STAT 4280/6280 **Project 3** **Name__Minjiao Yang_____**

Window of submission: December 2 to December 4, 2019 (by 11:59pm)

Milk contamination incident in Oahu, Hawaii:

In April 1982 (in fact, towards the end of March 1982), Hawaii's State Health Department announced that local dairy products contained 15 times the official acceptable adult level of *heptachlor* pesticide. Eight recalls of milk and dairy products on Oahu were announced during April of 1982. The press provided a tremendous amount of coverage on the contamination crisis and milk quality. Their reports contained both negative and positive information regarding milk quality. This impacted the sales of milk, so much so that the average daily milk consumption dropped considerably in April 1982 compared with the previous months. Although the government and industry assured the public that available milk was safe after each recall and that there was no evidence to suggest further contamination, consumers became reluctant to buy locally produced milk. Milk consumption (measured in sales of milk) **returned slowly to normal levels** after the incident; see Liu et al. (1998) (paper attached) for more details.

I have attached a SAS data, which gives information on the sales, date, month, and year.

1) Give your best time series model for data prior to intervention.
2) Build at least two or three intervention models for the data up to April 1983. Use the intervention models to forecast the remaining three months (May 1983-July 1983) and compute the MAPE values for each. Then, select the best intervention model.
3) Give the forecast plot of the entire data, and also the forecast plot from January 1982 to July 1983.
4) State a conclusion summarizing (1) – (3) in about 500 words.

Instructions for handing in the final:

- **Upload your project in eLC in the Project 3 folder.**
- Attach this sheet to the project with your name in capitals
- **The final should have an Introduction. State your final intervention model on the first page of your project right after the introduction with AIC, BIC, and MAPE values for your final model.**
- **Then, give a detailed description not exceeding 6 pages about how you arrived at the conclusion; an overall summary table should be included.**
- Copy Paste the entire SAS program in an Appendix. Give plots to support your answers in an Appendix. Upload the Appendix in eLC folder.

Table of Contents

Introduction

In April 1982, Hawaii's State Health Department informed that the dairy products contained 15 times the official acceptable adult level of *heptachlor* pesticide. Eight recalls of milk and dairy products on Oahu were announced during April of 1982. The press provided a tremendous amount of coverage on the contamination crisis and milk quality. Their reports contained both negative and positive information regarding milk quality. This incident impacted the sales of milk so that the average daily milk consumption dropped rapidly in April 1982. Although the government and industry try to convince the public that available milk was safe, consumers became reluctant to buy locally produced milk. Milk consumption (measured in sales of milk) **returned slowly to normal levels** after the incident;

(reference: project instruction, sorry for not having time to write my own introduction, spending too much time on the analysis)

Summary

In April 1982, the incident of Milk containment impacts the sales of Milk in Oahu of Hawaii. The consumption of milk drops rapidly after April 1982. To analysis this intervention time series data of Milk consumption in Oahu, we first analyze the data of pre-intervention, which are the milk sales before the incident. And we find out the best time series model for the pre-intervention data is the seasonal model $((0,2),0,1)\text{x}(1,0,0)_{12}$. Next, we use the model from pre-intervention refit the original intervention data without the last three months of 1983 (May to July) and find out the best model consist of both Pulse and Step function since the sales of milk drops and produces a step change start at April 1982. So the final intervention model for our time series data is,

$$y_t = \frac{(\omega_0 - \omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2) P_t + \frac{\mu + (1-\theta_1 B) w_t}{(1-\varphi_1 B^2)(1-\Phi_{12} B^{12})}$$

Where I_t=1 for a date after April 1982 and 0 before the date, and P_t=1 for April 1982 and elsewhere. With AIC 2101, BIC 2122 and MAPE 0.014421. However, the model does not yield the least AIC and BIC, It has the smallest MAPE, and the smaller percentage error of the model indicates the higher prediction accuracy. Which is proven by the forecast plot of the model.

So based on the best intervention model, we predict the sales of milk from Feb.1977 to July 1983 and sales from January 1982 to July 1983. And we conclude that the sales of milk in Oahu drop rapidly after the milk containment incident in April 1982 and reach the minimum sale in May of 1982, then it raises again and slowly back to normal after August of 1983.

Pre-intervention Data Analysis

The period from Feb.1977 to March.1982 is assumed to be free of intervention effects and is used to estimate the noise model.

Stationary and Transformation of Pre-intervention Data

The pre-intervention data set contains 62 monthly records of milk consumption in Oahu from Feb. 1977 to March 1982.

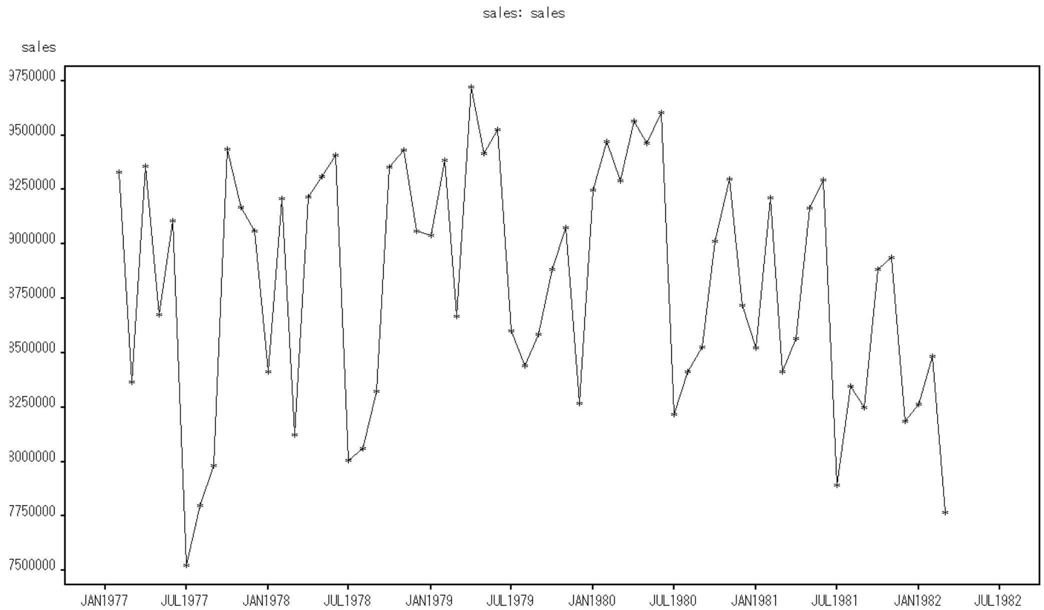

Figure 1. Times series graph of monthly milk consumption from Feb.1977 to March 1982

In figure 1, the time series of monthly milk consumption does not change over time, the series is stationary in mean, but it seems not stationary invariance. Otherwise, there's no seasonal trend for the data, taking the seasonal difference for the series is unnecessary. Next, we do the box-cox transformation to see if we need any transformation on data to achieve stationary invariance.

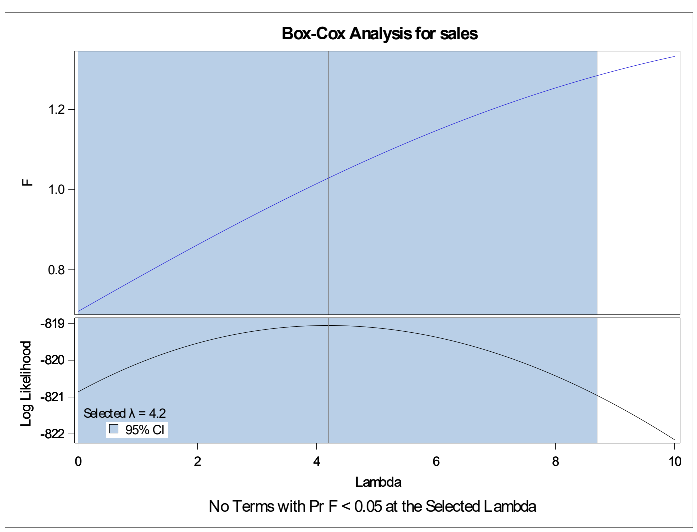

Figure 2. Box-Cox Transformation for sales of pre-intervention

However, BOX-COX transformation suggested the best lambda is 4.2. lambda=1 is included in a 95% Confidence Interval of log-likelihood indicating transformation on data is unnecessary. Then, we try simple differencing on data and see if this action is can make our series achieve stationary in variance.

