IMSE 586 PROJECT ASSIGNMENT

Part C: Time Series (Box-Jenkins) - Case Study on Nite's Rest Inc.

Team:

Suresh Ooty Krishnaswamy

Selvamani Masilamani

Contents

TABLES

FIGURES

Team Effort:

As the team comprises of two members, the effort between the members have been equal, each contributing to this part of project (Part C) at 50%.

"**The Project Team has not given nor received any aid on this assignment**"

Data:

Case Study data – Monthly Average Number of occupied rooms for 168 months (1963 – 1977)

A sample of data is as follows.

t	yt
1	501
2	488
3	504
4	578
5	545
6	632
7	728
8	…

C1:

Develop and diagnose the same model as their Model 5 and report your results.

a. Show ACF (Autocorrelation Function) and PACF (Partial ACF) of the differenced time series used for the model and comment on your findings

b. Perform residual analysis of the model results and comment on your findings

c. Show your results in Table 12.2 format of the original case study and discuss how your Minitab results compare with the Model 5 results of the original case study

Solution:

A time series plot on the raw data indicate that there is a seasonality, in every 12 time period (in this case 12 month). Lowest being the months 2 & 11 (Feb & Nov) and highest being months 7 & 8 (indicating that the hotel was at a summer holiday location).

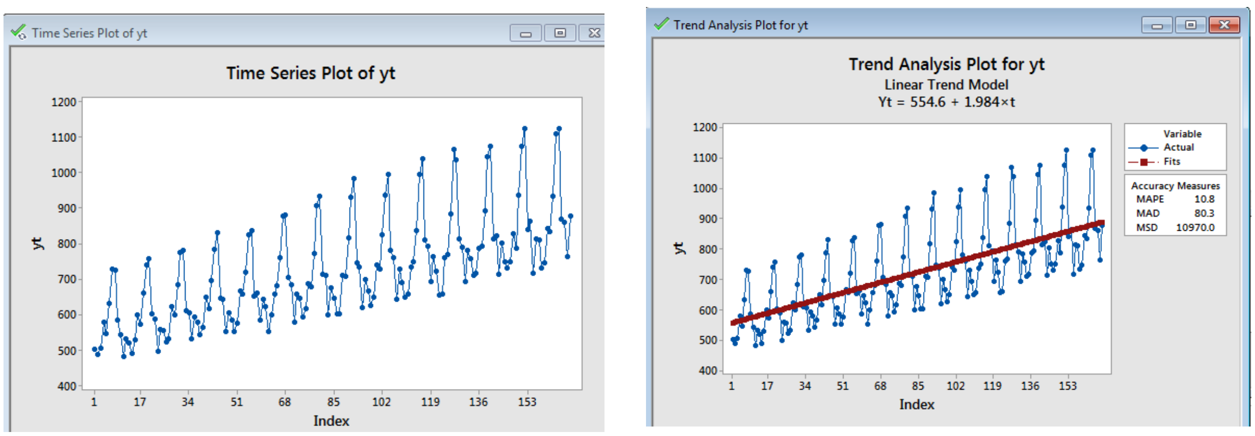


FIGURE 1: Time Series and Trend Analysis on Actual data

A growing trend is evident from the Trend Analysis plot for y_t. This indicates that the data is candidate for transformation such as Natural Log (yt).

Time-series plot on the transformed data.

