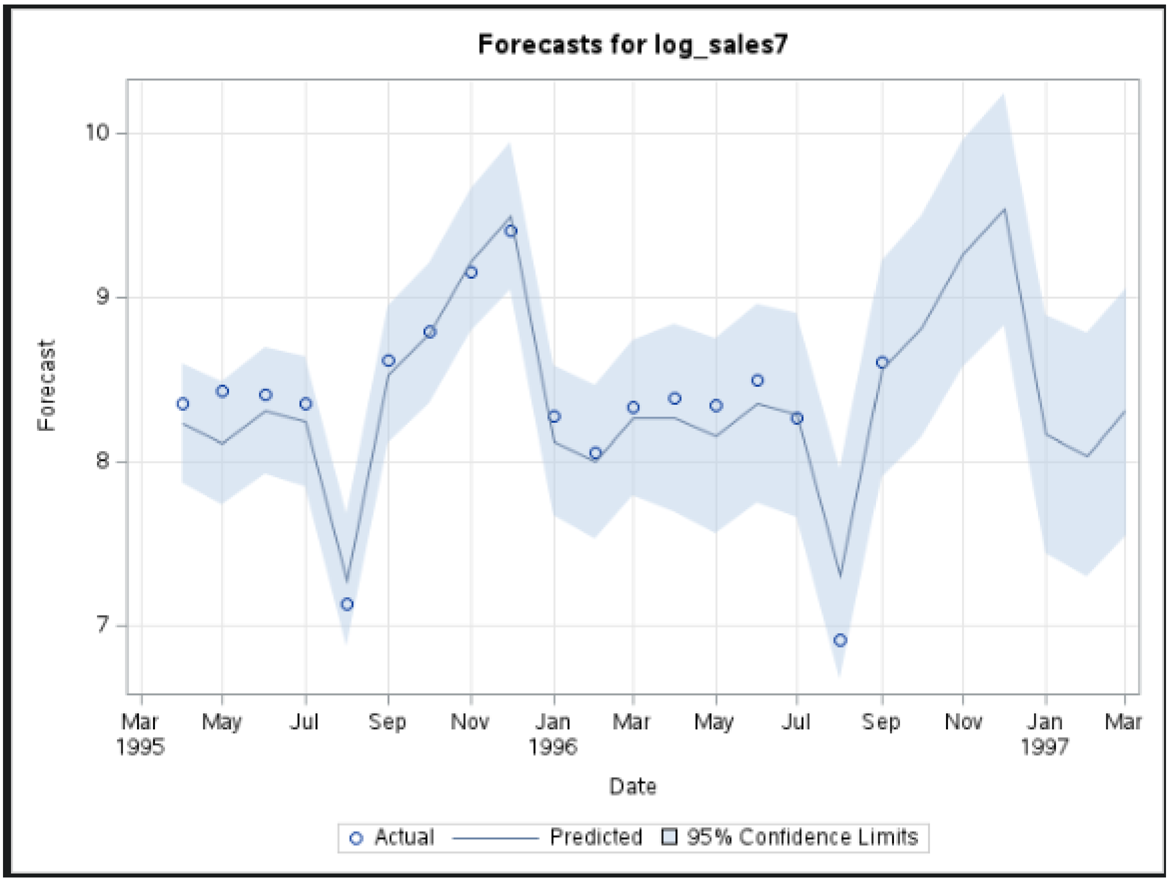
Forecasting Stage:

Forecasts for variable log_sales7				
Obs	Forecast	Std Error	95% Confidence Limits	
106	8.7716	0.1863	8.4065	9.1366
107	9.1630	0.1926	8.7856	9.5405
108	9.4354	0.1987	9.0460	9.8248
109	8.2531	0.2046	7.8521	8.6540
110	8.0232	0.2103	7.6109	8.4355
111	8.3225	0.2159	7.8992	8.7457
112	8.3874	0.2214	7.9535	8.8213
113	8.3866	0.2267	7.9422	8.8310
114	8.4774	0.2319	8.0228	8.9320
115	8.3056	0.2370	7.8410	8.7701
116	7.0036	0.2420	6.5292	7.4779
117	8.6223	0.2469	8.1384	9.1062