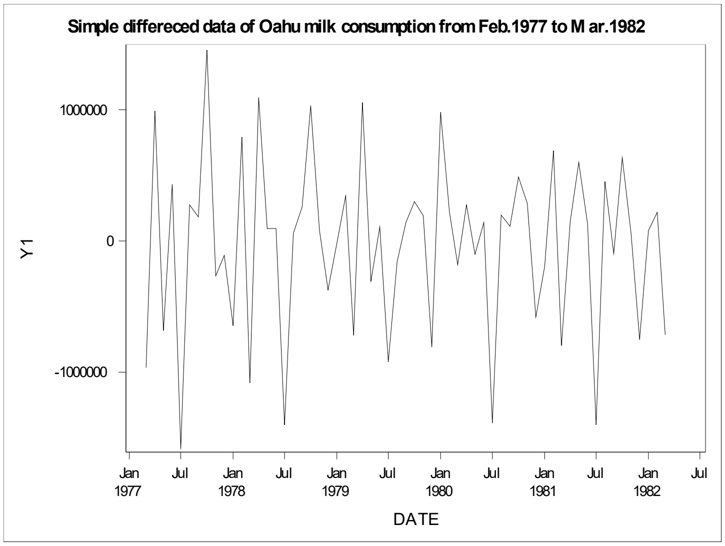

Figure 4. Graph of Simple differenced pre-intervention data

The simple differenced data is also not stationary in variance. So, we will use our original data to do the analysis. Let us call the series up to March 1982 as N_t, and then we need to find the best Noise model for our per-intervention data.

Noise Model Selection By AIC/BIC/LBP Test

1) First candidate models: $(0,0,2)x(1,0,0)_{12}$

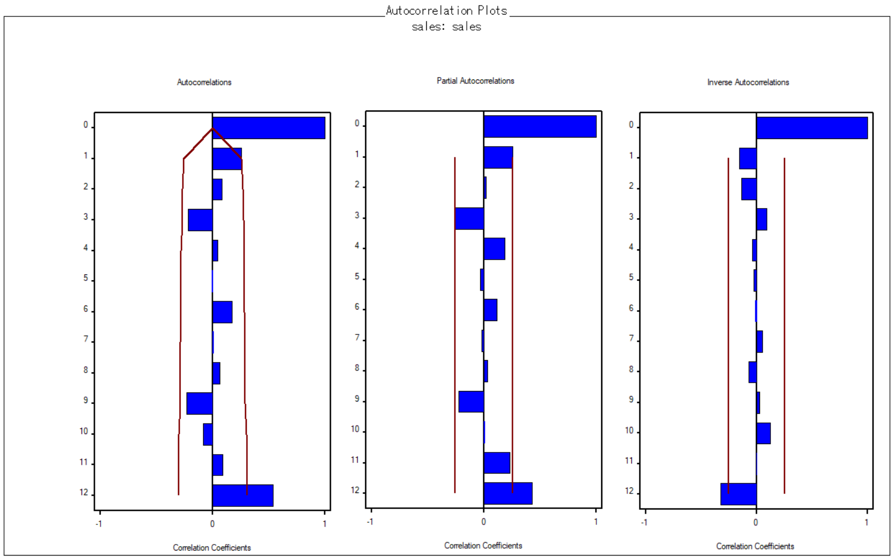

Figure 5. Autocorrelation plots of pre-intervention data (lag<12)

In the autocorrelation plots of the pre-intervention data, In the non-seasonal lags (lag<12), there are 2 significant spikes in PACF at lag1& 3, while ACF decays exponentially. This suggests a possible AR(1) and AR(3) term. On the other hand, there is 1 significant spike in ACF at lag 1, while PACF decays exponentially. This suggests a possible MA(1) term.

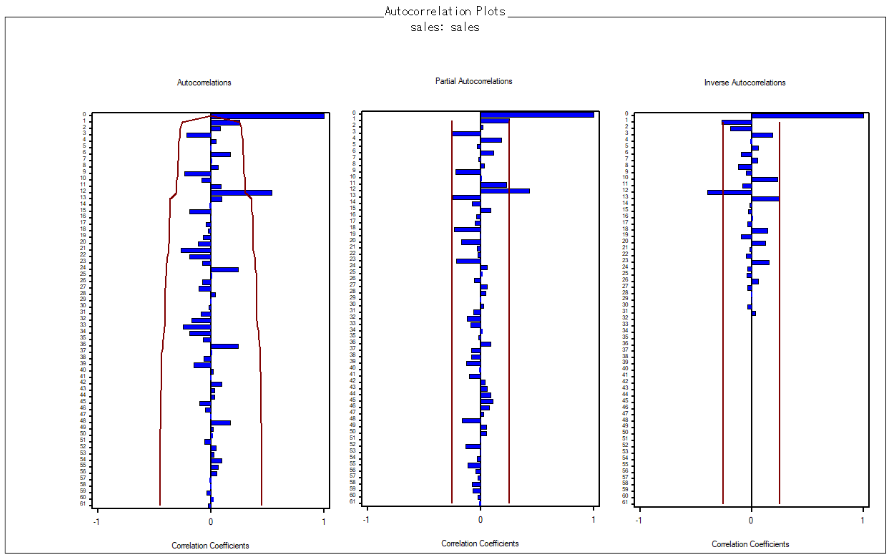

Figure 6. Autocorrelation plots of pre-intervention data (more lags)

In addition, there are spikes in the PACF at lag 12, while ACF decays exponentially at lag 12, 24, etc. This may be suggestive of a seasonal term AR(1) term. On the other hand, there is a spike in the ACF at lag 12, while PACF decays exponentially at lag 12, 24. This may be suggestive of a seasonal term MA(1).

Consequently, this initial analysis suggests that possible models for these data are ARIMA(1,0,0)x$(1,0,0)_{12}$, ARIMA(1,0,0)x$(0,0,1)_{12}$, ARIMA(3,0,0)x$(1,0,0)_{12}$ and ARIMA(3,0,0)x$(0,0,1)_{12}$. We fit these models, along with some variations on them (in total 20 models), compute the Akaike information criterion (AIC=2k-2ln(loglikelihood)), Bayesian information criterion (BIC=2nk-2ln(loglikelihood) and look at the Ljung–Box test (LBP-test) (details shown in Appendix). The models ARIMA(3,0,0)x$(1,0,0)_{12}$, ARIMA(2,0,1)x$(1,0,0)_{12}$, ARIMA(2,0,0)x$(1,0,0)_{12}$ and ARIMA(0,0,2)x$(1,0,0)_{12}$ are selected as the best four models based on AIC/BIC criteria (shown in table.1).

MODEL	AIC	BIC	LBP TEST$_{(LAG>0.05)}$	SIGNIFICANT VBLES
ARIMA(3,0,0)X$(1,0,0)_{12}$	1773	1783	All pass	AR(3) is not significant
ARIMA(2,0,1)X(1,0,0)12	1772	1782	All pass	AR(1) is not significant
ARIMA(2,0,0)X$(1,0,0)_{12}$	1773	1781	All pass	All significant
ARIMA(0,0,2)X$(1,0,0)_{12}$	1772	1780	All pass	All significant

Table 1. best 4 Models among 20 possible models (all models intercept)

Among the four models, ARIMA(0,0,2)x$(1,0,0)_{12}$ is selected as the first candidate model because:

- A preferred model should have the smallest AIC and BIC. ARIMA(0,0,2)x$(1,0,0)_{12}$ has the least AIC/BIC among 4 models.
- Residuals of a model should behave like a white noise series. LBP-test is to check if residuals are correlated with each other with null hypothesis "data are independently distributed". If the p-value of lags is all greater than 0.05 then it passes the LBP-test (accept null hypothesis), the residuals are not correlated with each other. ARIMA(0,0,2)x$(1,0,0)_{12}$ passed LBP test.
- All variables are significant. (p-value < 0.05)

So, the first candidate model is:

$$(1 - \Phi_{12}B^{12})y_t = \mu + (1 - \theta_1 B - \theta_2 B^2)w_t$$

And, the fitted model is:

$$(1 - 0.79664B^{12})y_t = 1762270 + (1 + 0.35085B + 0.39012B^2)w_t$$

A detailed summary of the model can be found in the Appendix.

