Time series analysis using SAS

Yu-Hsuan TING

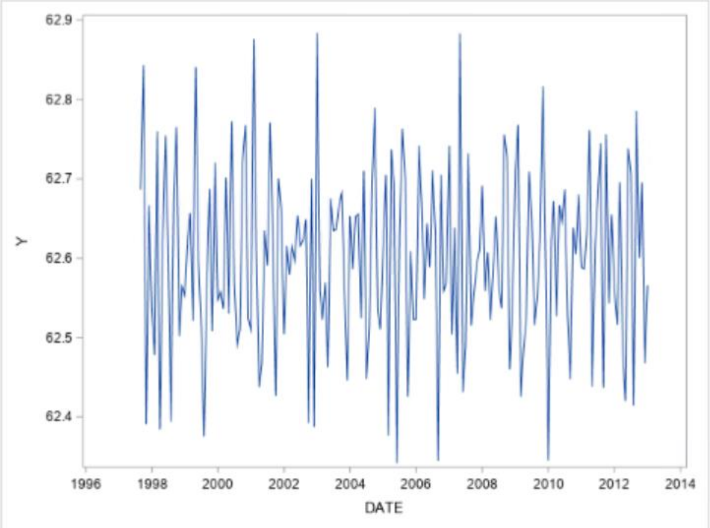
17th Jan 2020

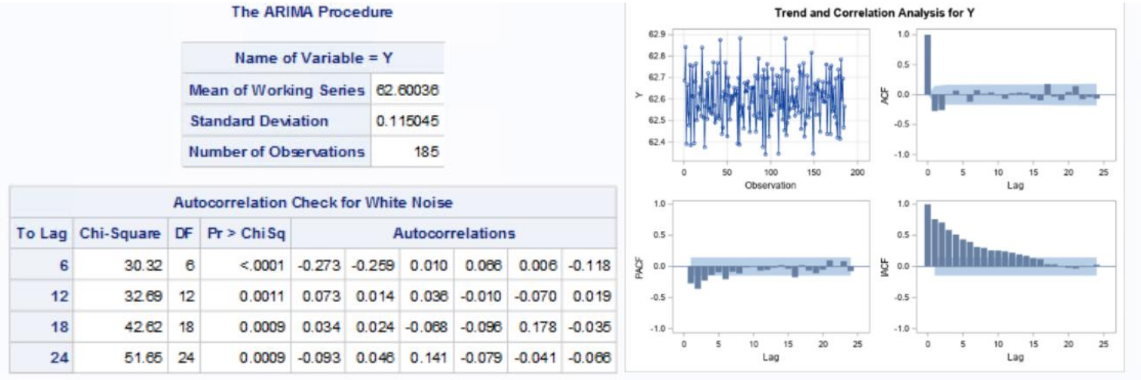
Contents

1. Part 1: Exercise

1.1. E1 dataset

You receive the SAS data set E1 from a colleague. Represent with a graph the timeseries and identify and estimate an appropriate model to fit the data. Justify your choice.





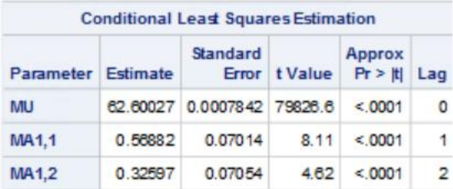
The ARIMA Procedure

Name of Variable = Y	
Mean of Working Series	62.60036
Standard Deviation	0.115045
Number of Observations	185

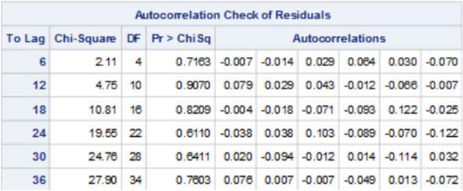
Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	30.32	6	<.0001	-0.273	-0.259	0.010	0.066	0.006	-0.118
12	32.69	12	0.0011	0.073	0.014	0.036	-0.010	-0.070	0.019
18	42.62	18	0.0009	0.034	0.024	-0.068	-0.096	0.178	-0.035
24	51.65	24	0.0009	-0.093	0.046	0.141	-0.079	-0.041	-0.066

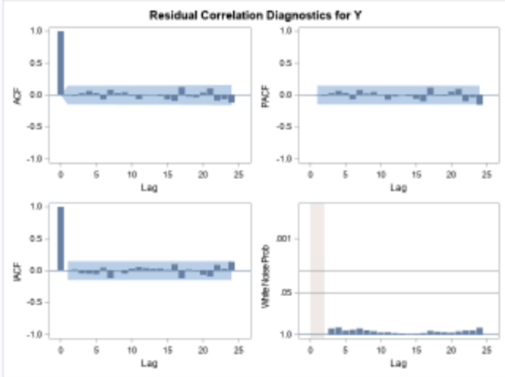
With simple plot of time series with a glance doesn’t seems to have any trend and seasonality. With Proc arima I see that there’s 185 observations without any missing data, with monthly interval. Then I check the autocorrelation with the lag p-value small, that is reject the H0 where there’s a autocorrelation in the data from the Ljung box test, so I know that our data is not a white noise I try to fit the ARMA model. I first identify the y variable from ploc arima. From ACF, PACF I see that PACF is decaying and ACF go only significant until lag 2 so I can assume it to be a MA2 model (q=2).

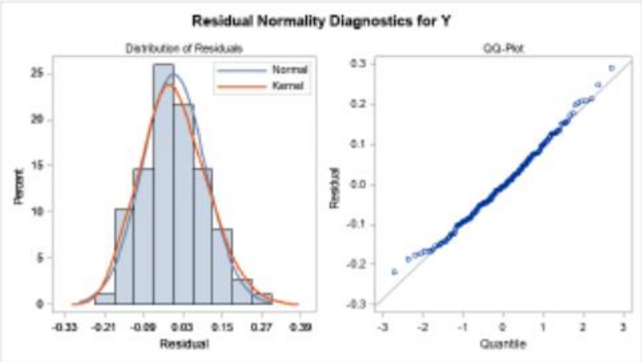


Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	62.60027	0.0007842	79826.6	<.0001	0
MA1,1	0.56882	0.07014	8.11	<.0001	1
MA1,2	0.32597	0.07054	4.62	<.0001	2



Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.11	4	0.7163	-0.007	-0.014	0.029	0.064	0.030	-0.070
12	4.75	10	0.9070	0.079	0.029	0.043	-0.012	-0.066	-0.007
18	10.81	16	0.8209	-0.004	-0.018	-0.071	-0.093	0.122	-0.025
24	19.55	22	0.6110	-0.038	0.038	0.103	-0.089	-0.070	-0.122
30	24.76	28	0.6411	0.020	-0.094	-0.012	0.014	-0.114	0.032
36	27.90	34	0.7603	0.076	0.007	-0.007	-0.049	0.013	-0.072

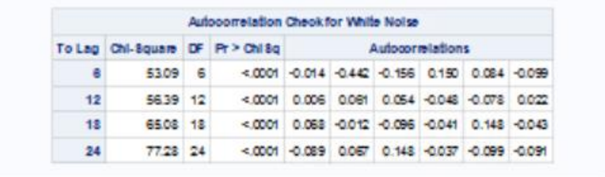




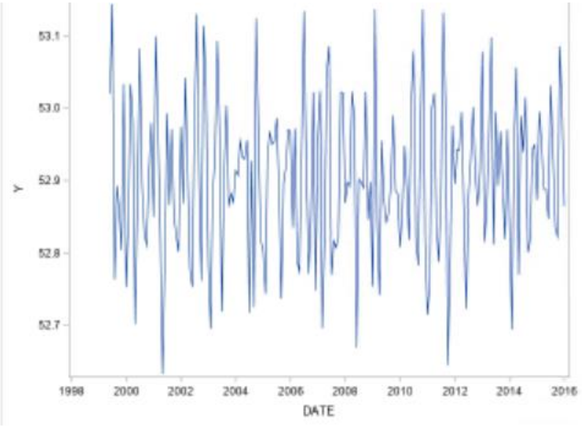
After put in estimation q=2, from the conditional least square estimation table I see that for all intercept, MA1 and MA2 parameters have small p-value here suggesting that they are all significant, the new fitted Autocorrelation check now is accepting Ljung box test (it is a white noise), no lags in PACF, ACF is significant, also for the white noise prob suggesting that I don't reject the white noise test, all these evidence saying that now the residuals is a white noise. Finally, for the residual normality diagnostics residuals fit Ill to the normal distribution and QQ plot. I can then say MA2 is a good model for this dataset.

