Thomas Pattara
STAT 560 Time Series Analysis

1. Analyze the attached final_data1 by three methods: ARIMA, exponential smoothing, and linear regression. The first column of final_data1 is the year; the second column is the coal production.

1.1 ARIMA model

1.1.1 *Fit an ARIMA model to this time series, excluding the last 10 observations. If you are selecting the best model from multiple models, please do NOT include the selection procedures, just show the summary of the best model, and by saying that this model has the least AIC (AIC=***) comparing with others (AIC=*** for ARIMA(p1,d1,q1), AIC=*** for ARIMA(p2,d2,q2), etc.).*

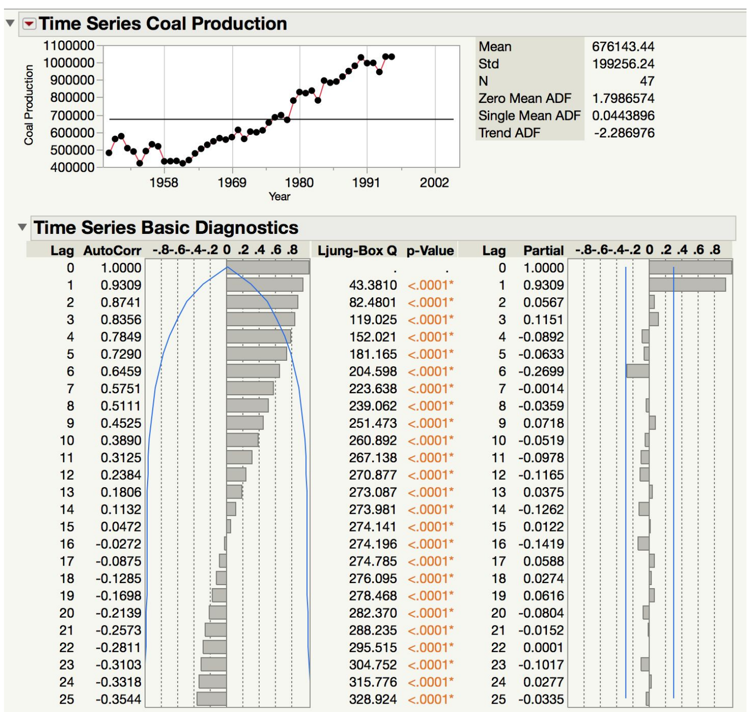


Figure 1 : Time Series Analysis of Coal Production data excluding last 10 observations.

Model: I(1)

Model Summary

DF	45	Stable	Yes
Sum of Squared Errors	8.5849e+10	Invertible	Yes
Variance Estimate	1907750003		
Standard Deviation	43677.7976		
Akaike's 'A' Information Criterion	1114.51407		
Schwarz's Bayesian Criterion	1116.34271		
RSquare	0.95301013		
RSquare Adj	0.95301013		
MAPE	5.18300434		
MAE	32532.311		
-2LogLikelihood	1112.51407		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
Intercept	0	12008.777	6369.819	1.89	0.0659	12008.7771

Forecast

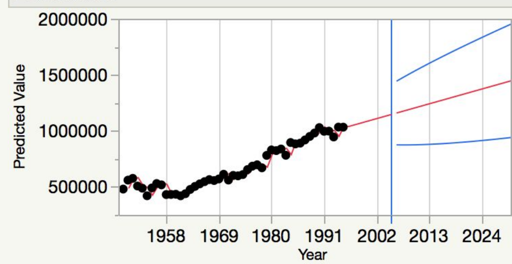


Figure 2: Summary of best fit

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.1568	1.2065	0.2720	1	-0.1568
2	-0.0506	1.3352	0.5129	2	-0.0771
3	-0.1819	3.0349	0.3863	3	-0.2089
4	-0.0023	3.0352	0.5520	4	-0.0803
5	0.2087	5.3814	0.3711	5	0.1761
6	0.0136	5.3916	0.4946	6	0.0461
7	-0.0036	5.3924	0.6122	7	0.0274
8	-0.0840	5.8024	0.6694	8	-0.0030
9	-0.0492	5.9471	0.7452	9	-0.0510
10	0.2325	9.2640	0.5072	10	0.1943
11	-0.1825	11.3642	0.4133	11	-0.1644
12	0.0149	11.3786	0.4968	12	-0.0476
13	-0.0581	11.6043	0.5603	13	0.0075
14	0.0437	11.7361	0.6275	14	-0.0051
15	0.1495	13.3289	0.5769	15	0.0963
16	-0.1155	14.3113	0.5755	16	-0.0478
17	0.0390	14.4272	0.6367	17	0.0494
18	-0.1098	15.3782	0.6359	18	-0.0331
19	0.0404	15.5119	0.6896	19	-0.0158
20	0.0997	16.3554	0.6943	20	0.0208
21	-0.1810	19.2485	0.5692	21	-0.1421
22	0.0031	19.2494	0.6299	22	-0.0793
23	-0.1246	20.7388	0.5970	23	-0.1159
24	0.1511	23.0302	0.5180	24	0.0613
25	0.0598	23.4060	0.5539	25	0.0277