Then we look at the residual diagnose of the ARIMA $(0,0,2)x(1,0,0)_{12}$.

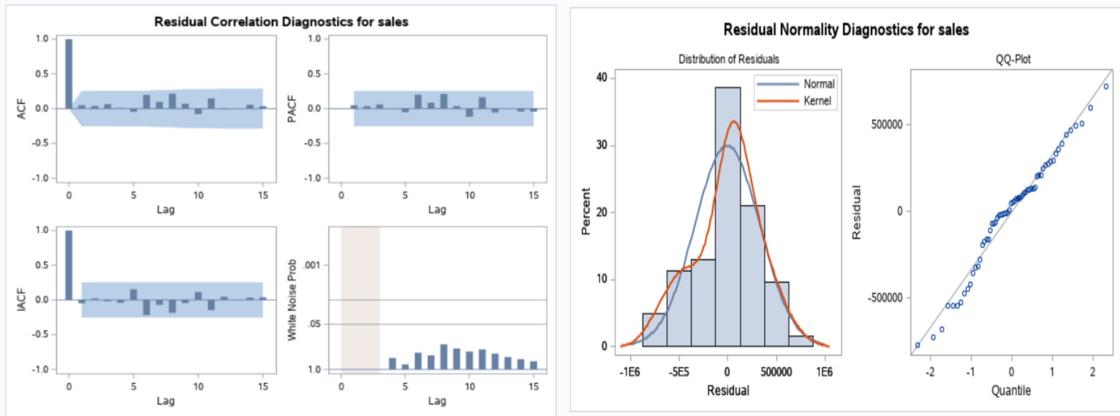

Figure 7. Residual Diagnose for ARIMA(0,0,2)x(1,0,0)_12

Residuals of ARIMA $(0,0,2)x(1,0,0)_{12}$ seems normally distributed in histogram and QQ-plot (figure 9). All spikes are within the significance limits in residual autocorrelation plots, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations.

2) 2^{nd} Candidate: ARIMA$((0,2),0,1)x(1,0,0)_{12}$

Among the best four models in the previous step, ARIMA$(2,0,1)x(1,0,0)_{12}$ also yields the lest AIC, but AR(1) term is not significant. We can improve the model by dropping the AR(3) term from the model. So the resulting models ARIMA$((0,2),0,1)x(1,0,0)_{12}$ is shown in table 2.

MODEL	AIC	BIC	LBP TEST$_{(LAG>0.05)}$	SIGNIFICANT VBLES
ARIMA((0,2),0,1)X(1,0,0)$_{12}$	1771	1779	All pass	All significant

Table 2. Developed model from ARIMA(2,0,1)x(1,0,0)_12 (with intercept)

So, the second candidate model is:

$$(1 - \varphi_1 B^2)(1 - \Phi_{12}B^{12})y_t = \mu + (1 - \theta_1 B)w_t$$

And, the fitted model is:

$$(1 - 0.44056B^2)(1 - 0.80730B^{12})y_t = 931986.4 + (1 + 0.39680B)w_t$$

A detailed summary of the model can be found in the Appendix.

Then we look at the residual diagnose of the ARIMA$((0,2),0,1)x(1,0,0)_{12}$.

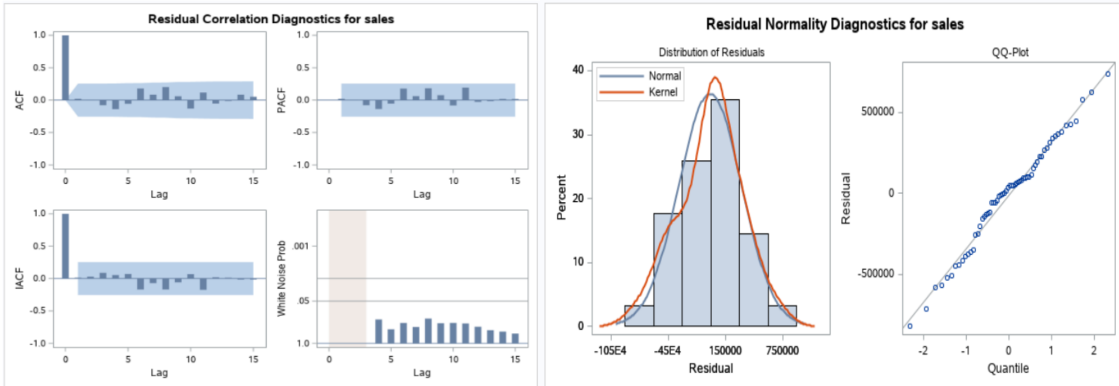

Figure 8. Residual Diagnose for ARIMA((0,2),0,1)x(1,0,0)_12

Residuals of ARIMA((0,2),0,1)x$(1,0,0)_{12}$ seems normally distributed in histogram and QQ-plot (figure 9). All spikes are within the significance limits in residual autocorrelation plots, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations.

3) Select the best model among two candidates

MODEL	AIC	BIC	RMSE	LBP TEST$_{(LAG>0.05)}$	SIGNIFICANT VBLES $_{(P<0.05)}$
ARIMA(0,0,2)X$(1,0,0)_{12}$	1772	1780		All pass	All significant
ARIMA((0,2),0,1)X$(1,0,0)_{12}$	1771	1779		All pass	All significant

Table 3. Candidates models comparison (all models with intercept)

Among two candidate models, ARIMA((0,2),0,1)x$(1,0,0)_{12}$ is selected as the best model because:

- It has the least AIC (1771) and BIC (1779) among 2 models
- It has a lower standard error (0.56).
- It passed the LBP Test at all lags.
- All variables are significant.
- The residuals are white noise. No ACF/PACF is significantly different from zero.
- QQ-plot and histogram of residual distribution indicate the residuals are asymptotically normally distributed.

So the best model for pre-intervention data is ARIMA((0,2),0,1)x$(1,0,0)_{12}$, we will use it as the noise model for our intervention model.

Intervention Model Selection

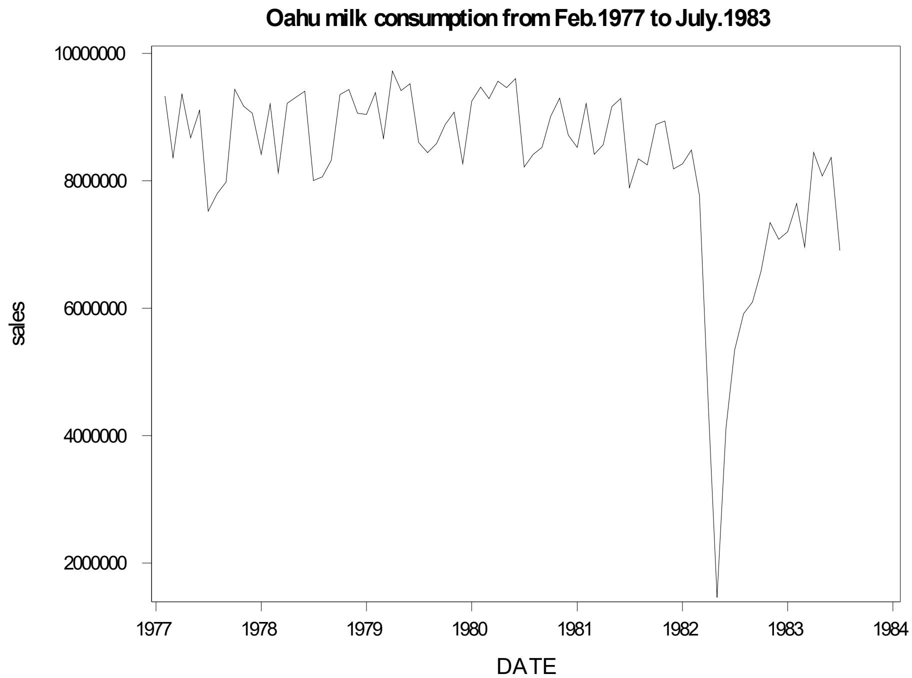

Figure 9. Graph of intervention data

Box and Tiao suggested that the incident of containment in Milk in Oahu impacted the sales of milk and make it suddenly dropped in April 1982 represent an intervention ***I***, which might be expected to produce a step-change in milk consumption starting at April 1982. And a pulse ***P***, which might be expected to occur in April 1982. So, we model our intervention data with step and pulse function, and the summary of four possible intervention models shown in table 4.