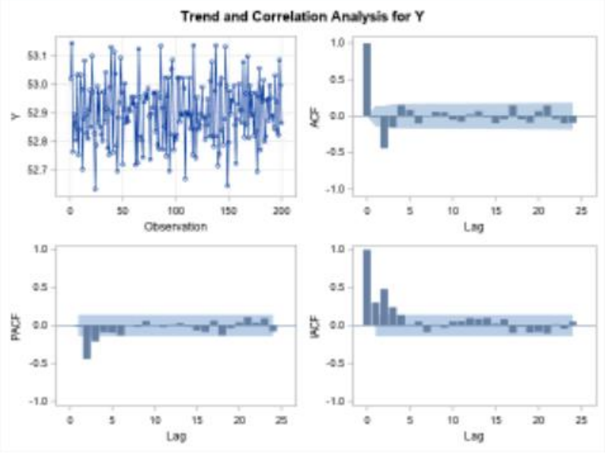
1.2. E2 dataset

Identify and estimate a relevant model for the variable Y in the SAS data set E2. You will use the Maximum Likelihood estimation method to obtain your model. Explain how you have decided which model to select.

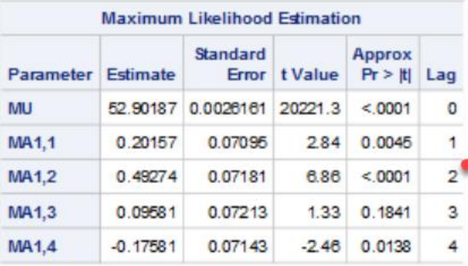


Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi Sq	Autocorrelations					
6	53.09	6	<.0001	-0.014	-0.442	-0.156	0.150	0.084	-0.099
12	56.39	12	<.0001	0.006	0.061	0.054	-0.048	-0.078	0.022
18	65.08	18	<.0001	0.068	-0.012	-0.096	-0.041	0.148	-0.043
24	77.28	24	<.0001	-0.089	0.067	0.148	-0.037	-0.099	-0.091



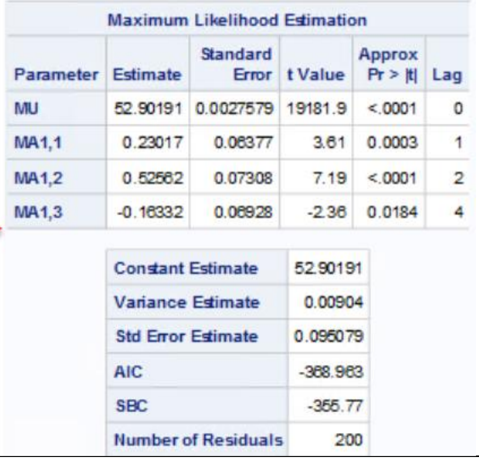


For e2 dataset the first glace after plotting the time series, I also don't see any significant trend and seasonality. I first try to identify the variable y from autocorrelation table I see that it is rejecting the Y as a white noise and both ACF and PACF are delaying with small significant lags, so I will fit the ARMA model to the data. From ACF PACF it is suggesting the model ARMA(2,1) but I can still test it with 3 different approach: ESACF, MINIC and SCAN options in proc arima. I got some candidates models for (p,q): (0,4) from esacf, (3,0) from MINIC, and (2,1), (3,0), (0,3) from SCAN. I will now try to fit them with the parameters and find the validated candidate models.



Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	52.90187	0.0026161	20221.3	<.0001	0
MA1,1	0.20157	0.07095	2.84	0.0045	1
MA1,2	0.49274	0.07181	6.86	<.0001	2
MA1,3	0.09581	0.07213	1.33	0.1841	3
MA1,4	-0.17581	0.07143	-2.46	0.0138	4

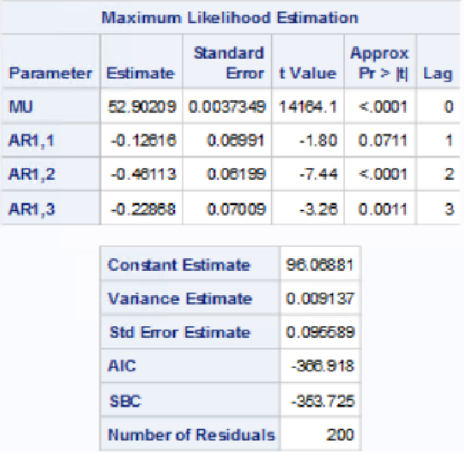


Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	52.90191	0.0027579	19181.9	<.0001	0
MA1,1	0.23017	0.06377	3.61	0.0003	1
MA1,2	0.52562	0.07308	7.19	<.0001	2
MA1,3	-0.16332	0.06928	-2.36	0.0184	4

Constant Estimate	52.90191
Variance Estimate	0.00904
Std Error Estimate	0.095079
AIC	-368.963
SBC	-355.77
Number of Residuals	200

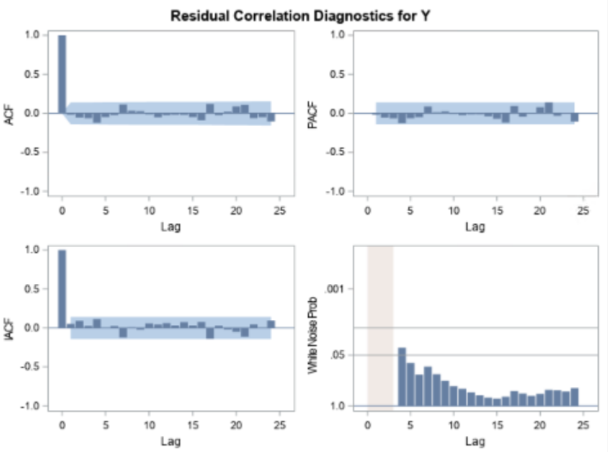
(0,4) For MA4 maximum likelihood suggesting that the lag 3 is not significant also for the autocorrelation check lag 6 reject white noise, so I remove lag 3. Now all the parameter are significant with autocorrelation not rejecting white noise test and residual is normally distributed, it is a valid model with AIC=-368,963



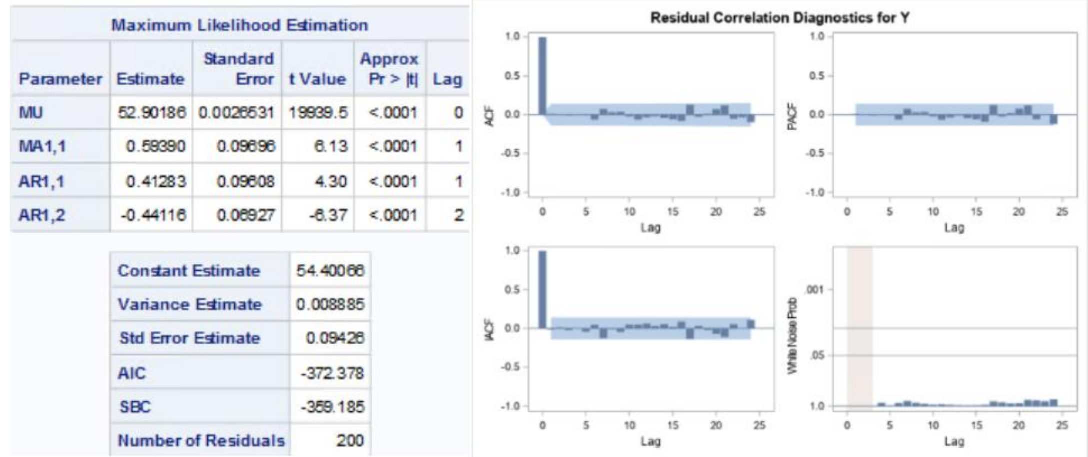
Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	52.90209	0.0037349	14164.1	<.0001	0
AR1,1	-0.12616	0.06991	-1.80	0.0711	1
AR1,2	-0.46113	0.06199	-7.44	<.0001	2
AR1,3	-0.22868	0.07009	-3.26	0.0011	3