Figure 3: ACF and PACF after ARIMA(0,1,0) modelling

Comparing various ARIMA models on the Coal production data (excluding last 10 observations), I found ARIMA(0,1,0) to be the best model. From Figure 1, since ACF has exponential decay and PACF cuts off at lag 1, we can say that AR would fit well for the data. Comparing the AIC's of different model, AR(1) has 1147.37 , first differencing I(1) has 1114.51 and ARI(1,1) 1115.001. Hence clearly, ARIMA(0,1,0) has the lowest AIC and should be the best model for the data.

Therefore, $(1-B)y_t = \delta + \varepsilon_t$

$= 12008.77 + \varepsilon_t$.

1.1.2 *Forecast the last 10 observations. Calculate the mean squared forecast error (MSE) and the mean absolute percent forecast error (MAPE). Please list only the 10 forecasts in a line, and MSE and MAPE in another line, do NOT include any other outputs.*

MSE = 1009498466

Table 1: MSE Calculation

	Year	Actual Coal Production	Predicted Coal Production	(Actual-Predicted)^2
1	1996	1063855.513	1044982.549	356188766.4
2	1997	1089931.788	1056991.326	1085074017
3	1998	1117535.167	1069000.103	2355652399
4	1999	1100431.428	1081008.881	377235347.5
5	2000	1073611.561	1093017.658	376596589.1
6	2001	1127688.806	1105026.435	513583063.9
7	2002	1094283.061	1117035.212	517660375.1
8	2003	1071752.573	1129043.989	3282306370
9	2004	1112098.87	1141052.766	838328111
10	2005	1133253.474	1153061.544	392359617.3
11	•	•		•
12	•	•	MSE	1009498466

MAPE = 2.64976837

Table 2: MAPE calculation

	Year	Actual Coal Production	Predicted Coal ...	|Actual-Predicted|	Absolute Percentage ...
1	1996	1063855.513	1044982.549	18872.9639	1.774015707
2	1997	1089931.788	1056991.326	32940.4617	3.022249838
3	1998	1117535.167	1069000.103	48535.0636	4.343045752
4	1999	1100431.428	1081008.881	19422.5474	1.764993884
5	2000	1073611.561	1093017.658	19406.0967	1.807552881
6	2001	1127688.806	1105026.435	22662.3711	2.009629871
7	2002	1094283.061	1117035.212	22752.151	2.079183331
8	2003	1071752.573	1129043.989	57291.4162	5.345582333
9	2004	1112098.87	1141052.766	28953.8963	2.603536168
10	2005	1133253.474	1153061.544	19808.0695	1.747894002
11	•	•	•		•
12	•	•	•	MAPE	2.649768377

Table 3: Forecast of last 10 observation

	Year	Predicted Coal Production
48	1996	1044982.5491
49	1997	1056991.3263
50	1998	1069000.1034
51	1999	1081008.8806
52	2000	1093017.6577
53	2001	1105026.4349
54	2002	1117035.212
55	2003	1129043.9892
56	2004	1141052.7663
57	2005	1153061.5435