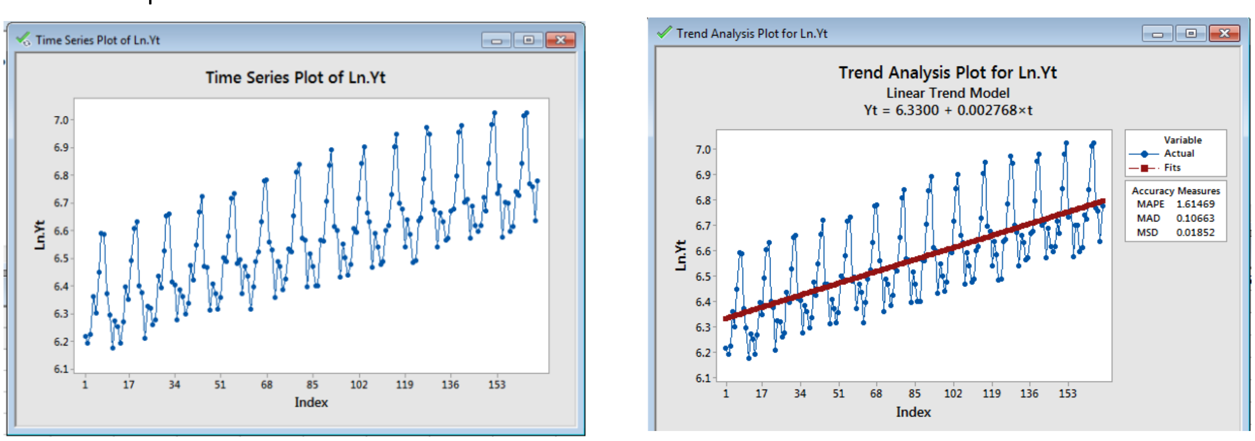


FIGURE 2: Time Series and Trend Analysis on Transformed data

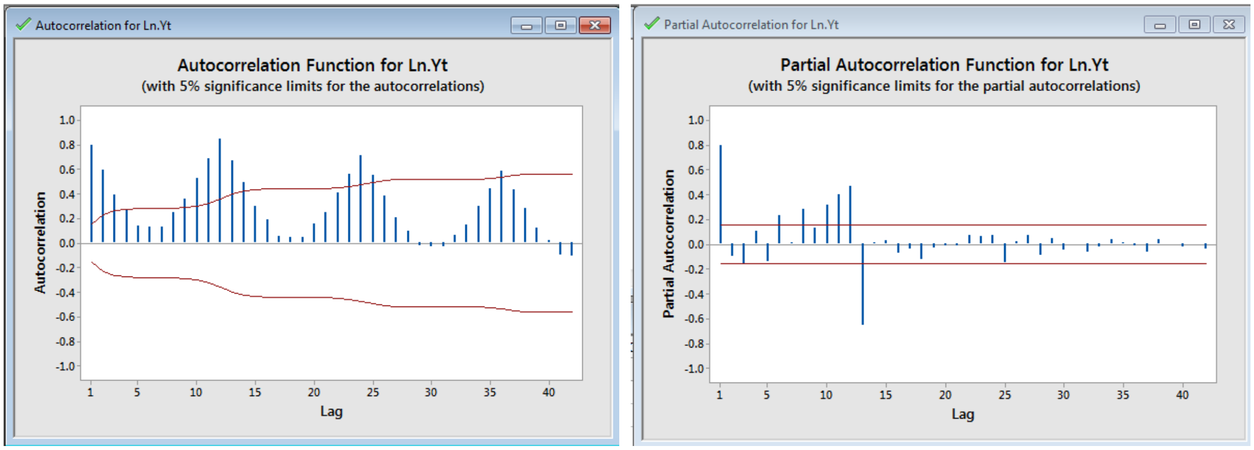


FIGURE 3: ACF & PACF on Transformed data

The ACF indicate that the data is non-stationery, where the autocorrelation is declining slowly with first three AR values are significant. The PACF at lag = 13 shows a significant value. To remove the seasonality, a difference of 12 is required to study the yearly pattern.

Applying the Model 5 ARIMA (3,0,0) $(0,1,1)^{12}$

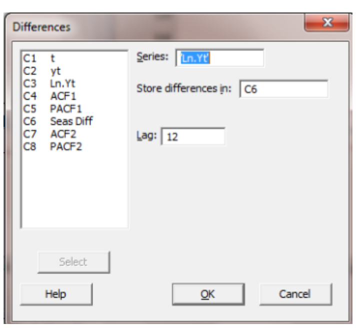
Model 5: $y^*_t = \delta + \phi_1 y^*_{t-1} + \phi_2 y^*_{t-2} + \phi_3 y^*_{t-3} + y^*_{t-12} - \phi_1 y^*_{t-13} - \phi_2 y^*_{t-14} - \phi_3 y^*_{t-15} - \theta_{1,12}\epsilon_{t-12} + \epsilon_t$

Operators: Seasonal moving average of order 1

Nonseasonal autoregressive of order 3

Differencing: $z_t = y^*_t - y^*_{t-12}$

Finding the seasonal difference (lag = 12), as the seasonality is identified for each year.



Applying the ACF and PACF on the seasonal difference data.

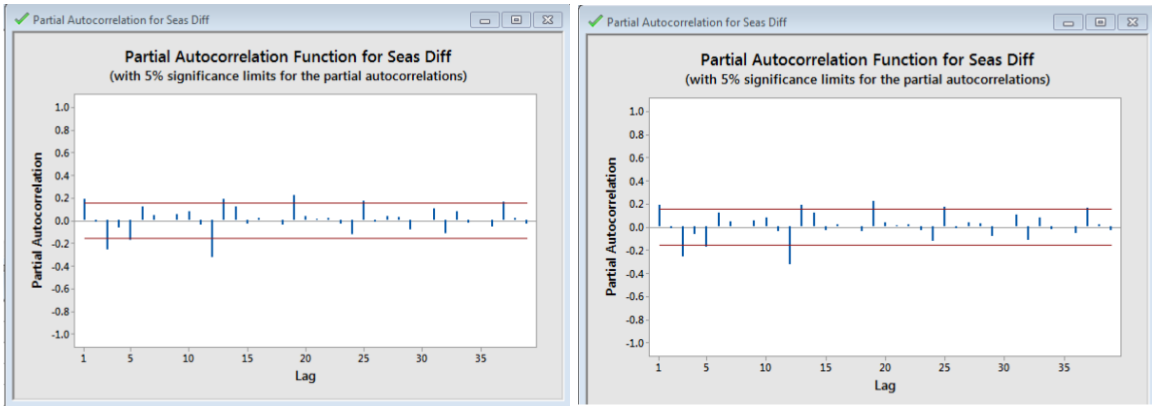
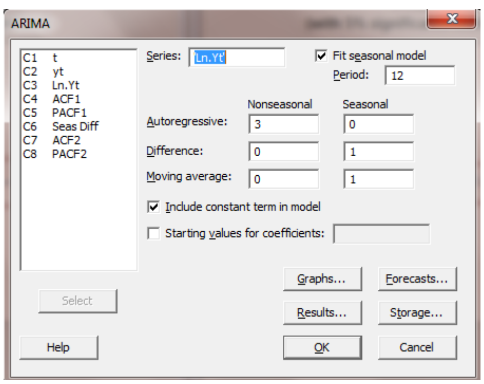


FIGURE 4: ACF & PACF on Transformed data

The ACF is significant for lags 3, 12,19 & 24. This indicates that the non-stationary characteristic is not removed by the difference. The PACF is significant for lags 3, 12,19 & 24.

ARIMA:

The ARIMA model 5: (3,0,0) $(0,1,1)^{12}$on Transformed data,



The ACF and PACF graph indicate that lag 10 has significant Autocorrelations and lag 10 & 14 partial autocorrelations (both at very little scale above the threshold for significance) in the ARIMA model.

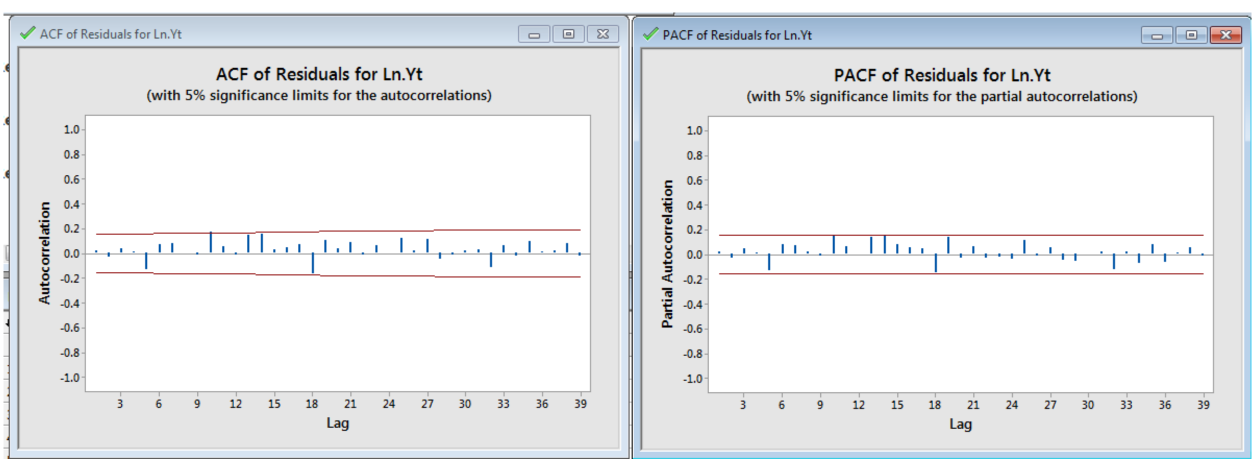


FIGURE 5: ACF & PACF using ARIMA

FIGURE 6: Residual plots using ARIMA

The residual plots satisfy the homoscedasticity, normality and independence assumptions.