NOISE MODEL	INTERVENTION	AIC	BIC	LBP TEST	SIGNIFICANT VBLES
$\frac{\mu + (1-\theta_1 B)w_t}{(1-\varphi_1 B^2)(1-\Phi_{12}B^{12})}$	$\frac{(\omega_0-\omega_1 B)B}{(1-\delta_1 B)} I_t + \omega_0 P_t$	2097	2115	All passed	All Significant
	$\frac{\omega_0 B}{(1-\delta_1 B)} P_t + (\omega_0 - \omega_1 B) I_t$	2097	2115	All passed	All Significant
	$\frac{(\omega_0-\omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2) P_t$	2101	2122	All passed	All Significant
	$\frac{\omega_0 B}{(1-\delta_1 B)} P_t + \frac{\omega_0}{(1-\delta_1 B)} I_t$	2107	2125	All passed	All Significant

Table 4 Four possible intervention models

Among the four models:

- The first two models have the least AIC(2097) and least BIC(2115).
- Residuals of a model should behave like a white noise series. LBP-test is to check if residuals are correlated with each other with null hypothesis "data are independently distributed". If the p-value of lags is all greater than 0.05 then it passes the LBP-test (accept null hypothesis), the residuals are not correlated with each other. All four models passed the LBP test.
- All variables of the four models are significant. (p-value < 0.05), this indicates the intervention did affect the series.

So,

1) 1st candidate intervention model:

$$y_t = \frac{(\omega_0-\omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2) P_t + \frac{\mu+(1-\theta_1 B)w_t}{(1-\varphi_1 B^2)(1-\Phi_{12} B^{12})}$$

Where I_t=1 for a date after April 1982 and 0 before the date, and P_t=1 for April 1982 and elsewhere.

A detailed summary of the model can be found in the Appendix.

Then we look at the residual diagnose of the 1st candidate.

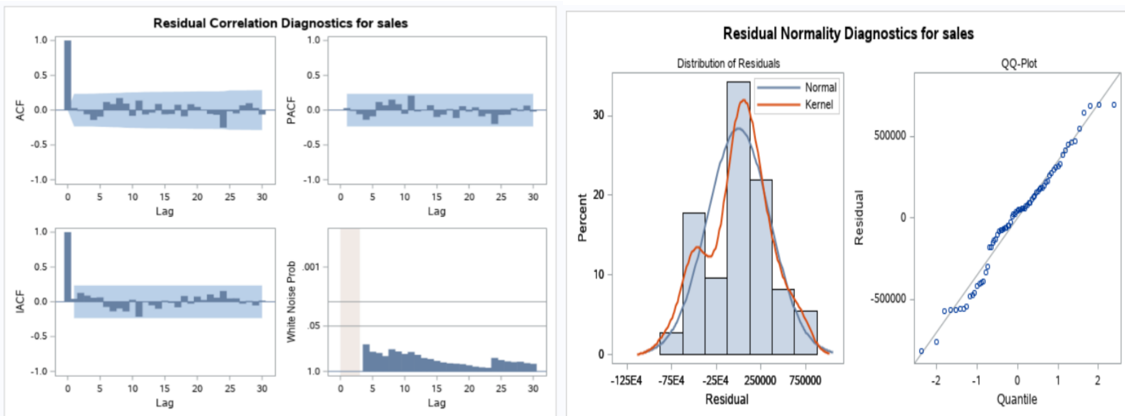

Figure 10. Residual Diagnose for candidate 1

Residuals of 1st candidate seem normally distributed in histogram and QQ-plot (figure 10). All spikes are within the significance limits in residual autocorrelation plots, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations.

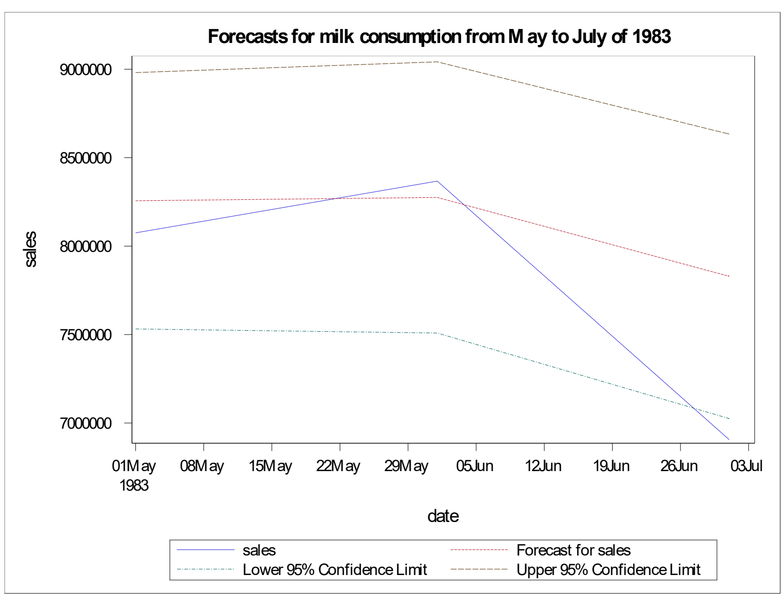

Figure 11. Forecasts from the 1st candidate model

Forecasts from 1st candidate for May to July of 1983 are shown in Figure.11. Forecasts have shown in red dashed lines, and blues lines represent upper and lower 95% confidence interval for forecast values. The forecasts follow the recent trend in the data, increases till the beginning of June of 1983 then dropdown. The prediction from the model does not close to the pattern of the observed sales of May to July of 1983.

2) 2nd Candidate intervention model:

$$y_t = \frac{(\omega_0 - \omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2) P_t + \frac{\mu + (1-\theta_1 B) w_t}{(1-\varphi_1 B^2)(1-\Phi_{12} B^{12})}$$

A detailed summary of the model can be found in the Appendix.

Then we look at the residual diagnose of the 2nd Candidate.

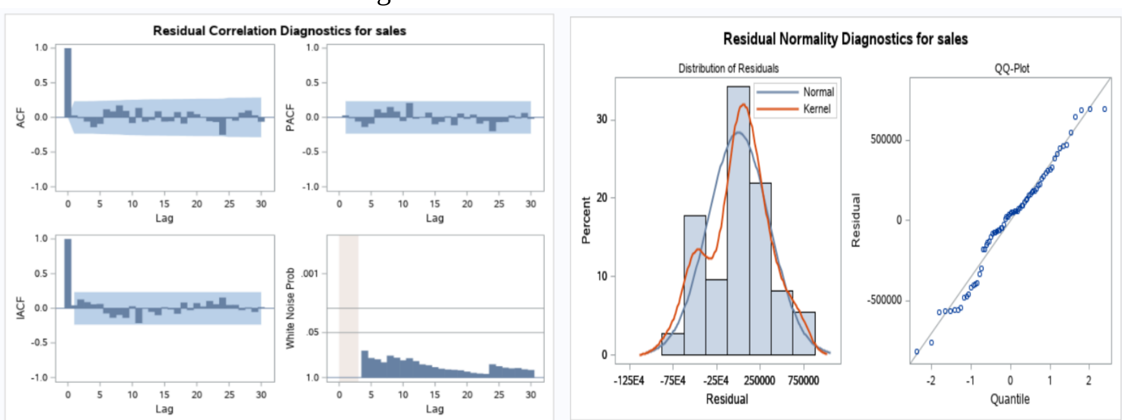

Figure 12. Residual Diagnose for 2nd candidate

Residuals of the 2nd candidate model seem normally distributed in histogram and QQ-plot (figure 12). All spikes are within the significance limits in residual autocorrelation plots, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations.