1.1.3 *Show how to obtain the 95% prediction intervals (PIs) for the forecasts in 1.1.2). Your answer should include how to obtain the linear filter $\psi$is for $yt+\tau$, the formula of calculating PIs and the calculated intervals.*

$$var[e_T(\tau)] = \sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2$$

$$= \sigma^2(\tau), \ \tau = 1,2,3\ldots$$

Prediction interval for $y_{T+\tau} = \hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} \sqrt{var[e_T(\tau)]}$

$$= \hat{y}_{T+\tau}(T) \pm Z_{\alpha/2} \sqrt{\sigma^2(\tau)}$$

Table 4: 95% prediction intervals (PIs) for the forecasts in 1.1.2

	Upper PI (0.95) Coal Production	Lower PI (0.95) Coal Production
48	1130589.4594	959375.63894
49	1178057.7797	935924.87285
50	1217275.6214	920724.58548
51	1252222.701	909795.06018
52	1284440.5283	901594.78717
53	1314719.6834	895333.18642
54	1343529.807	890540.61712
55	1371176.8961	886911.08229
56	1397873.497	884232.03571
57	1423774.3632	882348.72378

1.2 Exponential Smoothing method

1.2.1 *Use an exponential smoothing with the optimum value of λ to smooth the data, excluding the last 10 observations. Let the range of λ be [0.3, 1].*

Model: Double (Brown) Exponential Smoothing

Model Summary

DF	44	Stable	Yes
Sum of Squared Errors	8.8774e+10	Invertible	Yes
Variance Estimate	2017583333		
Standard Deviation	44917.517		
Akaike's 'A' Information Criterion	1095.3864		
Schwarz's Bayesian Criterion	1097.19306		
RSquare	0.94376042		
RSquare Adj	0.94376042		
MAPE	6.18291177		
MAE	37411.4116		
-2LogLikelihood	1093.3864		

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.31239095	0.0679824	4.60	<.0001*

Forecast

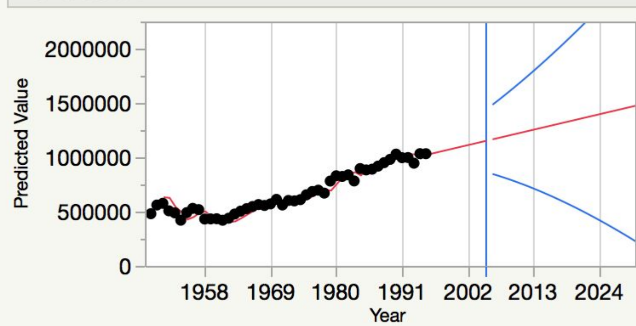
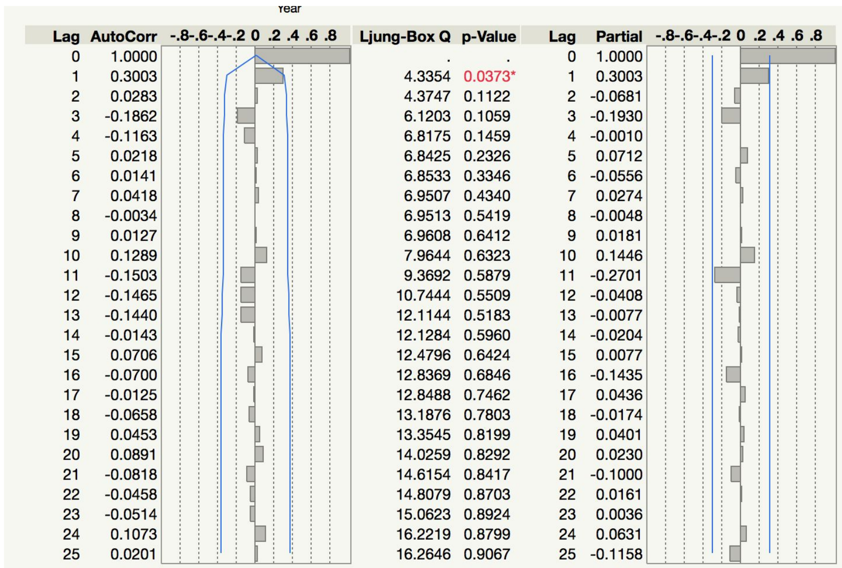


Figure 4: Exponential smoothing excluding the last 10 observations.