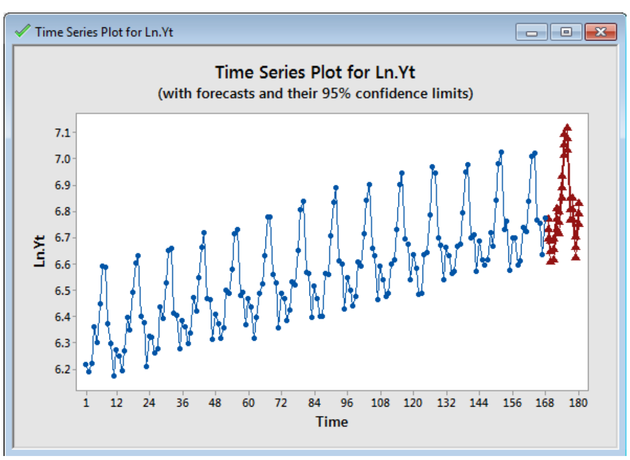


FIGURE 7: Time Series Forecast for 12 periods using ARIMA

ARIMA Model: Ln.Yt

```
Unable to reduce sum of squares any further

Final Estimates of Parameters

Type          Coef    SE Coef      T      P
AR   1      0.2792     0.0799   3.49  0.001
AR   2      0.1608     0.0823   1.95  0.053
AR   3     -0.2377     0.0799  -2.97  0.003
SMA  12     0.5956     0.0725   8.21  0.000
Constant 0.0263432  0.0006631  39.73  0.000


Differencing: 0 regular, 1 seasonal of order 12
Number of observations:  Original series 168, after differencing 156
Residuals:    SS =  0.0559476 (backforecasts excluded)
              MS =  0.0003705  DF = 151


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

Lag             12     24     36     48
Chi-Square    11.1   29.6   41.6   49.2
DF               7     19     31     43
P-Value      0.136  0.057  0.096  0.238
```

TABLE 12.2 *Comparison of Models 1–7*

	Model 1	2	3	4	5	6
Number of regular differences	0	0	0	0	0	0
Number of seasonal differences	0	1	1	1	1	1
Number of parameters	12	1	2	3	4	5
$\hat{\phi}_1$	—	—	.2977 (3.83)	.2779 (3.43)	.2922 (3.67)	.3518 (1.06)
$\hat{\phi}_2$	—	—	—	.1132 (1.40)	.1674 (2.04)	.1389 (1.18)
$\hat{\phi}_3$	—	—	—	—	−.2408 (−3.02)	−.2438 (−2.92)
$\hat{\theta}_1$	—	—	—	—	—	.0792 (0.23)
$\hat{\theta}_{1,12}$	—	.5509 (7.84)	.5962 (8.47)	.6552 (10.27)	.5917 (8.17)	.5896 (8.11)
$\hat{\delta}$	—	.0330	.0232	.0201	.0258	.0249
Box-Pierce χ^2(20 D.O.F.)	91.22	51.38	38.84	32.50	26.72	29.70
Significant autocorrelations	Lags 1, 2, 11, 12, 13	Lags 1, 5, 13, 14	Lags 3, 5, 18	Lags 3, 5, 18	Lag 10	Lag 10
s	.0217	.0206	.0197	.0197	.0192	.0193

ARIMA model 5 (3,0,0)(0,1,1)^{12}
Number of regular differences = 0
Number of seasonal difference = 1
Number of parameters = 4
ϕ_1 = 0.2792 (3.49)
ϕ_2 = 0.1608 (1.95)
ϕ_3 = -0.2377 (-2.97)
Θ_1 = ----
$\Theta_{1.12}$ = 0.5956 (8.21)
δ = 0.02634
χ_2 = 29.6 (19 DOF)
Significant autocorrelations = Lags 10

FIGURE 8: Model 5 comparison with Case Study

Observations:

1. ARIMA model indicate that all AR & MA coefficients are significant.
2. The chi-square value for 19 D.O.F came as 19.6
3. There exist a little difference observed between the case study values and Minitab values of Model 5.

C2:

Develop and diagnose the same model as their Model 6 ARIMA (3, 0, 1) (0, 1, 1)^{12}and report your results.

a. Perform residual analysis of the model results and comment on your findings

b. Show your results in Table 12.2 format of the original case study and discuss how your Minitab results compare with the Model 6 results of the original case study as well as your own Model 5 results

Solution:

Model 6: $y^*_t = \delta + \phi_1 y^*_{t-1} + \phi_2 y^*_{t-2} + \phi_3 y^*_{t-3} + y^*_{t-12} - \phi_1 y^*_{t-13} - \phi_2 y^*_{t-14} - \phi_3 y^*_{t-15} - \theta_1 \epsilon_{t-1} - \theta_{1,12}\epsilon_{t-12} + \theta_1\theta_{1,12}\epsilon_{t-13} + \epsilon_t$

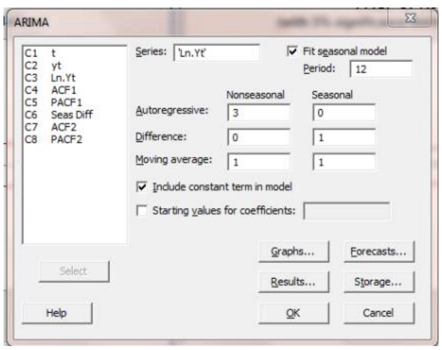
Operators: Seasonal moving-average of order 1