Constant Estimate	96.06881
Variance Estimate	0.009137
Std Error Estimate	0.095589
AIC	-366.918
SBC	-353.725
Number of Residuals	200



(3,0) For AR3 model, ML suggesting that lag 1 is not significant enough and not white noise test is almost past, I tried to remove lag 1. Even by removing the lag 1 the white noise test still not totally past so I stick to the original p=3 model as the white noise test almost past, it is almost a valid model with AIC=-366.918



Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	52.90186	0.0026531	19939.5	<.0001	0
MA1,1	0.59390	0.09696	6.13	<.0001	1
AR1,1	0.41283	0.09608	4.30	<.0001	1
AR1,2	-0.44116	0.06927	-6.37	<.0001	2

Constant Estimate	54.40066
Variance Estimate	0.008885
Std Error Estimate	0.09426
AIC	-372.378
SBC	-359.185
Number of Residuals	200

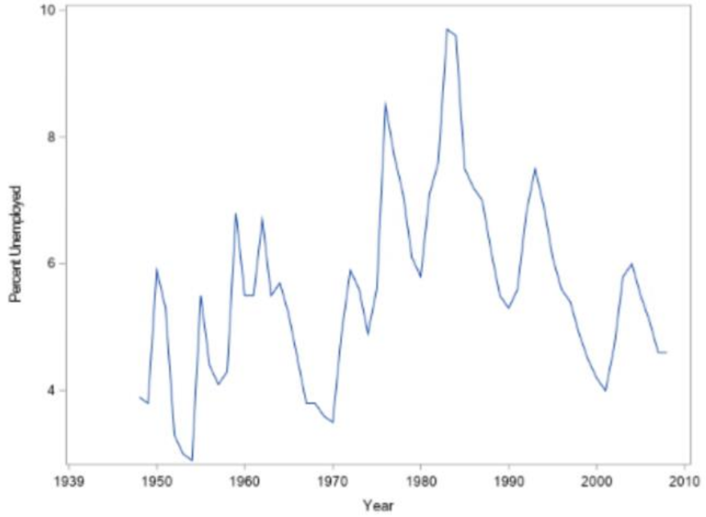
Residual Correlation Diagnostics for Y

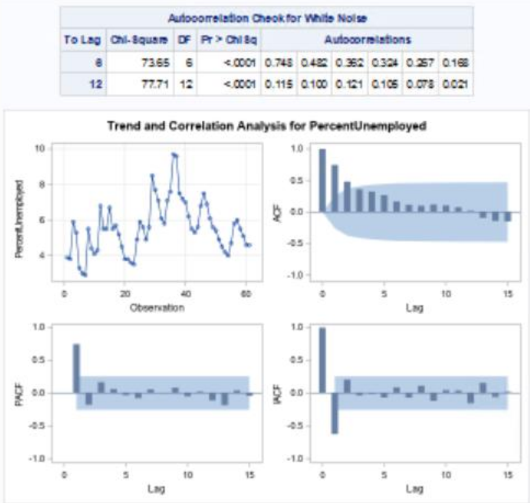
(2,1)	For ARMA (2,1) model, the parameters' p-value are all very small to be significant. The residuals also pass the white noise test and the normality check. It is a valid model with AIC=-372.378
(0,3)	Here the lag 3 parameter is not significant and the white noise test is rejected, as I have a good candidate above, I would not keep fitting it.

Therefore, in terms of AIC score and ML estimation and the white noise test, ARMA(2,1) is the best model for this dataset.

1.3. E3 dataset

Perform the Ljung-Box White Noise Probability test on the variable PercentUnemployed in the SAS data set E3. You should give the null and alternative hypothesis. What can you conclude from this test?

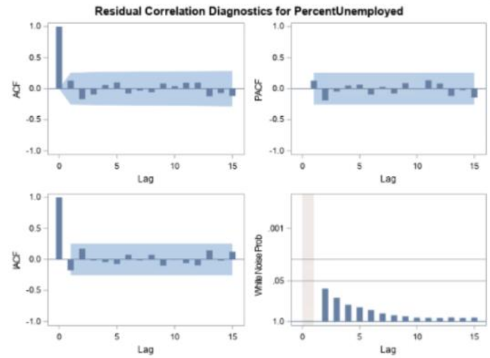


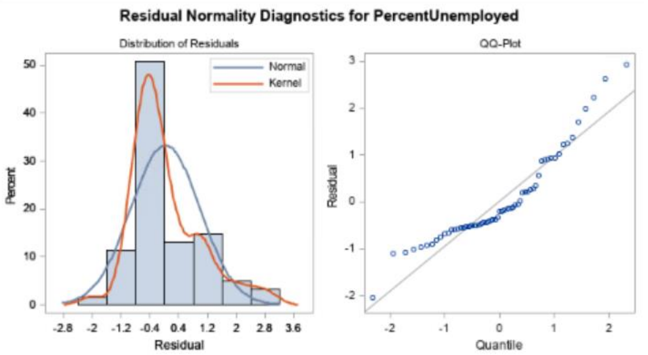


Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi Sq	Autocorrelations					
6	73.65	6	<.0001	0.748	0.482	0.362	0.324	0.257	0.168
12	77.71	12	<.0001	0.115	0.100	0.121	0.105	0.078	0.021

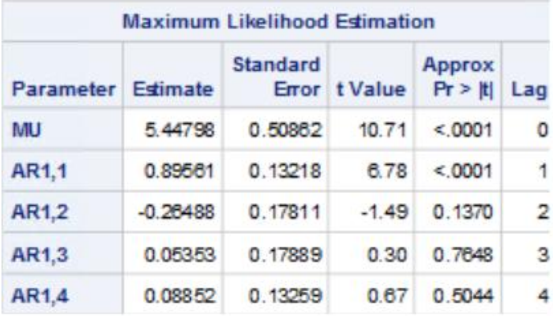
Trend and Correlation Analysis for PercentUnemployed

In e3 dataset I see that the data is rejecting the null hypothesis that it is a white noise and I cannot see clearly from the ACF and PACF plot that what is the model as above 2 exercises therefore I also try to check the best model with 3 different approach: ESACF, MINIC and SCAN options in proc arima. I got some candidates (p,q): (1,0), (4,0), (5,0), (0,2) from ESACF, (1,0), (0,2) from SCAN, (12,12) from MINIC but it is too big as lags so I would tried to fit the others first.



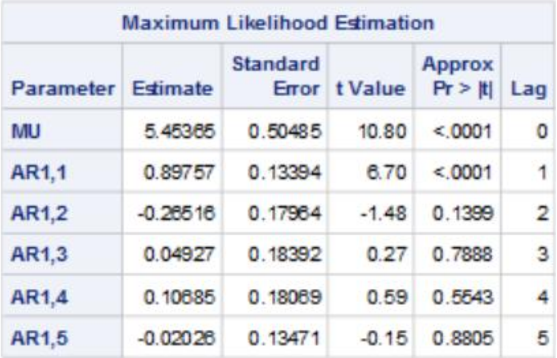


(1,0)	For this model the parameter AR1 is significant. The autocorrelation check and the white noise prob test suggesting that the residuals are white noise, but the normality of the residuals doesn't seems to fit quit Ill but not too far from fitting. So it is close to a valid model with AIC=171.9.