We find the optimum value of λ to smooth the data to be 0.31239



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.3003	4.3354	0.0373*	1	0.3003
2	0.0283	4.3747	0.1122	2	-0.0681
3	-0.1862	6.1203	0.1059	3	-0.1930
4	-0.1163	6.8175	0.1459	4	-0.0010
5	0.0218	6.8425	0.2326	5	0.0712
6	0.0141	6.8533	0.3346	6	-0.0556
7	0.0418	6.9507	0.4340	7	0.0274
8	-0.0034	6.9513	0.5419	8	-0.0048
9	0.0127	6.9608	0.6412	9	0.0181
10	0.1289	7.9644	0.6323	10	0.1446
11	-0.1503	9.3692	0.5879	11	-0.2701
12	-0.1465	10.7444	0.5509	12	-0.0408
13	-0.1440	12.1144	0.5183	13	-0.0077
14	-0.0143	12.1284	0.5960	14	-0.0204
15	0.0706	12.4796	0.6424	15	0.0077
16	-0.0700	12.8369	0.6846	16	-0.1435
17	-0.0125	12.8488	0.7462	17	0.0436
18	-0.0658	13.1876	0.7803	18	-0.0174
19	0.0453	13.3545	0.8199	19	0.0401
20	0.0891	14.0259	0.8292	20	0.0230
21	-0.0818	14.6154	0.8417	21	-0.1000
22	-0.0458	14.8079	0.8703	22	0.0161
23	-0.0514	15.0623	0.8924	23	0.0036
24	0.1073	16.2219	0.8799	24	0.0631
25	0.0201	16.2646	0.9067	25	-0.1158

Figure 5: ACF and PACF after exponential smoothing.

***1.2.2** Forecast the last 10 observations. Calculate the mean squared error (MSE) and the mean absolute percent forecast error (MAPE). Same requirement as question 1.1.2.*

MSE = 1154974117

	Year	Actual Coal Production	Predicted Coal Production	(Actual-Predicted)^2
1	1996	1063855.513	1043681.814	406978115.2
2	1997	1089931.788	1056604.18	1110729448
3	1998	1117535.167	1069526.546	2304827719
4	1999	1100431.428	1082448.911	323370903.3
5	2000	1073611.561	1095371.277	473485244.8
6	2001	1127688.806	1108293.643	376172355.6
7	2002	1094283.061	1121216.009	725383661
8	2003	1071752.573	1134138.374	3891988179
9	2004	1112098.87	1147060.74	1222332340
10	2005	1133253.474	1159983.106	714473200.1
11	•	•		•
12	•	•	MSE	1154974117

Table 5: MSE calculation

MAPE = 2.841539722

Table 6: MAPE calculation

	Year	Actual Coal Production	Predicted Coal Production	| Actual-Predicted |	Absolute Percentage ...
1	1996	1063855.513	1043681.814	20173.6986	1.896281812
2	1997	1089931.788	1056604.18	33327.6079	3.057770061
3	1998	1117535.167	1069526.546	48008.6213	4.295938304
4	1999	1100431.428	1082448.911	17982.5166	1.634133317
5	2000	1073611.561	1095371.277	21759.7161	2.026777364
6	2001	1127688.806	1108293.643	19395.1632	1.719903851
7	2002	1094283.061	1121216.009	26932.9475	2.461241379
8	2003	1071752.573	1134138.374	62385.8011	5.820914516
9	2004	1112098.87	1147060.74	34961.8698	3.143773521
10	2005	1133253.474	1159983.106	26729.6315	2.358663098
11	•	•			•
12	•	•		MAPE	2.841539722

Table 7: Forecast of last 10 observation

	Year	Actual Coal Production	Predicted Coal Production
1	1996	1063855.513	1043681.814
2	1997	1089931.788	1056604.18
3	1998	1117535.167	1069526.546
4	1999	1100431.428	1082448.911
5	2000	1073611.561	1095371.277
6	2001	1127688.806	1108293.643
7	2002	1094283.061	1121216.009
8	2003	1071752.573	1134138.374
9	2004	1112098.87	1147060.74
10	2005	1133253.474	1159983.106

1.2.3 *Show how to obtain the 95% prediction intervals for the forecasts in 1.2.2). Your answer should include the formula of calculating PIs and the calculated intervals.*

100(1 − α/2) percent prediction interval for any Coal Production at time τ is

$$\left(2+\frac{\lambda}{1-\lambda}\tau\right)\hat{y}_T^{(1)}-\left(1+\frac{\lambda}{1-\lambda}\tau\right)\hat{y}_T^{(2)}\pm Z_{\alpha/2}\frac{c_\tau}{c_1}\hat{\sigma}_e,$$

where

$$c_i^2=1+\frac{\lambda}{(2-\lambda)^3}[(10-14\lambda+5\lambda^2)+2i\lambda(4-3\lambda)+2i^2\lambda^2].$$