Nonseasonal autoregressive of order 3

Nonseasonal moving-average of order 1

Differencing: $z_t = y^*_t - y^*_{t-12}$

Applying ARIMA Model 6 - (3, 0, 1) (0, 1, 1)^{12} on Ln.Yt



The ACF and PACF graph indicate that lag 10 has significant Autocorrelations or partial autocorrelations (both at very little scale above the threshold for significance) in the ARIMA model

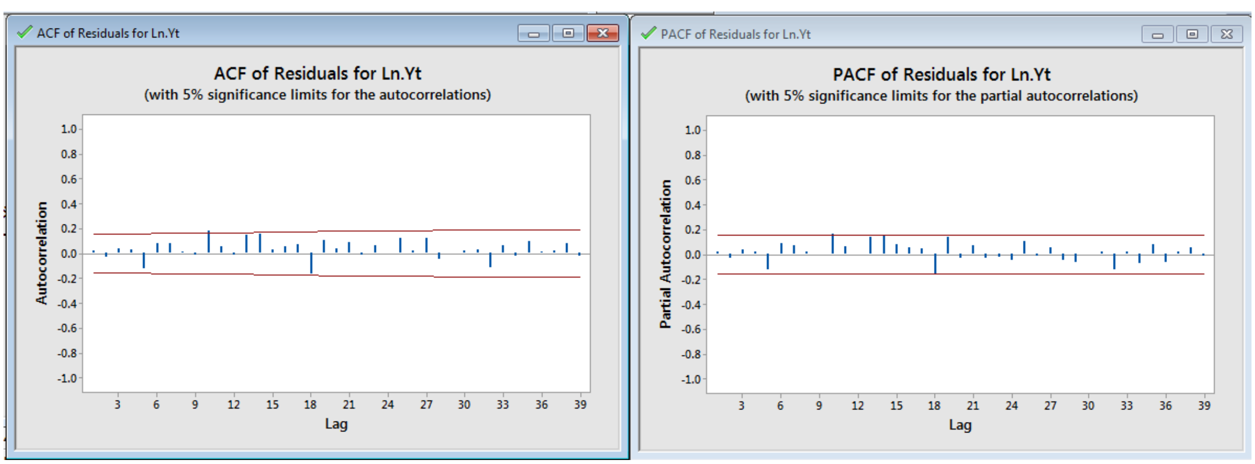


FIGURE 9: ACF & PACF using ARIMA

FIGURE 10: Residual plots using ARIMA

The residual plots satisfy the homoscedasticity, normality and independence assumptions.

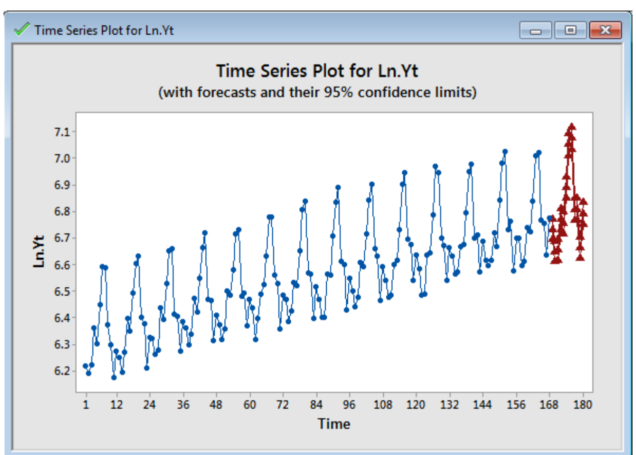


FIGURE 11: Time Series Forecast for 12 periods using ARIMA

ARIMA Model: Ln.Yt

```
Unable to reduce sum of squares any further


Final Estimates of Parameters

Type          Coef   SE Coef      T      P
AR   1      0.3437    0.3328   1.03  0.303
AR   2      0.1404    0.1173   1.20  0.233
AR   3     -0.2447    0.0840  -2.91  0.004
MA   1      0.0734    0.3430   0.21  0.831
SMA  12     0.5840    0.0738   7.91  0.000
Constant 0.0251345 0.0006327 39.72   0.000


Differencing: 0 regular, 1 seasonal of order 12
Number of observations:  Original series 168, after differencing 156
Residuals:    SS =  0.0561541 (backforecasts excluded)
              MS =  0.0003744  DF = 150


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

Lag            12     24     36     48
Chi-Square   11.3   30.2   42.4   49.9
DF              6     18     30     42
P-Value     0.080  0.035  0.066  0.188
```

TABLE 12.2 *Comparison of Models 1–7*

	Model 1	2	3	4	5	6
Number of regular differences	0	0	0	0	0	0
Number of seasonal differences	0	1	1	1	1	1
Number of parameters	12	1	2	3	4	5
$\hat{\phi}_1$	—	—	.2977 (3.83)	.2779 (3.43)	.2922 (3.67)	.3518 (1.06)
$\hat{\phi}_2$	—	—	—	.1132 (1.40)	.1674 (2.04)	.1389 (1.18)
$\hat{\phi}_3$	—	—	—	—	−.2408 (−3.02)	−.2438 (−2.92)
$\hat{\theta}_1$	—	—	—	—	—	.0792 (0.23)
$\hat{\theta}_{1,12}$	—	.5509 (7.84)	.5962 (8.47)	.6552 (10.27)	.5917 (8.17)	.5896 (8.11)
$\hat{\delta}$	—	.0330	.0232	.0201	.0258	.0249
Box-Pierce χ^2(20 D.O.F.)	91.22	51.38	38.84	32.50	26.72	29.70
Significant autocorrelations	Lags 1, 2, 11, 12, 13	Lags 1, 5, 13, 14	Lags 3, 5, 18	Lags 3, 5, 18	Lag 10	Lag 10
s	.0217	.0206	.0197	.0197	.0192	.0193

ARIMA model 6 (3,0,1)(0,1,1)^{12}
Number of regular differences = 0
Number of seasonal difference = 1
Number of parameters = 5
ϕ_1 = 0.3437 (1.03)
ϕ_2 = 0.1404 (1.20)
ϕ_3 = -0.2447 (-2.91)
Θ_1 = 0.0734 (0.21)
$\Theta_{1.12}$ = 0.5840 (7.91)
δ = 0.0251345
χ_2 = 30.2 (18 DOF)
Significant autocorrelations = Lags 10

FIGURE 12: Model 6

comparison with Case Study

Observations:

1. The estimated parameter for `MA 1 0.0734 0.3430 0.21 0.831` indicate that this is insignificant.
2. The Chi-square came as 30.2 for 18 D.O.F, against 29.7 for 20 D.O.F value from the case study.
3. A little difference in estimates noted between the case study and the Minitab estimates for model 6.