Forecasts for variable log_sales7						
Obs	Forecast	Std Error	95% Confidence Limits		Actual	Residual
86	7.9810	0.1863	7.6159	8.3461	7.9237	-0.0573
87	8.1442	0.1926	7.7668	8.5216	8.2651	0.1210
88	8.2021	0.1987	7.8128	8.5915	8.3608	0.1586
89	8.0729	0.2046	7.6719	8.4738	8.4360	0.3631
90	8.2825	0.2103	7.8702	8.6947	8.4060	0.1236
91	8.2176	0.2159	7.7944	8.6409	8.3507	0.1330
92	7.2574	0.2214	6.8235	7.6913	7.1381	-0.1193
93	8.5190	0.2267	8.0746	8.9633	8.6217	0.1028
94	8.7659	0.2319	8.3113	9.2205	8.7919	0.0261
95	9.2238	0.2370	8.7593	9.6884	9.1539	-0.0699
96	9.5073	0.2420	9.0330	9.9816	9.4149	-0.0924
97	8.0542	0.2469	7.5703	8.5381	8.2810	0.2267
98	7.9660	0.3010	7.3760	8.5560	8.0596	0.0936
99	8.1324	0.3118	7.5213	8.7435	8.3373	0.2049
100	8.2104	0.3222	7.5789	8.8419	8.3866	0.1762
101	8.0306	0.3323	7.3794	8.6819	8.3471	0.3165
102	8.2876	0.3421	7.6172	8.9581	8.4994	0.2118
103	8.2658	0.3516	7.5767	8.9549	8.2682	0.0024
104	7.2758	0.3608	6.5686	7.9831	6.9207	-0.3552
105	8.5442	0.3699	7.8193	9.2691	8.6083	0.0641
106	8.7830	0.3787	8.0408	9.5252	.	.
107	9.2380	0.3873	8.4790	9.9971	.	.
108	9.5253	0.3957	8.7497	10.3008	.	.
109	8.1316	0.4040	7.3398	8.9233	.	.

Holdout Method

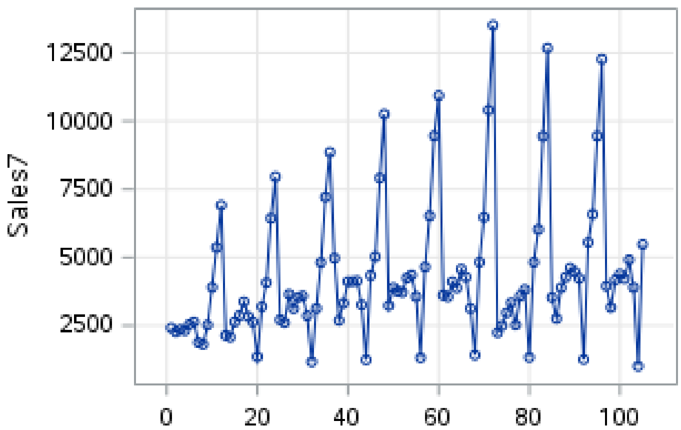


As the prediction generated by the model ARIMA (0, 1, 1) (0, 1, 1)12 during the training session was found to be within the 95% confidence interval, it may be expected that forecasts generated from this model will be quite accurate.