Table 8: 95% prediction intervals for the forecasts in 1.2.2

	Upper PI (0.95) Coal Production	Lower PI (0.95) Coal Production
48	1131718.53	955645.09881
49	1160411.0945	952797.26562
50	1191264.8312	947788.26028
51	1223980.1426	940917.68023
52	1258331.8103	932410.74392
53	1294151.0348	922436.25071
54	1331309.3591	911122.65784
55	1369706.8246	898569.92367
56	1409263.7698	884857.70978
57	1449915.2362	870050.9748

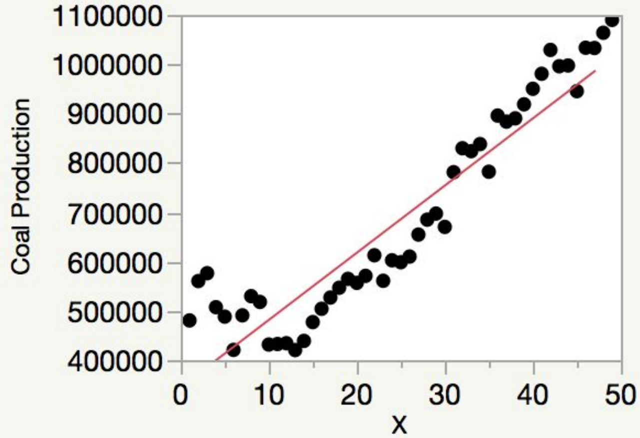
1.3 Linear regression model

1.3.1 *Use a linear regression model to fit the data, excluding the last 10 observations.*

Coal Production (y_t) = 349159.7 + 13624.322(X) , where X is the row order.

Response Coal Production

Regression Plot



Summary of Fit

RSquare	0.860248
RSquare Adj	0.857143
Root Mean Square Error	76126.06
Mean of Response	676143.4
Observations (or Sum Wgts)	47

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.6053e+12	1.605e+12	276.9994
Error	45	2.6078e+11	5.7952e+9	Prob > F
C. Total	46	1.866e+12		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	349159.7	22567.44	15.47	<.0001*
X	13624.322	818.6071	16.64	<.0001*

Figure 6: Linear fit excluding the last 10 observations.

1.3.2 *Forecast the last 10 observations. Calculate the mean squared forecast error (MSE) and the mean absolute percent forecast error (MAPE).*

MSE = 2412579187

Table 9: MSE calculation

	Year	Actual Coal Production	Predicted Coal Production	(Actual - Predicted)^2
1	1996	1063855.513	1003127.168	3687931850
2	1997	1089931.788	1016751.491	5355355942
3	1998	1117535.167	1030375.813	7596753059
4	1999	1100431.428	1044000.135	3184490852
5	2000	1073611.561	1057624.457	255587497.5
6	2001	1127688.806	1071248.779	3185476636
7	2002	1094283.061	1084873.101	88547341.56
8	2003	1071752.573	1098497.423	715287022.9
9	2004	1112098.87	1112121.746	523.2930754
10	2005	1133253.474	1125746.068	56361149.35
11	•	•		•
12	•	•	MSE	2412579187

MAPE = 3.586

Table 10: MAPE calculation

	Year	Actual Coal Production	redicted Coal Production	| Actual - Predicted |	Absolute Percentage ...
1	1996	1063855.513	1003127.168	60728.3447	5.708326362
2	1997	1089931.788	1016751.491	73180.2975	6.714208935
3	1998	1117535.167	1030375.813	87159.3544	7.799249364
4	1999	1100431.428	1044000.135	56431.2932	5.128106283
5	2000	1073611.561	1057624.457	15987.1041	1.489095748
6	2001	1127688.806	1071248.779	56440.0269	5.00492925
7	2002	1094283.061	1084873.101	9409.9597	0.859920073
8	2003	1071752.573	1098497.423	26744.8504	2.495431415
9	2004	1112098.87	1112121.746	22.8756	0.002056975
10	2005	1133253.474	1125746.068	7507.4063	0.662464883
11	•	•	•		•
12	•	•	•	MAPE	3.586378929

Table 11: Forecast of last 10 observations

	Year	Coal Production	Predicted Coal ...
48	1996	1063855.513	1003127.1683
49	1997	1089931.788	1016751.4905
50	1998	1117535.167	1030375.8126
51	1999	1100431.428	1044000.1348
52	2000	1073611.561	1057624.4569
53	2001	1127688.806	1071248.7791
54	2002	1094283.061	1084873.1013
55	2003	1071752.573	1098497.4234
56	2004	1112098.87	1112121.7456
57	2005	1133253.474	1125746.0677