TABLE 1 : COMPARISON OF C1 & C2

	Model 5 – C1	Model 6 – C2
Model	ARIMA model 5 (3,0,0)(0,1,1)^{12}	ARIMA model 6 (3,0,1)(0,1,1)^{12}
Number of regular differences	0	0
Number of seasonal difference	1	1

Number of parameters	4	5
ϕ_1 =	0.2792 (3.49)	0.3437 (1.03)
ϕ_2 =	0.1608 (1.95)	0.1404 (1.20)
ϕ_3 =	-0.2377 (-2.97)	-0.2447 (-2.91)
Θ_1 =	----	0.0734 (0.21)
$\Theta_{1.12}$ =	0.5956 (8.21)	0.5840 (7.91)
δ =	0.02634	0.0251345
Chi Square	29.6 (19 DOF)	30.2 (18 DOF)
Significant autocorrelations	Lags 10	Lags 10
S		

Forecast values	`Forecasts from period 168`
	
	`95% Limits`
	`Period Forecast Lower Upper Actual`
	`169 6.73188 6.69415 6.76962`
	`170 6.64894 6.60976 6.68812`
	`171 6.65657 6.61637 6.69677`
	`172 6.77210 6.73162 6.81258`
	`173 6.75753 6.71698 6.79808`
	`174 6.89236 6.85165 6.93307`
	`175 7.05358 7.01287 7.09429`
	`176 7.07709 7.03638 7.11780`
	`177 6.80902 6.76831 6.84974`
	`178 6.80886 6.76814 6.84957`
	`179 6.66404 6.62333 6.70476`
	`180 6.79216 6.75144 6.83288`
		`Forecasts from period 168`
		
		`95% Limits`
		`Period Forecast Lower Upper Actual`
		`169 6.73210 6.69417 6.77003`
		`170 6.64933 6.61004 6.68862`
		`171 6.65655 6.61628 6.69683`
		`172 6.77163 6.73107 6.81219`
		`173 6.75671 6.71604 6.79738`
		`174 6.89164 6.85079 6.93249`
		`175 7.05328 7.01243 7.09413`
		`176 7.07681 7.03596 7.11767`
		`177 6.80867 6.76781 6.84953`
		`178 6.80862 6.76776 6.84949`
		`179 6.66400 6.62314 6.70486`
		`180 6.79239 6.75152 6.83325`

Statistics	`Differencing: 0 regular, 1 seasonal of order 12`
	`Number of observations: Original series 168, after differencing 156`
	`Residuals: SS = 0.0559476 (backforecasts excluded)`
	`MS = 0.0003705 DF = 151`
	
	`Modified Box-Pierce (Ljung-Box) Chi-Square statistic`
	
	`Lag 12 24 36 48`
	`Chi-Square 11.1 29.6 41.6 49.2`
	`DF 7 19 31 43`
	`P-Value 0.136 0.057 0.096 0.238`
		`Differencing: 0 regular, 1 seasonal of order 12`
		`Number of observations: Original series 168, after differencing 156`
		`Residuals: SS = 0.0561541 (backforecasts excluded)`
		`MS = 0.0003744 DF = 150`
		
		`Modified Box-Pierce (Ljung-Box) Chi-Square statistic`
		
		`Lag 12 24 36 48`
		`Chi-Square 11.3 30.2 42.4 49.9`
		`DF 6 18 30 42`
		`P-Value 0.080 0.035 0.066 0.188`

Findings	All coefficients are significant	MA is not significant

C3:

Develop and diagnose one other model based on the results of Models 5 and 6 for the purpose of coming out with a better model.

a. Show ACF (Autocorrelation Function) and PACF (Partial ACF) of the differenced time series used for the model and comment on your findings if it is different from Model 5 differencing used

b. Perform residual analysis of the model results and comment on your findings

c. Show your results in Table 12.2 format of the original case study and discuss how your Minitab results compare with your own Model 5 and 6 results

Solution:

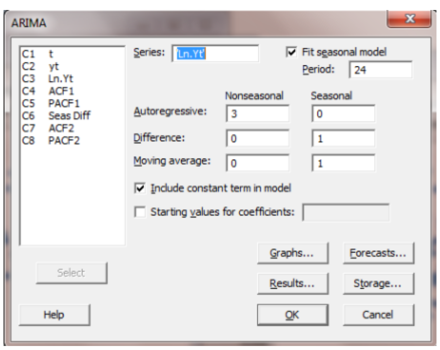
Model 6, has an insignificant MA estimate, leading to believe that MAs are not good to predict this series. Hence, several trials have been tried on Model 5 by introducting MA or with 1, 2 differences or by raising the AR level to 4, 5 on non-seasonal. MAs did not help in any means, so few trials were made using 6 months seasonality and 24 months seasonality. In any case, the following 2 models seem to have better estimates out of all trials.

Two models tried:

1. Modified ARIMA model 5 - $(3,0,0)(0,1,1)^{24}$
2. Modified ARIMA model 5 - $(3,0,0)(0,2,1)^{24}$

Model 1:

24 months seasonality on Model 5, with 1 difference to seasonality.



Residual plots for this model.