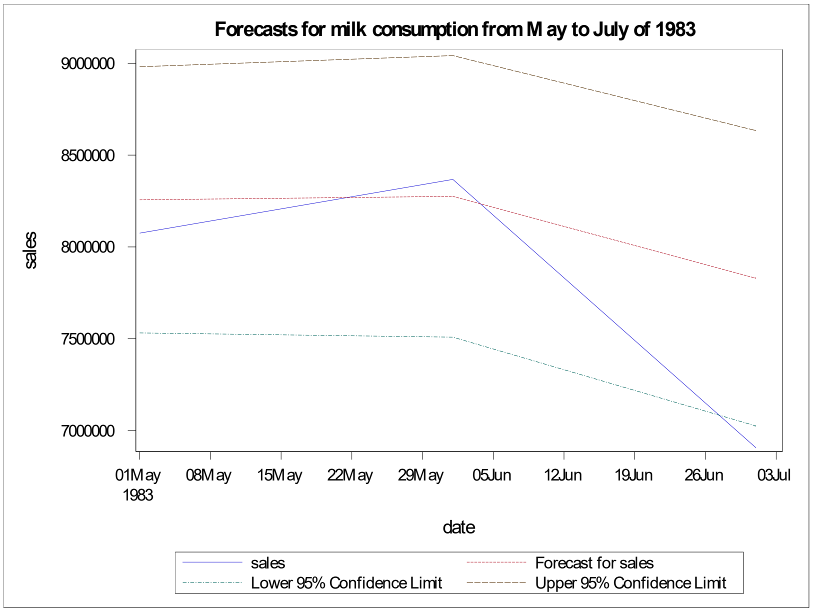

Figure 13. Forecasts from the 2nd candidate model

Forecasts from 2nd candidate for May to July of 1983 are shown in Figure.11. Forecasts have shown in red dashed lines, and blues lines represent upper and lower 95% confidence interval for forecast values. The forecasts follow the recent trend in the data, increases till the beginning of June of 1983 then dropdown. The prediction from the model does not close to the pattern of the observed sales of May to July of 1983.

3) 3rd Candidate intervention model:

$$y_t = \frac{(\omega_0 - \omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2) P_t + \frac{\mu + (1-\theta_1 B) w_t}{(1-\varphi_1 B^2)(1-\Phi_{12} B^{12})}$$

A detailed summary of the model can be found in the Appendix.

Then we look at the residual diagnose of the 3rd candidate.

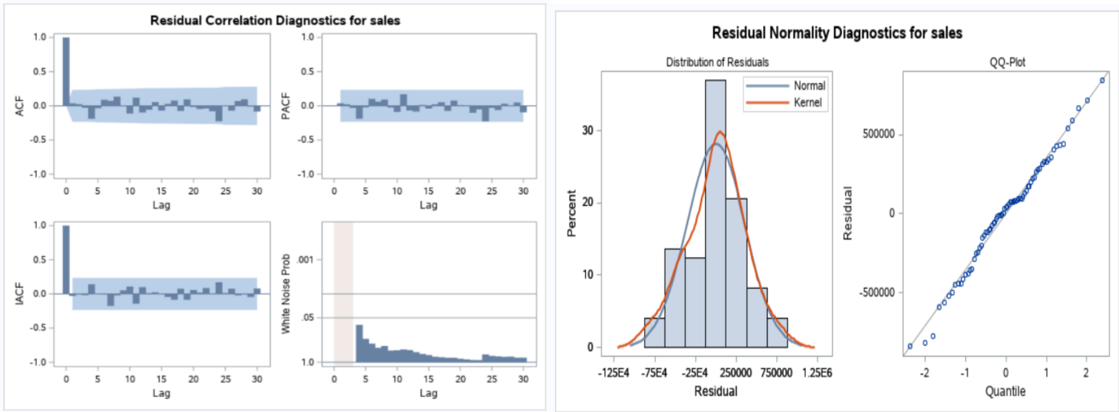

Figure 14. Residual Diagnose for 3rd candidate

Residuals of the 3rd candidate model seem normally distributed in histogram and QQ-plot (figure 14). All spikes are within the significance limits in residual autocorrelation plots, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations.

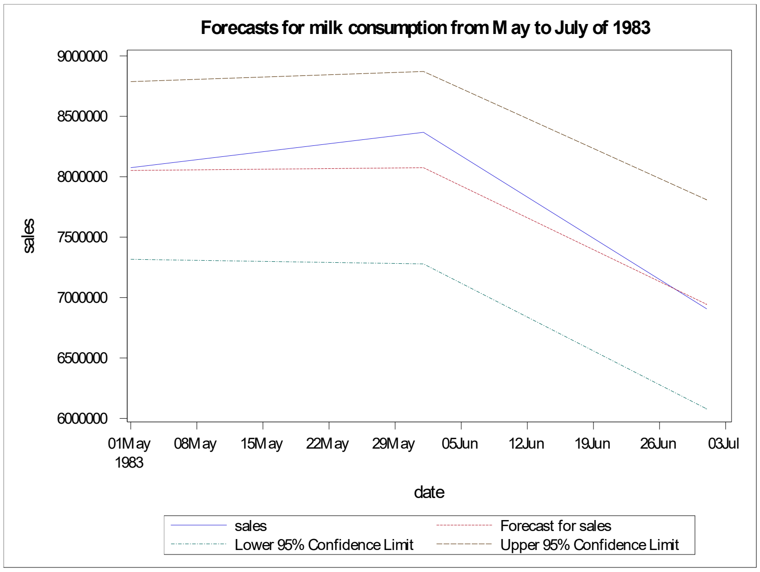

Figure 15. Forecasts from the 3rd candidate model

Forecasts from 3rd candidate for May to July of 1983 are shown in Figure.11. Forecasts have shown in red dashed lines, and blues lines represent upper and lower 95% confidence interval for forecast values. The forecasts follow the recent trend in the data, increases till the beginning of June of 1983 then dropdown. The prediction from the model is the closest to the observed sale.

4) 4th Candidate intervention model:

$$y_t = \frac{(\omega_0 - \omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2) P_t + \frac{\mu + (1-\theta_1 B) w_t}{(1-\varphi_1 B^2)(1-\Phi_{12} B^{12})}$$

A detailed summary of the model can be found in the Appendix.

Then we look at the residual diagnose of the 4th candidate.

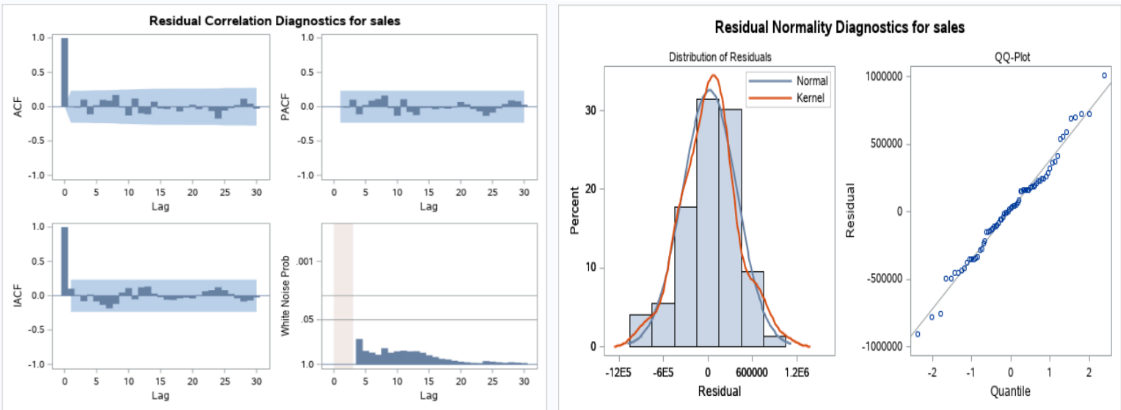

Figure 16. Residual Diagnose for 4th candidate

Residuals of the 4th candidate seem normally distributed in histogram and QQ-plot (figure 16). All spikes are within the significance limits in residual autocorrelation plots, so the residuals appear to be white noise. The Ljung-Box test also shows that the residuals have no remaining autocorrelations.