Maximum Likelihood Estimation

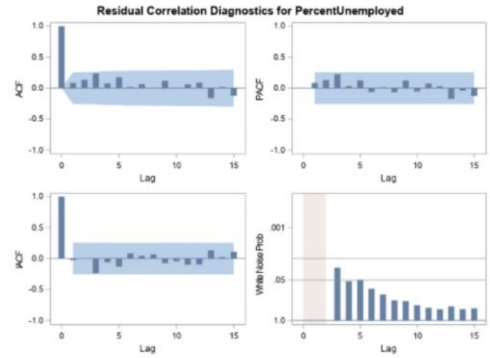
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	5.44798	0.50862	10.71	<.0001	0
AR1,1	0.89561	0.13218	6.78	<.0001	1
AR1,2	-0.26488	0.17811	-1.49	0.1370	2
AR1,3	0.05353	0.17889	0.30	0.7648	3
AR1,4	0.08852	0.13259	0.67	0.5044	4

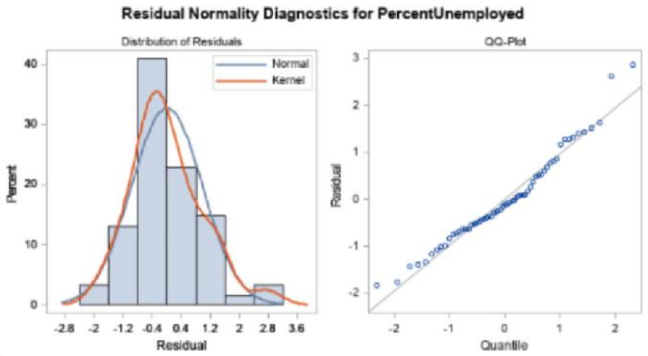


Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	5.45365	0.50485	10.80	<.0001	0
AR1,1	0.89757	0.13394	6.70	<.0001	1
AR1,2	-0.26516	0.17964	-1.48	0.1399	2
AR1,3	0.04927	0.18392	0.27	0.7888	3
AR1,4	0.10685	0.18069	0.59	0.5543	4
AR1,5	-0.02026	0.13471	-0.15	0.8805	5

(4,0)	For this model I can see from the graph above that the lag 2,3,4 are not significant parameters therefore if I remove them, it would be the same as AR1 model. It is an invalid model.
(5,0)	This model has the same result as AR4 that only lag 1 is significant so it is also an invalid model





(0,2)	For MA2 model both lag parameter are significant, auto correlation check is suggesting the that the residuals are white noise, but if for the white noise prob plot, lag one is passing the limit but it is very close. On the other hand the residuals normality of this model fit much better then AR1 model. So I said it is also close to valid model with AIC=175.74

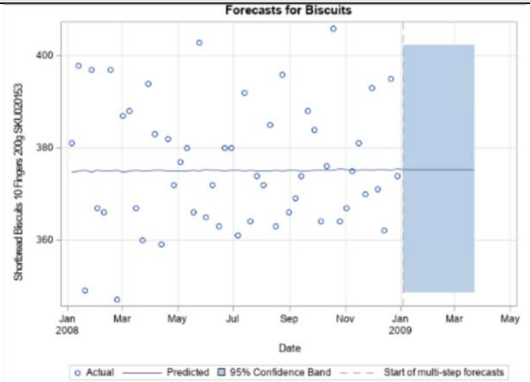
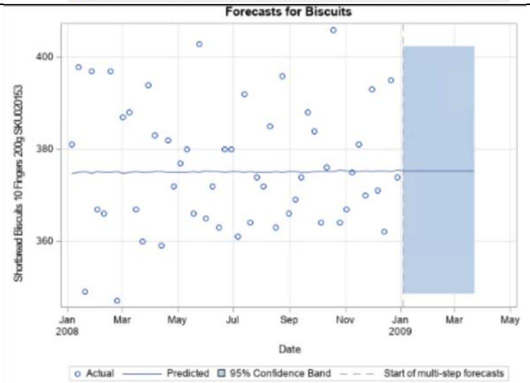
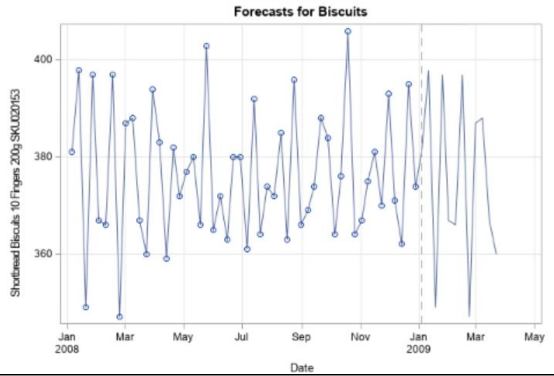
To conclude I have 2 almost validate models: AR1 and MA2, if one have to be chosen I would choose AR1 model as the final model since its residuals didn't reject Ljung-Box White Noise

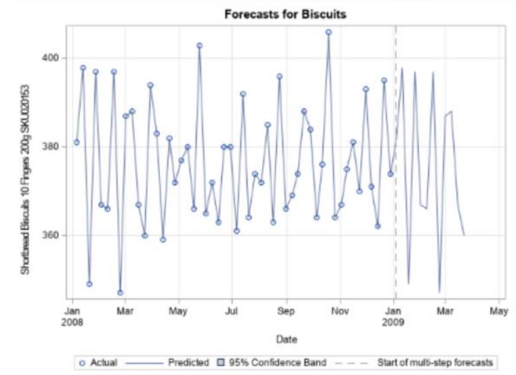
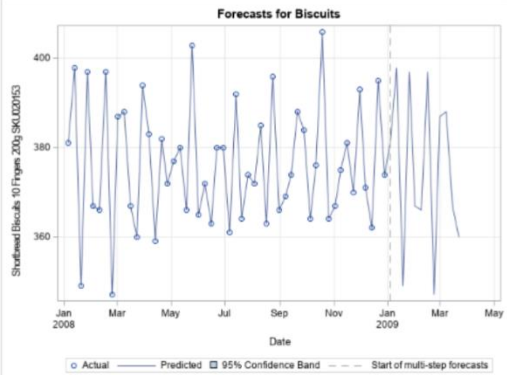
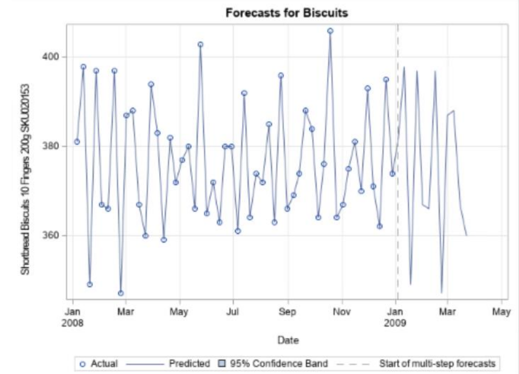
Probability test (with null hypothesis that the residuals are white noise it means that it is not time dependence. Also for that AR1 model's AIC is smaller then MA2 model.

1.4. E4 dataset

Using the PROC ESM in SAS, generate a forecast for the next 12 periods for the variable Biscuits in the SAS data set E4 with the model of your choice. Justify your choice.

When I first plot the model it seems like there's a seasonality in the biscuits data. The interval for this data is lekly data. I've tried multiple models in proc esm shown as below.

model	plot	MAPE score
Double		--
Simple		--
Addseasonal		0

Multseasonal		1.16E-8
Addwinters		0
Winters		2.14E-8

For double and simple model the curve for forecast doesn't fit Ill the data and also the confidence interval for them are too large to consider, so I would just compare the others, when I see other models they all fit really Ill the model, especially with addwinter and add seasonal they all have 0 score for MAPE.

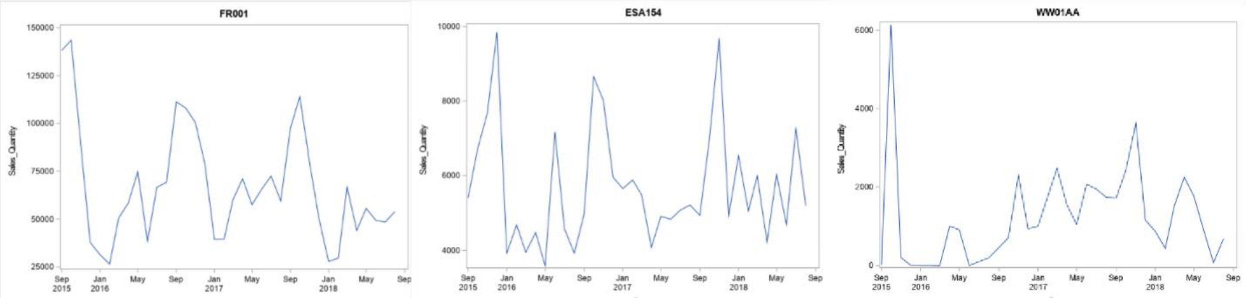
2. Part 2: Case study

The Sales department asked you to provide a statistical forecast for 3 key products for the next 16 months (last forecast in December 2019). You managed to extract the relevant data in the file DSTI_SAS_ETS_Evaluation_Part2.csv.