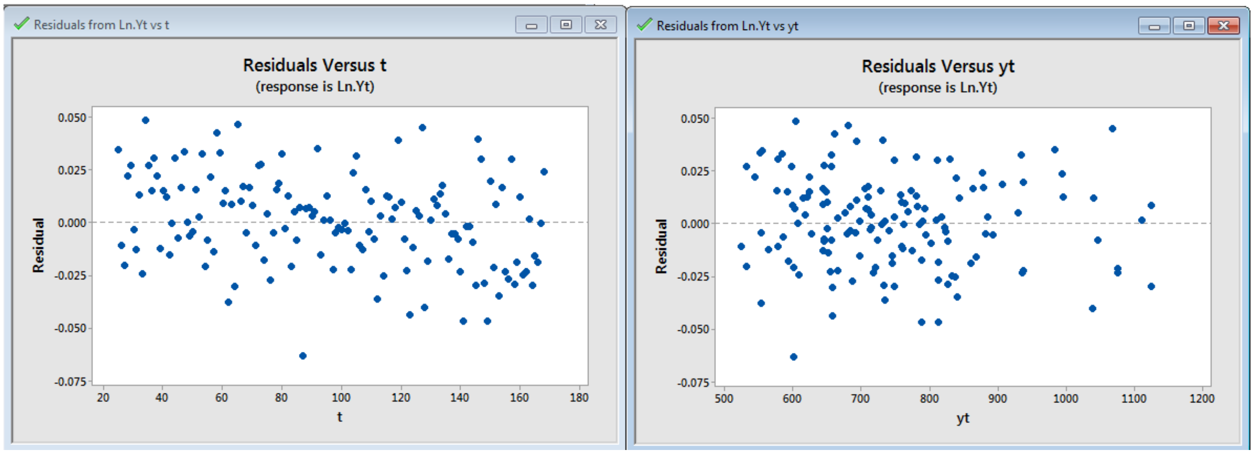


FIGURE 13: Residual plots for Model 1

The ACF & PACF had lag = 12 as significant.

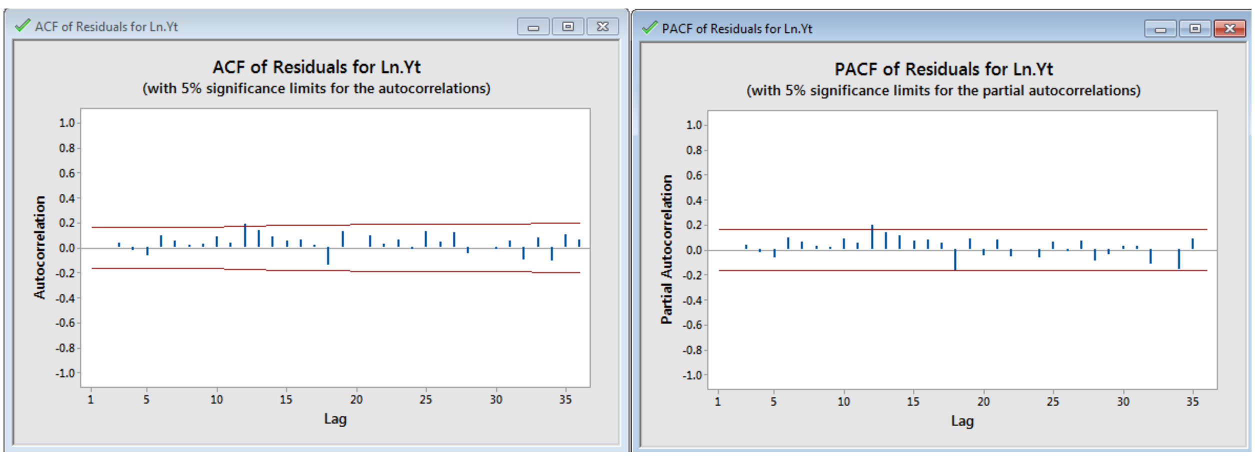


FIGURE 14: Residual plots for Model 1

The residual plot and Q-Q plot show that the assumptions of independence, homoscedasticity and normality assumptions.

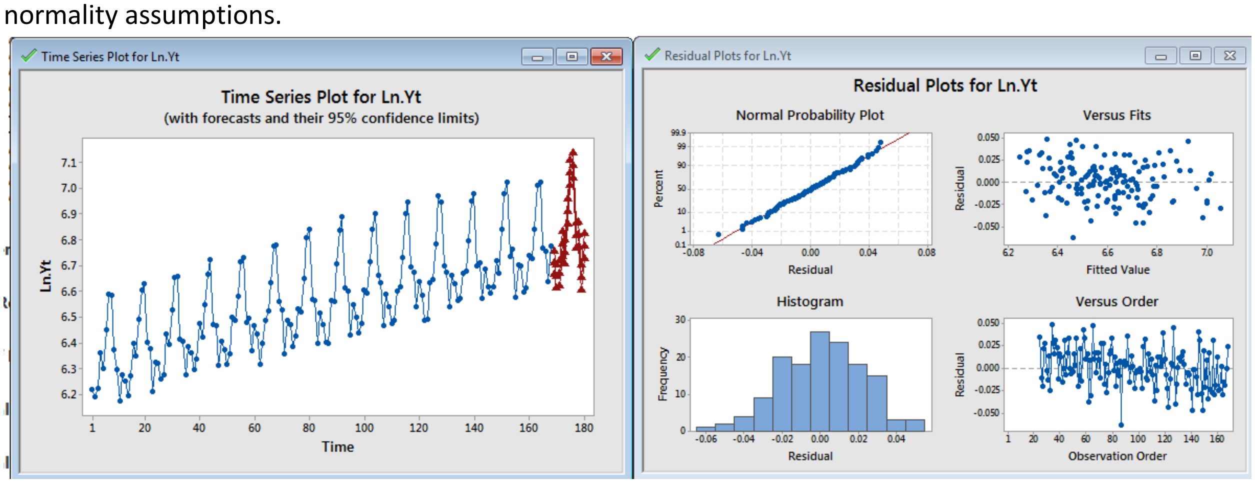


FIGURE 15: Time series plot & 4 residual plots

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
AR	1	0.3578	0.0826	4.33	0.000
AR	2	0.2274	0.0854	2.66	0.009
AR	3	-0.2664	0.0833	-3.20	0.002
SMA	24	0.4328	0.0902	4.80	0.000
Constant		0.045116	0.001196	37.74	0.000

```
Differencing: 0 regular, 1 seasonal of order 24
Number of observations:  Original series 168, after differencing 144
Residuals:    SS =  0.0674266 (backforecasts excluded)
              MS =  0.0004851  DF = 139
```

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

Lag	12	24	36	48
Chi-Square	10.6	24.6	39.7	47.2
DF	7	19	31	43
P-Value	0.158	0.176	0.135	0.303