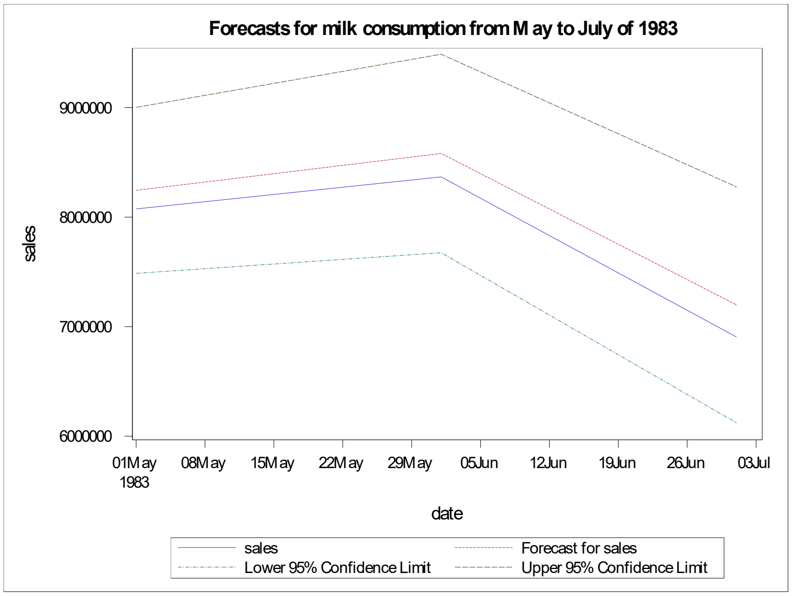

Figure 17. Forecasts from the 4th candidate model

Forecasts from the 4th candidate model for May to July of 1983 are shown in Figure.11. Forecasts have shown in red dashed lines, and blues lines represent upper and lower 95% confidence interval for forecast values. The forecasts follow the recent trend in the data, increases till the beginning of June of 1983 then dropdown. The trend of prediction sales from the model follows the observed sales but with a downshift.

Best Intervention Model Selection MAPE

The mean absolute *percentage* error (MAPE) is also often useful for comparing the model, which expresses the prediction accuracy of the forecasting model as a percentage of the error. By computing the MAPE value for the three best intervention models from the previous step using the data from January 2004 to December 2004, The 3rd candidate model is selected as the best model with the least MAPE value, although its AIC and BIC is larger than first two. With the least MAPE, the percentage error of the model is smaller, then the prediction accuracy of the 3nd intervention model is higher than others, which is also shown in the forecast plot of the 3rd model.

NOISE MODEL	INTERVENTION	MAPE
	$\frac{(\omega_0-\omega_1 B)B}{(1-\delta_1 B)} I_t + \omega_0 P_t$	0.055715

$\dfrac{\mu + (1-\theta_1 B)w_t}{(1-\varphi_1 B^2)(1-\Phi_{12}B^{12})}$	$\dfrac{\omega_0 B}{(1-\delta_1 B)}P_t + (\omega_0 - \omega_1 B)I_t$	0.055715
	$\frac{(\omega_0 - \omega_2 B^2)}{(1-\delta_2 B^2)} I_t + (\omega_0 - \omega_2 B^2)P_t$	0.014421
	$\frac{\omega_0 B}{(1-\delta_1 B)} P_t + \frac{\omega_0}{(1-\delta_1 B)} I_t$	0.029688

Table 5. MAPE values of competing intervention models

Forecast

Forecasts from the best model for entire data and for Jan.1982 to July 1983 are shown in Figure.18 and 19 Forecasts have shown in red dashed lines, and blues lines represent upper and lower 95% confidence interval for forecast values.

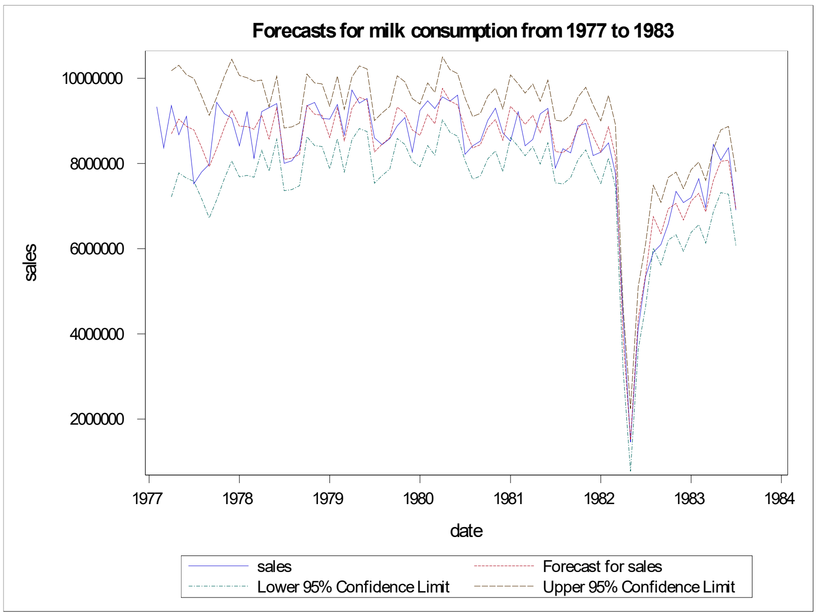

Figure 18. Forecast of the best intervention model for entire data

In figure 18, the forecast for the entire data follows the trend of observed data, drops rapidly after the milk containment incident in April 1982 then slowly rises again and slowly back to normal after mid of 1982.

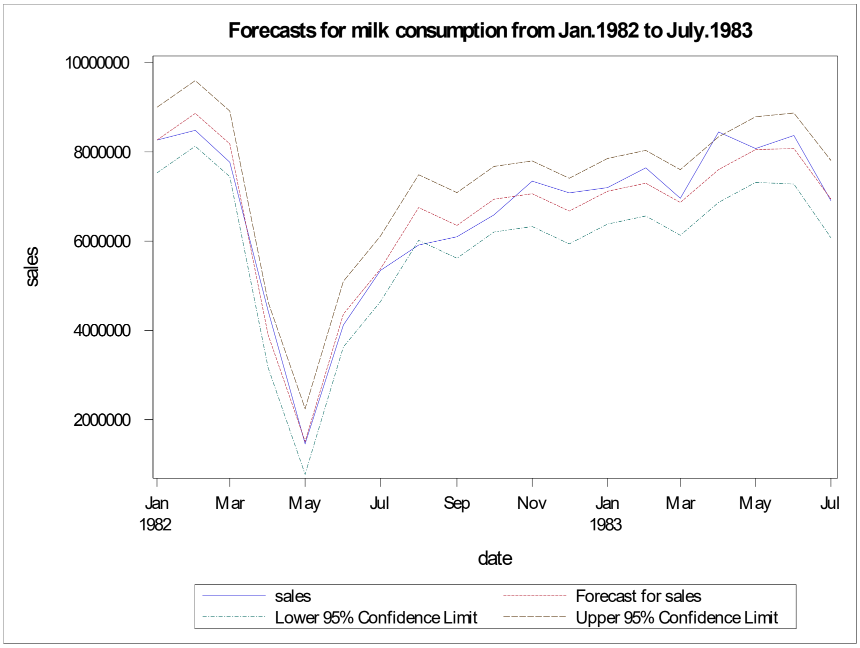

Figure 19. Forecast of the best intervention model for Jan 1982 to July 1983

In figure 19, the forecast for Jan 1982 to July 1983, we can take a closer look at the trend of sales for after milk containment incident. It's clear to see that after-sales drop and reach the minimum sale in May of 1982, it raises again and slowly back to normal after August of 1983.