Using all what you have learned in Times Series in SAS, generate a forecast for the 3 different products. You will explain all the steps you have folloId to choose the models and you will write a quick report for the Sales department to understand the sales evolution of these products.

2.1. Modeling choosing steps

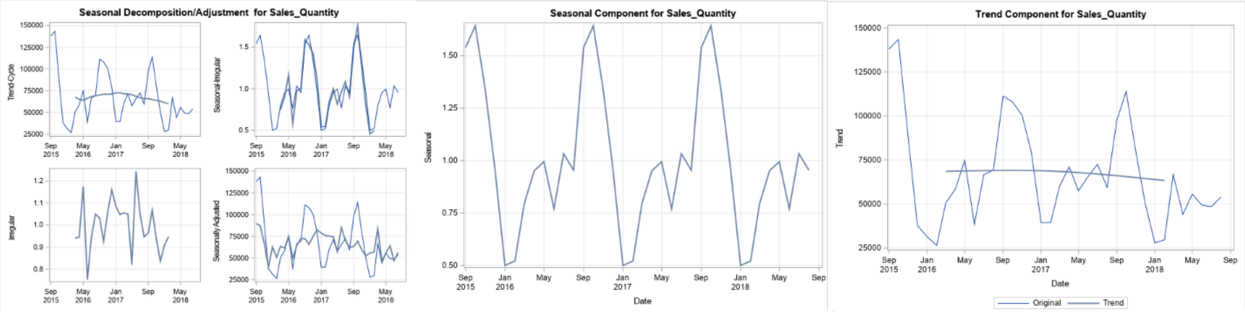
Plot 3 product with monthly data from September 2015 to August 2018 (3 years data). After having a glance of the data I would start modeling separately for 3 different data. Note that there are 3 missing value for product WW01AA I would I can do an average of t-1 and t+1, for t= Jan2016, Jul2016 and Sep2016.



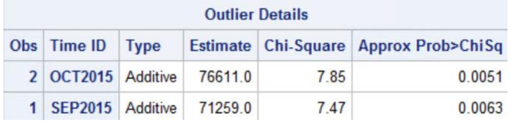
2.2. Product 1 : FR001

Modeling

Step 1: time series analyze, identify outliers

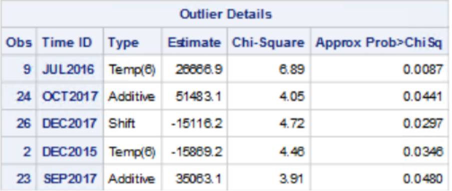


From the decomp graph it seems that this dataset had seasonality and without trend.



Outlier Details

Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
2	OCT2015	Additive	76611.0	7.85	0.0051
1	SEP2015	Additive	71259.0	7.47	0.0063



Outlier Details

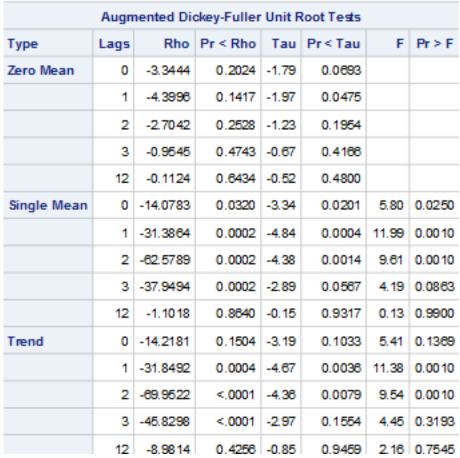
Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
9	JUL2016	Temp(6)	26666.9	6.89	0.0087
24	OCT2017	Additive	51483.1	4.05	0.0441
26	DEC2017	Shift	-15116.2	4.72	0.0297
2	DEC2015	Temp(6)	-15869.2	4.46	0.0346
23	SEP2017	Additive	35063.1	3.91	0.0480

If we only see the additive outliers, It seems like the data of October and September 2015 are the outliers without differencing. (will be removed for ESM models). There are some outliers that I would identify them in the modeling process with pro arimas.

Step 2: identify stationarity

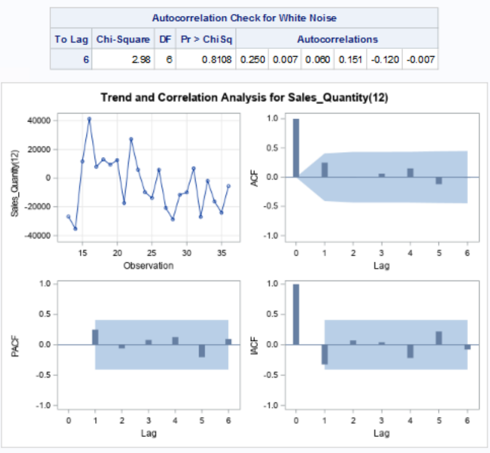
Stochastic approach for seasonality:

I first try to see if the differencing of the data is needed by ADF test (with null hypothesis differencing is needed) the result showing that I should differenciate by 3 and 12 since lag 3, 12 is accepting the null hypothesis also due to that the seasonality might exist.



Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-3.3444	0.2024	-1.79	0.0693		
	1	-4.3996	0.1417	-1.97	0.0475		
	2	-2.7042	0.2528	-1.23	0.1954		
	3	-0.9545	0.4743	-0.67	0.4166		
	12	-0.1124	0.6434	-0.52	0.4800		
Single Mean	0	-14.0783	0.0320	-3.34	0.0201	5.80	0.0250
	1	-31.3864	0.0002	-4.84	0.0004	11.99	0.0010
	2	-62.5789	0.0002	-4.38	0.0014	9.61	0.0010
	3	-37.9494	0.0002	-2.89	0.0567	4.19	0.0863
	12	-1.1018	0.8640	-0.15	0.9317	0.13	0.9900
Trend	0	-14.2181	0.1504	-3.19	0.1033	5.41	0.1369
	1	-31.8492	0.0004	-4.67	0.0036	11.38	0.0010
	2	-69.9522	<.0001	-4.36	0.0079	9.54	0.0010
	3	-45.8298	<.0001	-2.97	0.1554	4.45	0.3193
	12	-8.9814	0.4256	-0.85	0.9459	2.16	0.7545

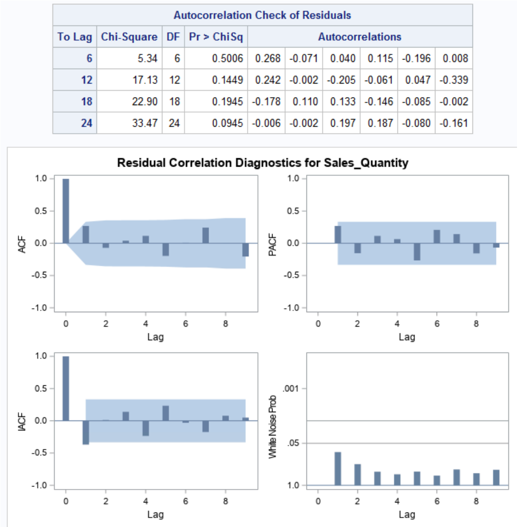


Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.98	6	0.8108	0.250	0.007	0.060	0.151	-0.120	-0.007

Trend and Correlation Analysis for Sales_Quantity(12)

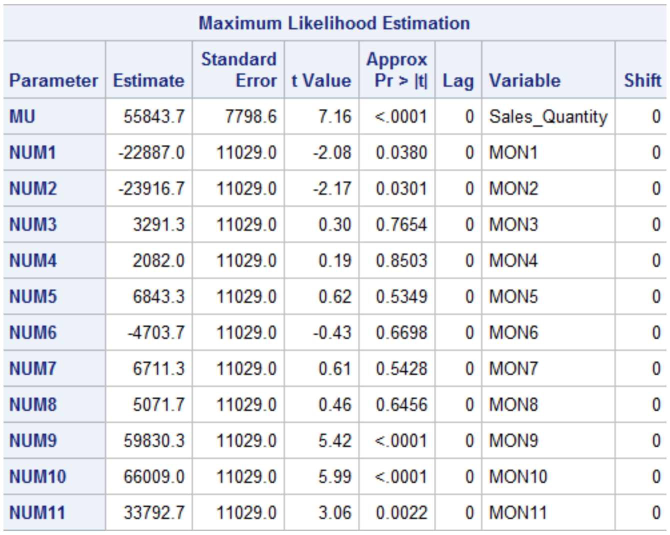
Deterministic approach by create dummy



Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.34	6	0.5006	0.268	-0.071	0.040	0.115	-0.196	0.008
12	17.13	12	0.1449	0.242	-0.002	-0.205	-0.061	0.047	-0.339
18	22.90	18	0.1945	-0.178	0.110	0.133	-0.146	-0.085	-0.002
24	33.47	24	0.0945	-0.006	-0.002	0.197	0.187	-0.080	-0.161