Forecasts from period 168

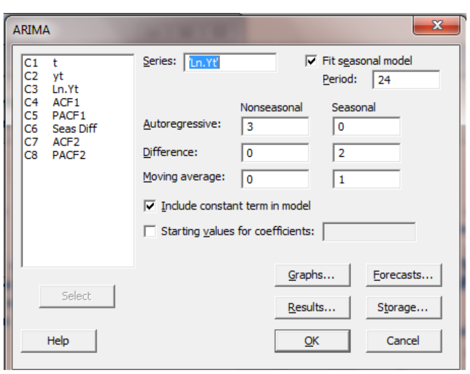
		95% Limits		
Period	Forecast	Lower	Upper	Actual
169	6.71017	6.66699	6.75334	
170	6.65854	6.61269	6.70440	
171	6.66961	6.62125	6.71796	
172	6.78391	6.73549	6.83233	
173	6.75321	6.70476	6.80165	
174	6.91043	6.86171	6.95915	
175	7.06098	7.01223	7.10973	
176	7.08859	7.03982	7.13736	
177	6.81622	6.76745	6.86499	
178	6.82211	6.77334	6.87088	
179	6.65390	6.60512	6.70267	
180	6.77732	6.72854	6.82609	

FINDINGS:

1. Parameter estimates show AR1, AR2, AR3, SMA 24 and constant as significant
2. Chi-square statistic show lag 36 with lowest P-value
3. The model 1 has lag 12 autocorrelation significant. (It seems that it is very difficult to get a ACF, PACF plot with all the ARs within significant line)
4. When the model was modified to have 1 seasonal of order 36, the lowest P-value of Chi-square moved to 48 month season, but ACF & PACF brought 12, 13 beyond the significant line.
5. Therefore, we concluded that this could be one of the best models.

Model 2:

24 Months seasonality with 2 differences, on a Model 5.



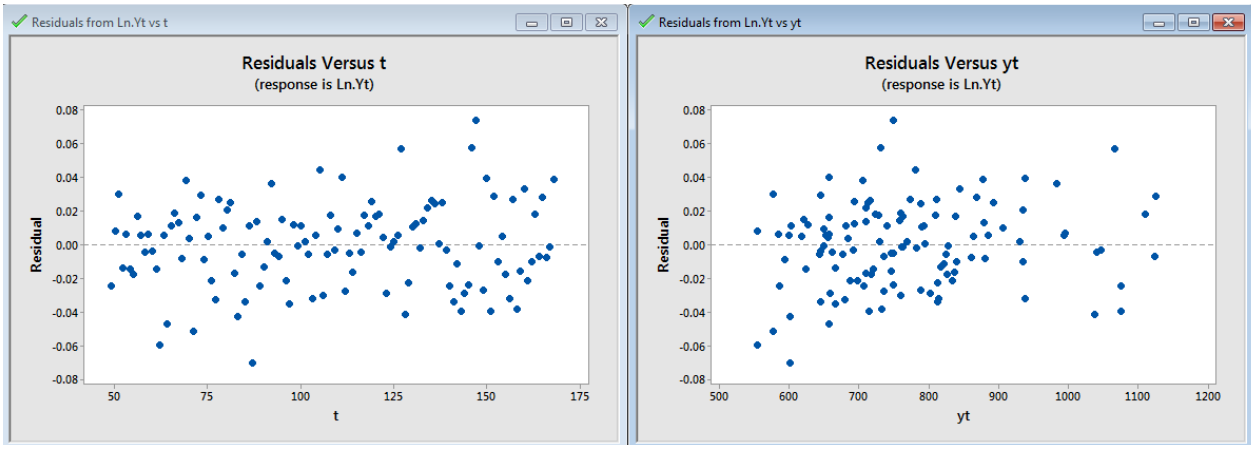


FIGURE 16: Residual plots for Model 2

The ACF & PACF plots show lag 24 as significant.