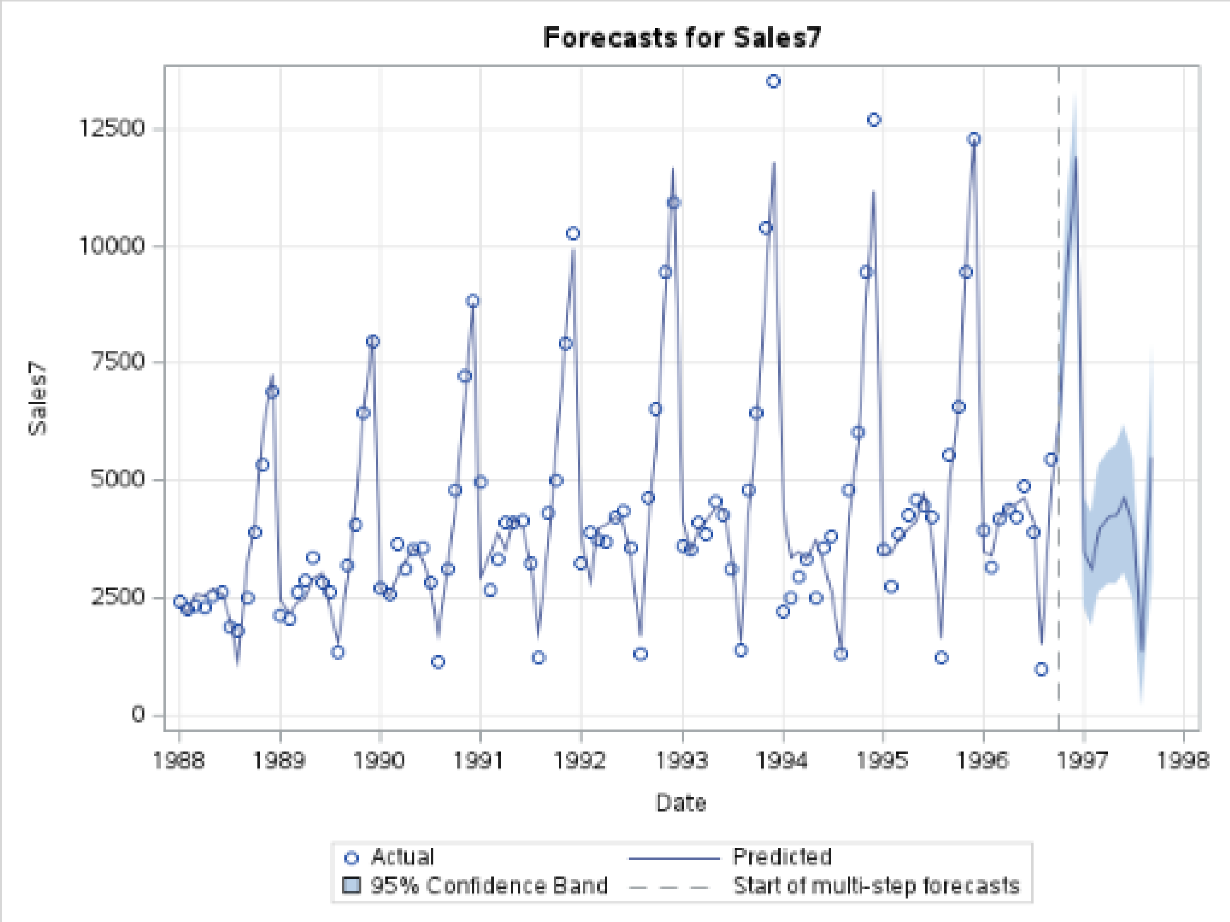
Exponential Smoothing Method

From the sales graph it is clear that, the size of the seasonal fluctuations increases as the level of the series increases it follows multiplicative seasonality.

Hence, we are going to apply holt's winters method (Inácio, S. (2023) Time Series – Exponential Smoothing Models).



Variable Information	
Name	Sales7
Label	Sales7
First	JAN1988
Last	SEP1996
Number of Observations Read	105



The Forecast of exponential Smoothing method is very identical to the ARIMA (0, 1, 1) (0, 1, 1)$_{12}$.

Forecast Summary												
Variable	OCT1996	NOV1996	DEC1996	JAN1997	FEB1997	MAR1997	APR1997	MAY1997	JUN1997	JUL1997	AUG1997	SEP1997
Sales7	6261.518	9368.195	11896.46	3503.248	3122.305	3991.826	4217.487	4260.986	4635.488	3915.486	1358.101	5498.948

Forecast Summation				
Variable	Forecast	Standard Error	Confidence Limits	
Sales7	62030	6206.74	49865	74195

Statistical Test Result:

MSE	RMSE	UMSE	URMSE	MAPE	MAE	RSQUARE
325404.4414	570.4423208	334975.1603	578.7703865	11.87870761	407.4135271	0.949614354

The mean of the absolute percent errors is 11.87.

Conclusion

We performed analysis in this report utilising ARIMA and the Exponential Smoothing Technique. To implement ARIMA method first we had to transform the data into stationary Then we evaluated different models to identify the best fit, and ultimately, we forecasted using the best fitted model, ARIMA (0, 1, 1) (0, 1, 1)12. Finally, to verify the accuracy of the forecast, we used the holdout approach. Finally, due to the multiplicative seasonality of the graph, we employed Holt Winters' approach for Exponential Smoothing Methods.

References

Inácio, S. (2023). ‘Time Series –Exponential Smoothing Models’ MM704 Forecasting and Credit Risk Analysis. School of ATE, University of Brighton.