1.3.3 *Show how to obtain the 95% prediction intervals for the forecasts in 1.3.2). Your answer should include the formula of calculating PIs and the calculated intervals.*

Prediction interval =

$$\hat{y} \pm t_{n-2} s_y \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}}$$

Table 12: 95% prediction intervals for the forecasts in 1.3.2

	Year	Coal Production	Predicted Coal ...	Lower 95% Indiv Coal ...	Upper 95% Indiv Coal ...
48	1996	1063855.513	1003127.1683	843206.01157	1163048.325
49	1997	1089931.788	1016751.4905	856414.41324	1177088.5677
50	1998	1117535.167	1030375.8126	869606.98136	1191144.6439
51	1999	1100431.428	1044000.1348	882783.84313	1205216.4264
52	2000	1073611.561	1057624.4569	895945.12895	1219303.7849
53	2001	1127688.806	1071248.7791	909090.97226	1233406.5859
54	2002	1094283.061	1084873.1013	922221.50935	1247524.6932
55	2003	1071752.573	1098497.4234	935336.87917	1261657.9677
56	2004	1112098.87	1112121.7456	948437.22321	1275806.268
57	2005	1133253.474	1125746.0677	961522.68529	1289969.4502

2. *The data final_data_560 contains 7 years of monthly data on the number of airline miles flown in the United Kingdom. This is a seasonal data. Develop an appropriate ARIMA model (Same requirement as question 1.1.1). And calculate the prediction intervals for the eighth year (next 12 months).*

Model: Seasonal ARIMA(0, 1, 1)(0, 1, 1)12

Model Summary

DF	68	Stable	Yes
Sum of Squared Errors	43.6184603	Invertible	Yes
Variance Estimate	0.64144795		
Standard Deviation	0.80090445		
Akaike's 'A' Information Criterion	179.947612		
Schwarz's Bayesian Criterion	186.735651		
RSquare	0.84604664		
RSquare Adj	0.8415186		
MAPE	5.51406149		
MAE	0.60373546		
-2LogLikelihood	173.947612		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
MA1,1	1	1	0.80010946	0.1035678	7.73	<.0001*
MA2,12	2	12	0.62561362	0.1259218	4.97	<.0001*
Intercept	1	0	0.00755811	0.0112814	0.67	0.5052

Constant Estimate	Mu
0.00755811	0.00755811

Figure 12: Seasonal ARIMA (0,1,1) X $(0,1,1)_{12}$ model summary

Comparing to other models, this model has the lowest AIC. SARIMA(0,1,1) X $(0,1,1)_{12}$ has AIC value 179.94 , SARIMA(0,1,0) X $(0,1,0)_{12}$ has 213.54 and SARIMA(0,0,1) X $(0,0,1)_{12}$ has 271.24. Therefore, we select SARIMA(0,1,1) X $(0,1,1)_{12}$ and the fitted model is:

$$(1-B)\,(1-B^{12})\,y_t = (1 + 0.8001B)\,(1 + 0.62561B^{12})\,\varepsilon_t$$

Table 13: Prediction intervals for the eighth year (next 12 months)

	Upper PI (0.95) Miles, in Millions	Lower PI (0.95) Miles, in Millions
84	14.350268257	11.210780485
85	13.341378799	10.201891026
86	12.996702761	9.7951082562
87	15.527269509	12.264750349
88	14.955293895	11.632967125
89	16.018176385	12.637099772
90	17.592476403	14.1536535
91	17.045730338	13.550114962
92	17.360075528	13.808575739
93	18.801034816	15.194516459
94	15.722654832	12.061944716
95	14.492603191	10.778491932
96	15.746765374	11.980009961

Table 14: Predicted values for next 12 months

	Actual Miles, in Millions	Month	Predicted Miles, in Millions
85	•	12/1970	11.771634912
86	•	01/1971	11.395905508
87	•	03/1971	13.896009929
88	•	03/1971	13.29413051
89	•	04/1971	14.327638079
90	•	05/1971	15.873064951
91	•	06/1971	15.29792265
92	•	07/1971	15.584325633
93	•	08/1971	16.997775637
94	•	09/1971	13.892299774
95	•	10/1971	12.635547561
96	•	11/1971	13.863387668