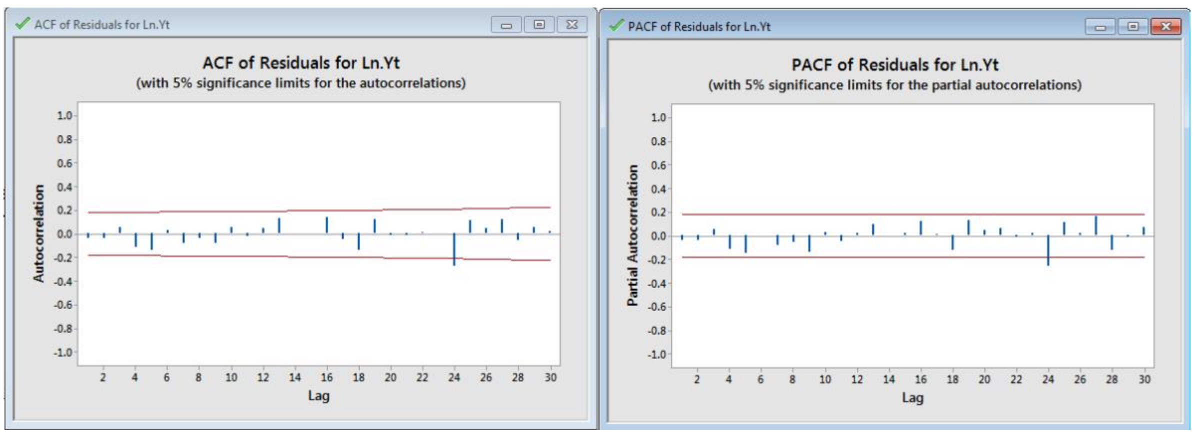


FIGURE 17: Residual plots for Model 2

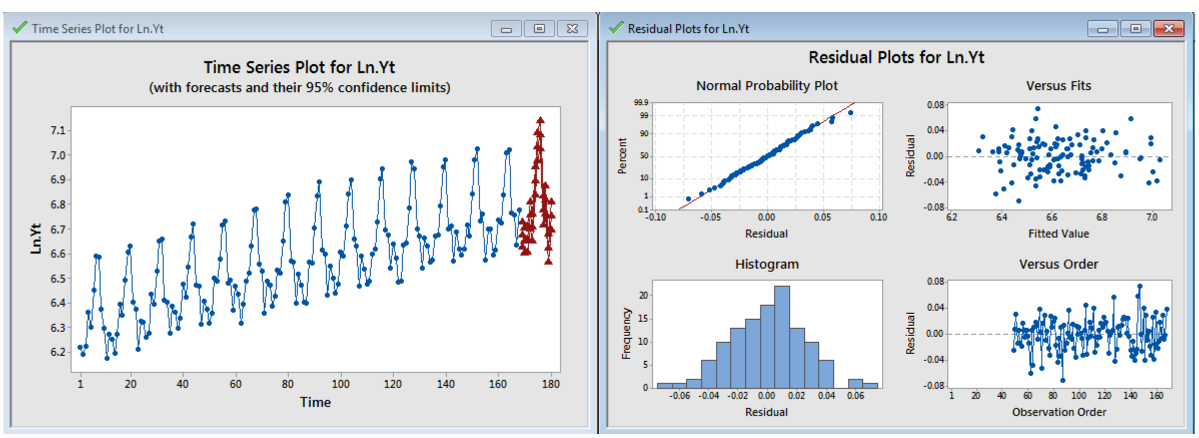


FIGURE 18: Time series plot & 4 residual plot

The residual plot and Q-Q plot show that the assumptions of independence, homoscedasticity and normality assumptions

```
Final Estimates of Parameters

Type           Coef     SE Coef       T      P
AR   1       0.2302      0.0841    2.74  0.007
AR   2       0.1860      0.0851    2.19  0.031
AR   3      -0.4256      0.0839   -5.07  0.000
SMA  24      0.7885      0.1072    7.35  0.000
Constant  -0.0048421  0.0007941  -6.10  0.000


Differencing: 0 regular, 2 seasonal of order 24
Number of observations:  Original series 168, after differencing 120
Residuals:    SS =  0.0762358 (backforecasts excluded)
              MS =  0.0006629  DF = 115


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

Lag             12      24      36      48
Chi-Square     8.0    30.3    44.6    58.9
DF               7      19      31      43
P-Value      0.330   0.048   0.054   0.054
```

```
Forecasts from period 168

                        95% Limits
Period  Forecast     Lower     Upper  Actual
   169   6.67663   6.62615   6.72710
   170   6.65243   6.60064   6.70423
   171   6.65809   6.60491   6.71128
   172   6.75595   6.70026   6.81164
   173   6.70683   6.65075   6.76290
   174   6.89529   6.83838   6.95220
   175   7.03620   6.97917   7.09322
   176   7.08351   7.02646   7.14056
   177   6.77570   6.71842   6.83299
   178   6.81986   6.76258   6.87715
   179   6.62369   6.56640   6.68097
   180   6.75351   6.69618   6.81083
```

FINDINGS:

1. The parameter estimates show AR1, AR2, AR3, SMA24 and constants as significant
2. Chi-square statistic show that Lag 24 with lowest P-value.
3. ACF & PACF had lag 24 as significant.

	24mnth, 1 diff	24mnth, 2 diff
Number of parameters	4	4
ϕ_1 =	0.3578 (4.33)	0.2302 (2.74)
ϕ_2 =	0.2274 (2.66)	0.1860 (2.19)
ϕ_3 =	-0.2664 (-3.20)	-0.4256 (-5.07)
Θ_1 =	-	-
$\Theta_{1.12}$ =	0.4328 (4.80)	0.7885 (7.35)
δ =	0.045116	-0.004841
Chi Square	24.6 (19 DOF)	30.3 (19 DOF)
Significant autocorrelations	Lags 12	Lags 24
S		

TABLE 12.2 *Comparison of Models 1–7*

	Model 1	2	3	4	5	6
Number of regular differences	0	0	0	0	0	0
Number of seasonal differences	0	1	1	1	1	1
Number of parameters	12	1	2	3	4	5
$\hat{\phi}_1$	—	—	.2977 (3.83)	.2779 (3.43)	.2922 (3.67)	.3518 (1.06)
$\hat{\phi}_2$	—	—	—	.1132 (1.40)	.1674 (2.04)	.1389 (1.18)
$\hat{\phi}_3$	—	—	—	—	−.2408 (−3.02)	−.2438 (−2.92)
$\hat{\theta}_1$	—	—	—	—	—	.0792 (0.23)
$\hat{\theta}_{1,12}$	—	.5509 (7.84)	.5962 (8.47)	.6552 (10.27)	.5917 (8.17)	.5896 (8.11)
$\hat{\delta}$	—	.0330	.0232	.0201	.0258	.0249
Box-Pierce χ^2(20 D.O.F.)	91.22	51.38	38.84	32.50	26.72	29.70
Significant autocorrelations	Lags 1, 2, 11, 12, 13	Lags 1, 5, 13, 14	Lags 3, 5, 18	Lags 3, 5, 18	Lag 10	Lag 10
s	.0217	.0206	.0197	.0197	.0192	.0193

TABLE 2 : COMPARISON OF C1, C2 & C3

Findings:

1. Generally, MA (theta 1) is not proving significant, or going to influence the time series forecasting.
2. The new models were very almost to Model 5, but when considering the Chi-square value for D.O.F 19, the 24th mnth, 2 diff model could be a better one. Model 6, may not be a good choice as the MA estimates showed up insignificant.

	24mnth, 1 diff	24mnth, 2 diff	Model 5	Model 6
Number of parameters	4	4	4	5
ϕ_1 =	0.3578 (4.33)	0.2302 (2.74)	0.2792 (3.49)	0.3437 (1.03)
ϕ_2 =	0.2274 (2.66)	0.1860 (2.19)	0.1608 (1.95)	0.1404 (1.20)
ϕ_3 =	-0.2664 (-3.20)	-0.4256 (-5.07)	-0.2377 (-2.97)	-0.2447 (-2.91)
Θ_1 =	-	-	----	0.0734 (0.21)
$\Theta_{1.12}$ =	0.4328 (4.80)	0.7885 (7.35)	0.5956 (8.21)	0.5840 (7.91)
δ =	0.045116	-0.004841	0.02634	0.0251345
Chi Square	24.6 (19 DOF)	30.3 (19 DOF)	29.6 (19 DOF)	30.2 (18 DOF)
Significant autocorrelations	Lags 12	Lags 24	Lags 10	Lags 10