Residual Correlation Diagnostics for Sales_Quantity



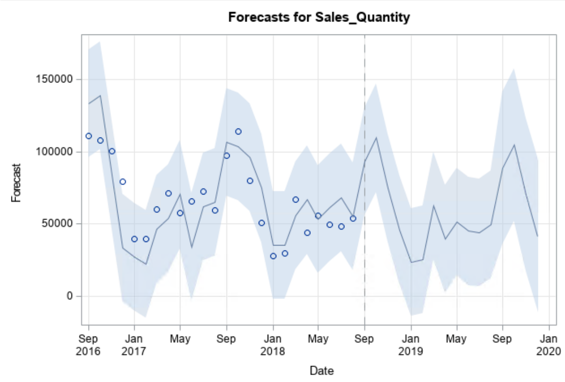
Maximum Likelihood Estimation

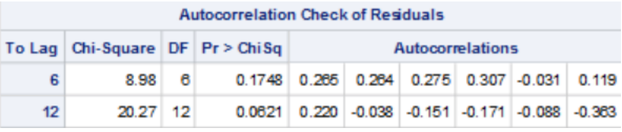
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	55843.7	7798.6	7.16	<.0001	0	Sales_Quantity	0
NUM1	-22887.0	11029.0	-2.08	0.0380	0	MON1	0
NUM2	-23916.7	11029.0	-2.17	0.0301	0	MON2	0
NUM3	3291.3	11029.0	0.30	0.7654	0	MON3	0
NUM4	2082.0	11029.0	0.19	0.8503	0	MON4	0
NUM5	6843.3	11029.0	0.62	0.5349	0	MON5	0
NUM6	-4703.7	11029.0	-0.43	0.6698	0	MON6	0
NUM7	6711.3	11029.0	0.61	0.5428	0	MON7	0
NUM8	5071.7	11029.0	0.46	0.6456	0	MON8	0
NUM9	59830.3	11029.0	5.42	<.0001	0	MON9	0
NUM10	66009.0	11029.0	5.99	<.0001	0	MON10	0
NUM11	33792.7	11029.0	3.06	0.0022	0	MON11	0

In this way after fitting the seasonality of the month the data become white noise for residuals, it is also important to see from the Maximum likelihood estimation that for this product January, February have an significant negative effect on the sales quantity whereas September to November has significant positive effect to the sales. This model is with AIC 796 therefore I chose the stochastic model with only fitting the monthly differencing.

Step 3: Identify parameters for the model and validate the model

If I tried differenciated the y value by 12(seasonality), the data is now a white noise, with the additive outlier December 2016. Removing I can start with the forecast it seems that the forecast is following Ill the past graph also with a small confidence interval. Here the AIC is 494.44. (candidate 1)

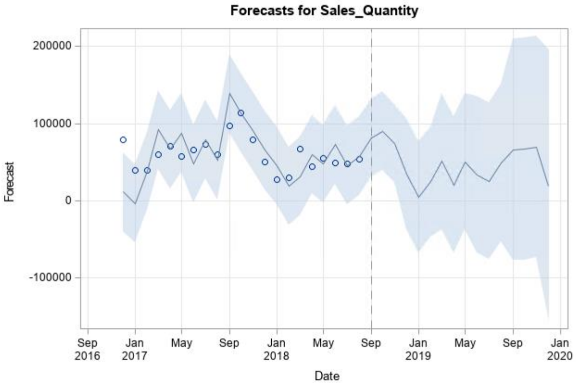


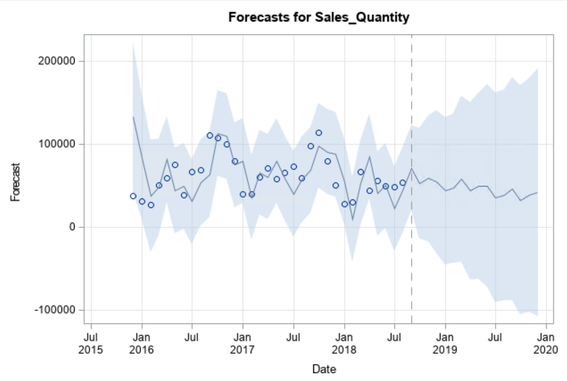


Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	8.98	6	0.1748	0.265	0.264	0.275	0.307	-0.031	0.119
12	20.27	12	0.0621	0.220	-0.038	-0.151	-0.171	-0.088	-0.363

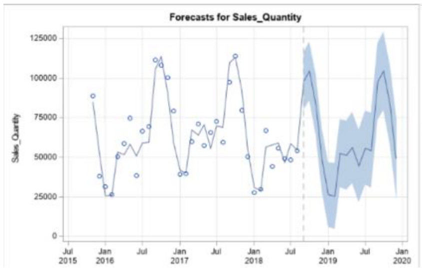
If I tried to differenciate 3 and 12 both, it is a white noise with active outliers are December 2016 june 2017 and October 2015. But the confidence interval of the model is too big that it is not a good candidate.

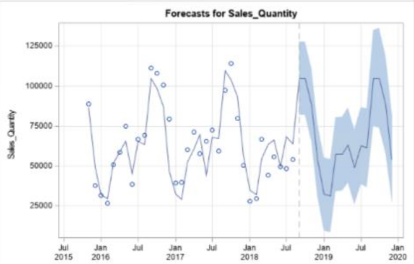


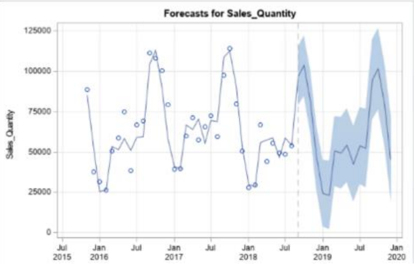


If I tried to only differenciate with 3, and the autocorrelation check is rejecting that it has a correlation so it is not a white noise I should fit ARMA model with 3 different approach: ESACF, MINIC and SCAN options in proc arima to find the good value for p and q, I got some candidate combinations (4,1), (0,2), (2,3) (1,4) (0,12), (12,3). For (1,4) and (2,3) model, only AR1 parameters seems to have important effect, so it's not a valid model, even if I tried to fit only AR1 model it is still not a white noise. (0,2) is not valid, since the parameters are not significant. Eventually for (1,4) seems like p=(1 3 4) are significant, it become a valid model with residuals white noise, therefore I use this model to do forecasting. In the end the confidence interval for the forecast is still big so I do not consider this model (graph above on the right).

Step 4: try ESM models

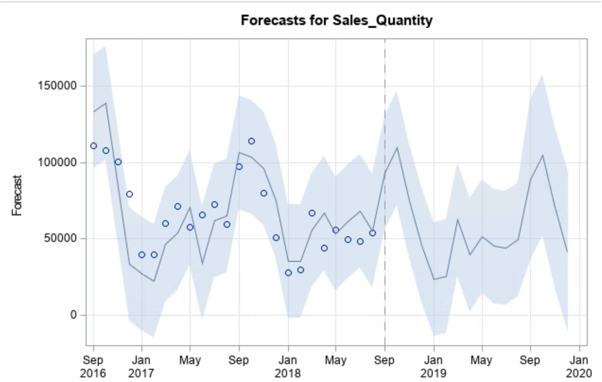


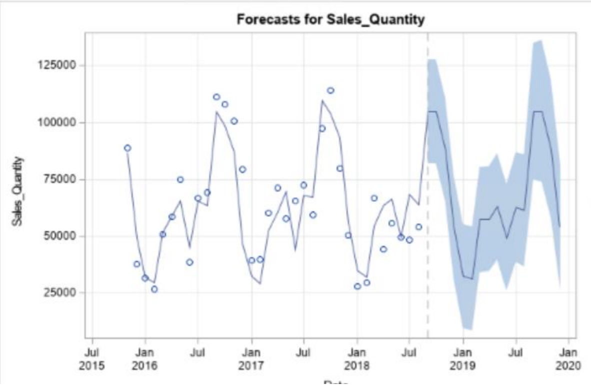




Here I first remove the what seems to be a outlier September and October 2015 data then to fit the model. In the model addseasonal (graph 1), addwinters (graph 2) and winters (graph 3) has smaller interval, with MAPE score: 12.65, 12.25 and 14.47, AIC score: 674, 673 and 676 separately. Therefore the ESM addwinters model is the best among all the ESM models.