Appendix

1) Summary of four possible noise models

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	8645191.6	309337.4	27.95	<.0001	0
MA1,1	-0.39680	0.11887	-3.34	0.0008	1
AR1,1	0.44056	0.11775	3.74	0.0002	2
AR2,1	0.80730	0.06569	12.29	<.0001	12

Constant Estimate	931986.4
Variance Estimate	1.139E11
Std Error Estimate	337507.5
AIC	1771.392
SBC	1779.901
Number of Residuals	62

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.23	3	0.2376	0.021	-0.004	-0.078	-0.140	-0.057	0.177
12	10.43	9	0.3169	0.080	0.199	0.058	-0.131	0.115	-0.056
18	14.33	15	0.5009	-0.015	0.079	0.047	0.054	0.021	-0.179
24	19.78	21	0.5350	0.026	0.017	0.059	0.025	-0.054	-0.211

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	8665894.2	212822.2	40.72	<.0001	0
MA1,1	-0.35085	0.11338	-3.09	0.0020	1
MA1,2	-0.39012	0.11160	-3.50	0.0005	2
AR1,1	0.79664	0.07012	11.36	<.0001	12

Constant Estimate	1762270
Variance Estimate	1.163E11
Std Error Estimate	341067
AIC	1772.042
SBC	1780.551
Number of Residuals	62

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.38	3	0.3367	0.048	0.037	0.063	0.010	-0.047	0.196
12	9.84	9	0.3635	0.094	0.212	0.070	-0.081	0.145	-0.012
18	12.47	15	0.6429	-0.006	0.052	0.031	0.037	0.022	-0.155
24	16.86	21	0.7197	0.037	-0.006	0.012	-0.014	-0.061	-0.192

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	8646069.2	308760.5	28.00	<.0001	0
AR1,1	0.34331	0.11468	2.99	0.0028	1
AR1,2	0.24259	0.11995	2.02	0.0431	2
AR2,1	0.81052	0.06412	12.64	<.0001	12

Constant Estimate	678396.2
Variance Estimate	1.174E11
Std Error Estimate	342620.8
AIC	1773.326
SBC	1781.835
Number of Residuals	62

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.70	3	0.0820	0.055	0.063	-0.178	-0.110	-0.098	0.192
12	12.62	9	0.1805	0.098	0.209	0.053	-0.104	0.090	-0.069
18	16.45	15	0.3526	0.004	0.059	0.075	0.046	0.022	-0.177
24	23.54	21	0.3159	0.024	0.004	0.086	0.003	-0.041	-0.242

2) Summary of possible intervention models

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.70	3	0.2957	0.029	-0.010	-0.060	-0.140	-0.093	0.116
12	10.18	9	0.3358	0.087	0.173	0.094	-0.081	0.137	-0.062
18	13.03	15	0.6002	-0.035	0.088	-0.060	-0.057	0.077	-0.088
24	21.78	21	0.4123	0.084	0.046	0.007	-0.055	-0.066	-0.250
30	24.44	27	0.6059	-0.016	-0.047	0.068	0.099	0.032	-0.062

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	8687236.6	248985.9	34.89	<.0001	0	sales	0
MA1,1	-0.34474	0.12064	-2.86	0.0043	1	sales	0
AR1,1	0.33526	0.12556	2.67	0.0076	2	sales	0
AR2,1	0.76157	0.07465	10.20	<.0001	12	sales	0
NUM1	-7606825.2	367088.5	-20.72	<.0001	0	I	1
NUM1,1	-7239505.5	372900.5	-19.41	<.0001	1	I	1
DEN1,1	0.60151	0.04645	12.95	<.0001	1	I	1
NUM2	-4255270.2	292766.0	-14.53	<.0001	0	P	0

Constant Estimate	1376865
Variance Estimate	1.367E11
Std Error Estimate	369728.1
AIC	2097.186
SBC	2115.51
Number of Residuals	73

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.70	3	0.2957	0.029	-0.010	-0.060	-0.140	-0.093	0.116
12	10.18	9	0.3358	0.087	0.173	0.094	-0.081	0.137	-0.062
18	13.03	15	0.6002	-0.036	0.088	-0.060	-0.057	0.077	-0.088
24	21.78	21	0.4123	0.084	0.046	0.007	-0.055	-0.066	-0.250
30	24.44	27	0.6058	-0.016	-0.047	0.068	0.099	0.032	-0.062

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	8687239.1	249019.1	34.89	<.0001	0	sales	0
MA1,1	-0.34474	0.12064	-2.86	0.0043	1	sales	0
AR1,1	0.33531	0.12556	2.67	0.0076	2	sales	0
AR2,1	0.76157	0.07465	10.20	<.0001	12	sales	0
NUM1	-6685025.9	440151.2	-15.19	<.0001	0	P	1
DEN1,1	0.60149	0.04644	12.95	<.0001	1	P	1
NUM2	-4255253.2	292762.6	-14.53	<.0001	0	I	0
NUM1,1	-3333451.8	352092.5	-9.47	<.0001	1	I	0

Constant Estimate	1376755
Variance Estimate	1.367E11
Std Error Estimate	369727.6
AIC	2097.186
SBC	2115.51
Number of Residuals	73

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.79	3	0.2853	0.037	0.023	-0.029	-0.189	-0.035	0.088
12	8.98	9	0.4391	0.072	0.134	-0.006	-0.113	0.120	-0.098
18	11.20	15	0.7380	-0.057	0.056	-0.068	0.029	0.076	-0.075
24	18.89	21	0.5921	0.094	-0.017	-0.047	-0.039	-0.079	-0.225
30	22.06	27	0.7343	0.003	-0.069	0.070	0.096	-0.010	-0.084

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	8696364.1	350918.8	24.78	<.0001	0	sales	0
MA1,1	-0.41555	0.12261	-3.39	0.0007	1	sales	0
AR1,1	0.46092	0.12697	3.63	0.0003	2	sales	0
AR2,1	0.79220	0.07417	10.68	<.0001	12	sales	0
NUM1	-7454924.7	378556.4	-19.69	<.0001	0	I	0
NUM1,1	-6398083.2	453651.8	-14.10	<.0001	2	I	0
DEN1,1	0.19501	0.06826	2.86	0.0043	2	I	0
NUM2	3223907.3	395390.0	8.15	<.0001	0	P	0
NUM1,1	2311138.0	413246.2	5.59	<.0001	2	P	0

Constant Estimate	974188.7
Variance Estimate	1.407E11
Std Error Estimate	375147.8
AIC	2101.865
SBC	2122.48
Number of Residuals	73

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.57	3	0.4627	-0.005	-0.012	0.101	-0.109	0.023	0.098
12	9.07	9	0.4308	0.084	0.168	-0.013	-0.129	0.118	-0.097
18	11.35	15	0.7271	-0.113	0.069	-0.028	-0.036	-0.006	-0.069
24	15.43	21	0.8007	-0.013	0.026	-0.036	0.008	-0.063	-0.174
30	17.94	27	0.9054	-0.057	-0.022	0.023	0.117	0.045	-0.028

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	9381658.0	936303.2	10.02	<.0001	0	sales	0
MA1,1	-0.65563	0.11200	-5.85	<.0001	1	sales	0
AR1,1	0.76846	0.10595	7.25	<.0001	2	sales	0
AR2,1	0.79574	0.07211	11.04	<.0001	12	sales	0
NUM1	-6018089.3	366537.3	-16.42	<.0001	0	P	1
DEN1,1	0.34544	0.05844	5.91	<.0001	1	P	1
NUM2	-4054791.3	289297.0	-14.02	<.0001	0	I	0
DEN1,1	-0.63903	0.06414	-9.96	<.0001	1	I	0

Constant Estimate	443688.8
Variance Estimate	1.495E11
Std Error Estimate	386678.5
AIC	2107.32
SBC	2125.643
Number of Residuals	73