Inácio, S. (2023). ‘Smoothing methods in SAS’, MM704 Forecasting and Credit Risk Analysis. School of ATE, University of Brighton.

Inácio, S. (2023). ‘Introduction ARIMA’, MM704 Forecasting and Credit Risk Analysis. School of ATE, University of Brighton.

SAS Customer Support Site | SAS support (2023). Available at: https://support.sas.com/documentation/onlinedoc/ets/132/esm.pdf (Accessed: March 31, 2023).

Inácio, S. (2023). ‘Estimation Residuals Forecasting’, MM704 Forecasting and Credit Risk Analysis. School of ATE, University of Brighton.

Inácio, S. (2023). ‘Times Series Characteristics’, MM704 Forecasting and Credit Risk Analysis. School of ATE, University of Brighton.

Appendix

```
data asses1;
set '/home/u63249191/Forecasting_risk/assesment1.sas7bdat';
run;

ODS PDF
        FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
  STYLE = EGDefault;
proc arima data=asses1;
identify var= sales7;
quit;
ODS PDF CLOSE;

data asses1;
set '/home/u63249191/Forecasting_risk/assesment1.sas7bdat';
log_sales7=log(sales7);
run;

ODS PDF
        FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
  STYLE = EGDefault;
proc arima data=asses1;
identify var=log_sales7;
quit;

ODS PDF CLOSE;

ODS PDF
        FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
```

```
   STYLE = EGDefault;
proc arima data=asses1;
identify var=log_sales7(12);
quit;
ODS PDF CLOSE;

ODS PDF
        FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
   STYLE = EGDefault;
proc arima data=asses1;
identify var=log_sales7(1,12);
quit;
ODS PDF CLOSE;

/* Alternative model - not effective*/

ODS PDF
        FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
   STYLE = EGDefault;
/* Alternative model - effective*/
proc arima data=asses1;
identify var=log_sales7(1,12);
estimate p = (1)(12) noconstant; /* ARIMA(1,1,0) (1,1,0)_s12 - reject/unstable*/
quit;/* Mu insignificant */
ODS PDF CLOSE;

ODS PDF
        FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
   STYLE = EGDefault;
```

```
/* Alternative model - effective - best*/
proc arima data=asses1
plots = (residual(smooth) forecast(forecasts));
identify var=log_sales7(1,12);
estimate q = (1)(12) noconstant; /*  ARIMA(0,1,1) (0,1,1)_s12 - Best fit*/
outlier;
quit;/* Mu insignificant */
ODS PDF CLOSE;

ODS PDF
	FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
   STYLE = EGDefault;
proc arima data=asses1;
identify var=log_sales7(1,12);
estimate q =(1) p = (12) noconstant; /*  ARIMA(0,1,1) (1,1,0)_s12 -2nd best fit*/
run;
ODS PDF CLOSE;

/* Forecasting model - multiplicative*/
proc arima data=asses1;
identify var=log_sales7(1,12);
estimate q = (1)(12) noconstant;
forecast lead=12 interval=month  out=table_forecast id= date alpha=0.05;
forecast back = 18 interval=month  out=table_forecast1 id= date alpha=0.05;
run;

ODS PDF
	FILE = '/home/u63249191/Forecasting_risk/Forecast_data/stationarity.pdf'
   STYLE = EGDefault;
```

```
proc arima data=asses1;
identify var=log_sales7(1,12);
estimate q =(1) p =(12) noconstant;
forecast lead=12 interval=month  out=table_forecast id= date alpha=0.05;
forecast back = 20 interval=month  out=table_forecast1 id= date alpha=0.05;
proc print data=table_forecast;
run;
quit;
ODS PDF CLOSE;


/* Exponential smoothening method */
ODS PDF
        FILE =
'/home/u63249191/Forecasting_risk/Forecast_data/exponential_smoothing.pdf'
   STYLE = EGDefault;
proc esm data=asses1 out=table_forecast lead=12 print=summary plot=forecasts
outstat=asses2;
id date interval=month ;
forecast sales7 /model=winters alpha=0.05 accumulate=total;
run;
proc print data=table_forecast;
run;
ODS PDF CLOSE;


/* Mape, why holt winters - multiplicative, arima and ESM have same output */
proc export
  data=asses2
  dbms=xlsx
  outfile="/home/u63249191/Forecasting_risk/Forecast_data/sheet.xlsx"
  replace;
```

run;