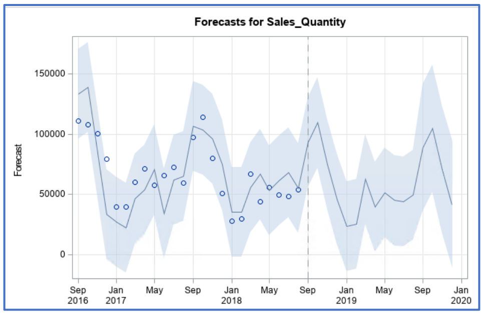
Step 5: Choose the best model





For ARIMA, I've chosen only fit the seasonality differencing model (on the left). And the ESM addwinters mode on the right. After comparing both AIC ARIMA model seems to fit better before continuing tuning the ARIMA model, ESM model perform better. But the deterministic model help as Ill to describe some general behavior monthly.

Quick summary: Sales report

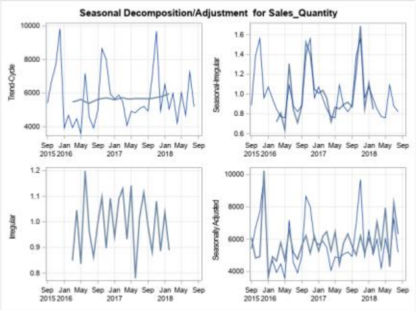


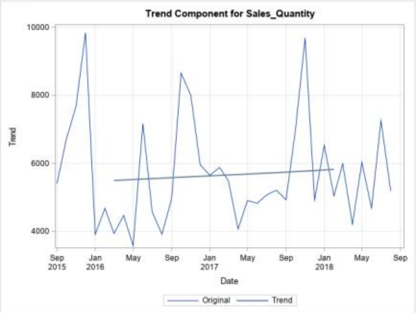
For product FR001 the model suggesting that in the end of the year September and October has a high peak for the sales quantity then it would start dropping very fast until the beginning of next year (January and February). It is important to check with the department to understand why there's this drop whether it's due to the internal or external reason and to take some action about this change (like some promotions) to increase the sales.

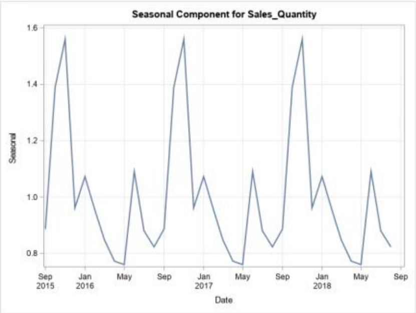
2.3. Product 2: ESA154

Modeling

Step 1: time series analyze, identify outliers





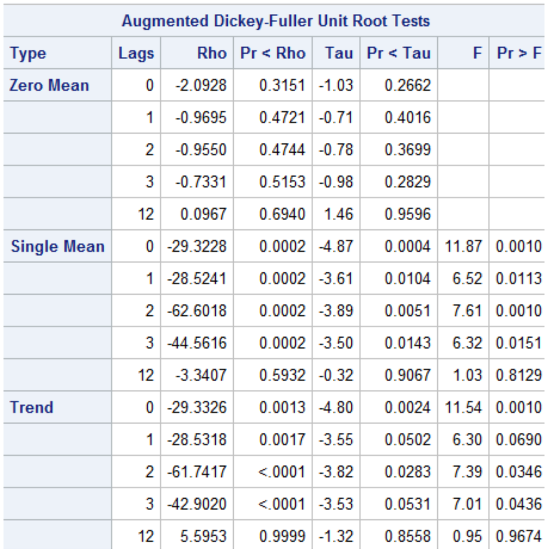


From the decomp graph it seems that this dataset had seasonality and without trend as Ill. No significant outliers seen.

Step 2: identify stationarity

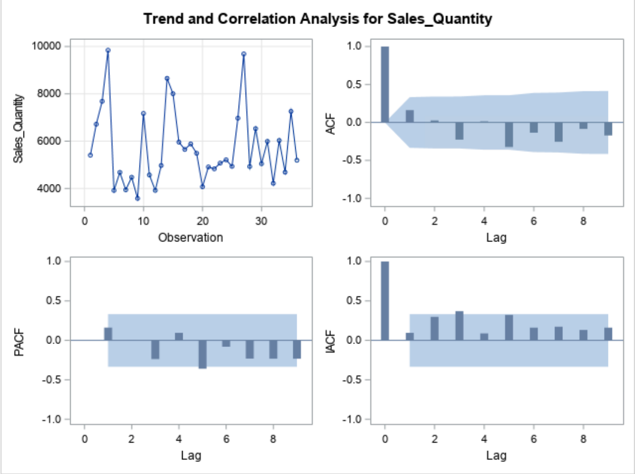
Stochastic approach for seasonality:

I first try to see if the differencing of the data is needed by ADF test (with null hypothesis differencing is needed) the result showing that I should differenciate by 12 since lag 12 is accepting the null hypothesis also due to that the seasonality might exist.

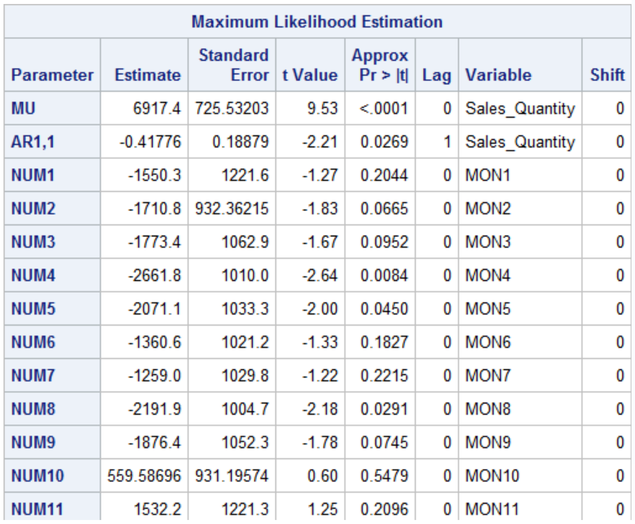


Augmented Dickey-Fuller Unit Root Tests

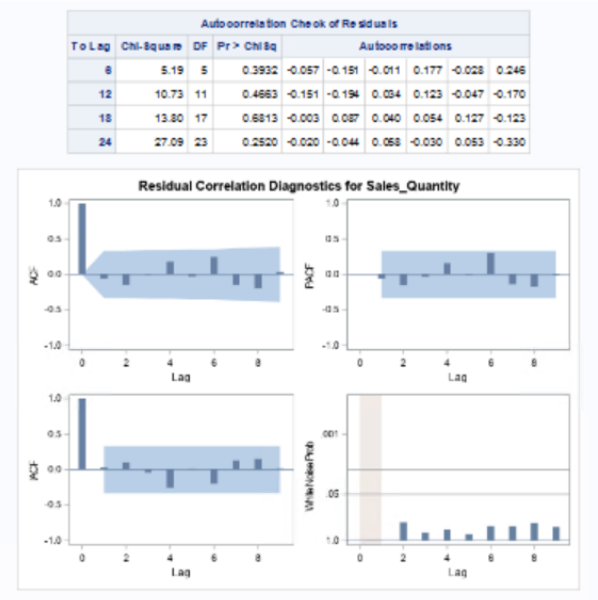
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-2.0928	0.3151	-1.03	0.2662		
	1	-0.9695	0.4721	-0.71	0.4016		
	2	-0.9550	0.4744	-0.78	0.3699		
	3	-0.7331	0.5153	-0.98	0.2829		
	12	0.0967	0.6940	1.46	0.9596		
Single Mean	0	-29.3228	0.0002	-4.87	0.0004	11.87	0.0010
	1	-28.5241	0.0002	-3.61	0.0104	6.52	0.0113
	2	-62.6018	0.0002	-3.89	0.0051	7.61	0.0010
	3	-44.5616	0.0002	-3.50	0.0143	6.32	0.0151
	12	-3.3407	0.5932	-0.32	0.9067	1.03	0.8129
Trend	0	-29.3326	0.0013	-4.80	0.0024	11.54	0.0010
	1	-28.5318	0.0017	-3.55	0.0502	6.30	0.0690
	2	-61.7417	<.0001	-3.82	0.0283	7.39	0.0346
	3	-42.9020	<.0001	-3.53	0.0531	7.01	0.0436
	12	5.5953	0.9999	-1.32	0.8558	0.95	0.9674