SAS Code

```
data PROJECT3.milk1;
set PROJECT3.milk;
if time>62 then delete;
keep date sales;
run;
quit;

Proc sgplot DATA = PROJECT3.MILK;
 SERIES X =date Y = sales;
 TITLE 'Oahu milk consumption from Feb.1977 to July.1983';
```

```
run;
quit;

Proc sgplot DATA = PROJECT3.MILK1;
 SERIES X =date Y = sales;
 TITLE 'Oahu milk consumption from Feb.1977 to Mar.1982';
run;
quit;

proc arima data=PROJECT3.MILK1 plots;
identify var=sales scan nlag=60;
run;

proc univariate data=PROJECT3.MILK1;
  var sales;
  histogram / normal (fill mu =est sigma = est);
qqplot sales/ normal(mu=est sigma=est) square;
run;
quit;

title2 'Identify noise model for the values before Mar.1982';
proc arima data=PROJECT3.MILK1;
  identify var=sales;
  estimate p=(1)(12) q=0 printall grid method=ml plot; *(1,0,0)x(1,0,0)_12;
  *estimate p=(1) q=(0)(12) printall grid method=ml plot;*(1,0,0)x(0,0,1)_12;
  *estimate p=(0)(12) q=1 printall grid method=ml plot;*(0,0,1)x(1,0,0)_12;
  *estimate p=0 q=(1)(12) printall grid method=ml plot;*(0,0,1)x(0,0,1)_12;
   estimate p=(1,2,3)(12) q=0 printall grid method=ml plot;*(3,0,0)x(1,0,0)_12;
  *estimate p=(1,2,3) q=(0)(12) printall grid method=ml plot;*(3,0,0)x(0,0,1)_12;
  *estimate p=(0)(12) q=(1,2,3) printall grid method=ml plot;*(0,0,3)x(1,0,0)_12;
```

```
  *estimate p=0 q=(1,2,3)(12) printall grid method=ml plot;*(0,0,3)x(0,0,1)_12;
  *estimate p=(1)(12) q=(1,2) printall grid method=ml plot;*(1,0,2)x(1,0,0)_12;
  *estimate p=(1) q=(1,2)(12) printall grid method=ml plot;*(1,0,2)x(0,0,1)_12;
  estimate p=(1,2)(12) q=1 printall grid method=ml plot;*(2,0,1)x(1,0,0)_12;
  *estimate p=(1,2) q=(1)(12) printall grid method=ml plot;*(2,0,1)x(0,0,1)_12;
   *estimate p=(0)(12) q=0 printall grid method=ml plot;*(0,0,0)x(1,0,0)_12;
  *estimate p=0 q=(0)(12) printall grid method=ml plot;*(0,0,0)x(0,0,1)_12;
   estimate p=(1,2)(12) q=0 printall grid method=ml plot;*(2,0,0)x(1,0,0)_12;
  *estimate p=(1,2) q=(0)(12) printall grid method=ml plot;*(2,0,0)x(0,0,1)_12;
   estimate p=(0)(12) q=(1,2) printall grid method=ml plot;*(0,0,2)x(1,0,0)_12;
  *estimate p=0 q=(1,2)(12) printall grid method=ml plot;*(0,0,2)x(0,0,1)_12;
  *estimate p=(1)(12) q=1 printall grid method=ml plot;*(1,0,1)x(1,0,0)_12;
  *estimate p=(1) q=(1,2)(12) printall grid method=ml plot;*(1,0,1)x(0,0,1)_12;
run;
quit;

title2 'best noise model for the values before Mar.1982';
proc arima data=PROJECT3.MILK1;
  identify var=sales;
  estimate p=(0,2)(12) q=1 printall grid method=ml plot;
   estimate p=(1,2)(12) q=0 printall grid method=ml plot;
     estimate p=(0)(12) q=(1,2) printall grid method=ml plot;
run;
quit;


data PROJECT3.Milk2;
		set PROJECT3.Milk;
		keep sales;
*Here I am retaining only the sales column in my Milk data;
```

```
run;

quit;


data PROJECT3.Milk2;

        set PROJECT3.Milk2;

        retain date '01jan77'd;

        date =intnx('month', date,1);

        format date monyy.;

        P=(date='01apr82'd);

        I=(date>= '01apr82'd);

        time=_n_;

*Run this code and look at the data, then you will know what this did to Milk2 data;

run;

quit;


data PROJECT3.MilktoApr83;

        set PROJECT3.Milk2;

        if time > 75 then sales='.';

*You need to create this kind of data where the last three values are identified as missing before fitting the intervention models, only the SAS will give you forecast for the last three values which you would need to compute MAPE;

run;

quit;


proc arima data=PROJECT3.MilktoApr83;

  identify var=sales crosscorr=(I P) nlag=30;

  estimate p=(0,2)(12) q=1 input=(1$(1)/(1)I P) printall grid method=ml plot;

  forecast out=PROJECT3.SARIMA id=time alpha=0.05 lead=3;

run;

quit;

proc arima data=PROJECT3.MilktoApr83;
```

```
  identify var=sales crosscorr=(I P) nlag=30;
  estimate p=(0,2)(12) q=1 input=(1$/(1)P (1)I) printall grid method=ml plot;
  forecast out=PROJECT3.SARIMA id=time alpha=0.05 lead=3;
run;
quit;

proc arima data=PROJECT3.MilktoApr83;
  identify var=sales crosscorr=(I P) nlag=30;
  estimate p=(0,2)(12) q=1 input=((2)/(2)I (2)P) printall grid method=ml plot;
  forecast out=PROJECT3.SARIMA id=time alpha=0.05 lead=3;
run;
quit;

proc arima data=PROJECT3.MilktoApr83;
  identify var=sales crosscorr=(I P) nlag=30;
  estimate p=(0,2)(12) q=1 input=(1$/(1)P 0$/(1)I) printall grid method=ml plot;
  forecast out=PROJECT3.SARIMA id=time alpha=0.05 lead=3;
run;
quit;
data PROJECT3.SARIMA;
 merge PROJECT3.SARIMA PROJECT3.Milk2;
run;
quit;

Proc sgplot DATA = PROJECT3.SARIMA;
where time>75;
 SERIES X =date Y = sales;
 SERIES X =date Y = forecast;
 SERIES X =date Y = L95;
 SERIES X =date Y = U95;
```

```
 TITLE 'Forecasts for milk consumption from May to July of 1983';
run;
quit;


title2 "MAPE Computations for SARIMA ((0,2),0,1)x(0,0,1)_12 for Sale with Intercept";
data forecastbest1;
 set PROJECT3.SARIMA;
 where time > 75;
 keep forecast1 time;
run;
quit;

data PartCompbest1;
 set PROJECT3.Milk2;
 where time > 75;
 keep sales time;
run;
quit;


data MAPEbest1;
 merge forecastbest1 PartCompbest1;
 ratio1 = (abs(sales-forecast1))/sales;
proc print;
run;
quit;

proc means data=MAPEbest1 noprint;
 var ratio1;
```

```
output out=mape1 mean=mape1;
run;

proc print;
title2 "MAPE value for SARIMA ((0,2),0,1)x(0,0,1)_12 for Sale with Intercept";
run;
quit;
proc arima data=PROJECT3.MilktoApr83;
  identify var=sales crosscorr=(I P) nlag=30;
  estimate p=(0,2)(12) q=1 input=((2)/(2)I (2)P) printall grid method=ml plot;
  forecast out=PROJECT3.SARIMA id=time alpha=0.05 lead=78;
run;
quit;

data PROJECT3.SARIMA;
 merge PROJECT3.SARIMA PROJECT3.Milk2;
run;
quit;

Proc sgplot DATA = PROJECT3.SARIMA;
 SERIES X =date Y = sales;
 SERIES X =date Y = forecast;
 SERIES X =date Y = L95;
 SERIES X =date Y = U95;
 TITLE 'Forecasts for milk consumption from 1977 to 1983';
run;
quit;

proc arima data=PROJECT3.MilktoApr83;
  identify var=sales crosscorr=(I P) nlag=30;
```

```
  estimate p=(0,2)(12) q=1 input=((2)/(2)I (2)P) printall grid method=ml plot;
  forecast out=PROJECT3.SARIMA id=time alpha=0.05 lead=19;
run;
quit;

data PROJECT3.SARIMA;
 merge PROJECT3.SARIMA PROJECT3.Milk2;
run;
quit;

Proc sgplot DATA = PROJECT3.SARIMA;
 where time >59;
 SERIES X =date Y = sales;
 SERIES X =date Y = forecast;
 SERIES X =date Y = L95;
 SERIES X =date Y = U95;
 TITLE 'Forecasts for milk consumption from Jan.1982 to July.1983';
run;
quit;
```