Deterministic approach by create dummy



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	6917.4	725.53203	9.53	<.0001	0	Sales_Quantity	0
AR1,1	-0.41776	0.18879	-2.21	0.0269	1	Sales_Quantity	0
NUM1	-1550.3	1221.6	-1.27	0.2044	0	MON1	0
NUM2	-1710.8	932.36215	-1.83	0.0665	0	MON2	0
NUM3	-1773.4	1062.9	-1.67	0.0952	0	MON3	0
NUM4	-2661.8	1010.0	-2.64	0.0084	0	MON4	0
NUM5	-2071.1	1033.3	-2.00	0.0450	0	MON5	0
NUM6	-1360.6	1021.2	-1.33	0.1827	0	MON6	0
NUM7	-1259.0	1029.8	-1.22	0.2215	0	MON7	0
NUM8	-2191.9	1004.7	-2.18	0.0291	0	MON8	0
NUM9	-1876.4	1052.3	-1.78	0.0745	0	MON9	0
NUM10	559.58696	931.19574	0.60	0.5479	0	MON10	0
NUM11	1532.2	1221.3	1.25	0.2096	0	MON11	0

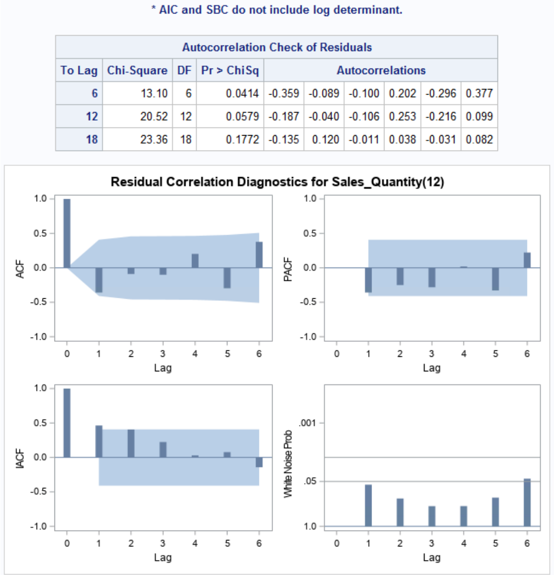


Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.19	5	0.3932	-0.057	-0.151	-0.011	0.177	-0.028	0.246
12	10.73	11	0.4663	-0.151	-0.194	0.034	0.123	-0.047	-0.170
18	13.80	17	0.6813	-0.003	0.087	0.040	0.054	0.127	-0.123
24	27.09	23	0.2520	-0.020	-0.044	0.058	-0.030	0.053	-0.330

Residual Correlation Diagnostics for Sales_Quantity

In this model, after fitting AR1 and seasonality to the model it became a validate white noise residuals model. From the maximum likelihood estimation I can see that for product 2 I see that in February, April, May and August are the month that have significant negative effects to the sales. The AIC of this model is 619.

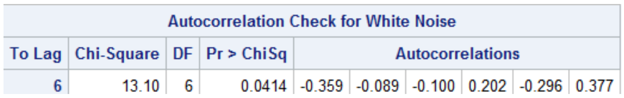
Step 3: Identify parameters for the model and validate the model



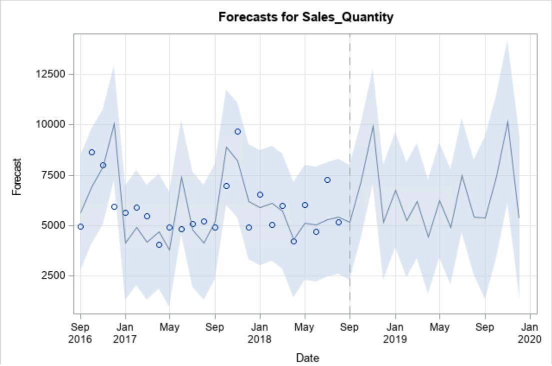
* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	13.10	6	0.0414	-0.359	-0.089	-0.100	0.202	-0.296	0.377
12	20.52	12	0.0579	-0.187	-0.040	-0.106	0.253	-0.216	0.099
18	23.36	18	0.1772	-0.135	0.120	-0.011	0.038	-0.031	0.082

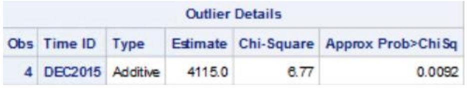
Residual Correlation Diagnostics for Sales_Quantity(12)



Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	13.10	6	0.0414	-0.359	-0.089	-0.100	0.202	-0.296	0.377

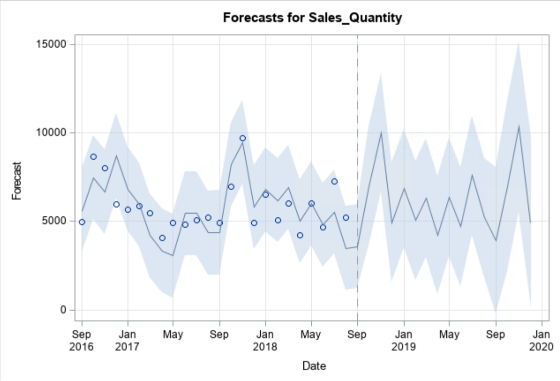


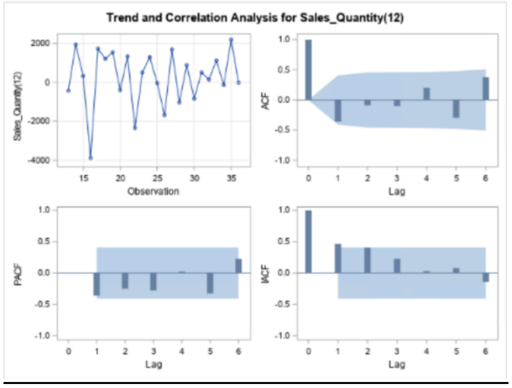
Forecasts for Sales_Quantity



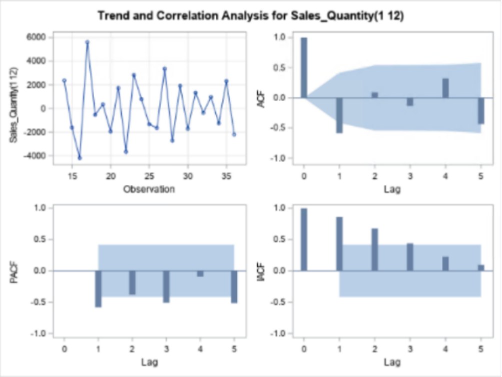
Outlier Details					
Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
4	DEC2015	Additive	4115.0	6.77	0.0092

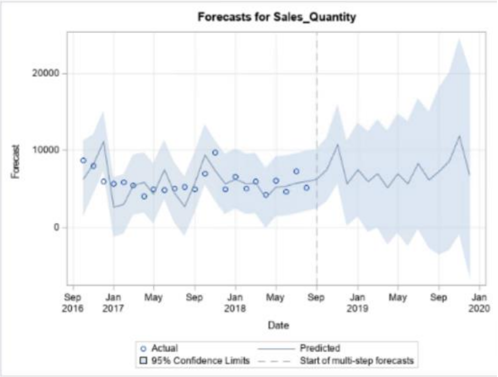
Seems like seasonal differencing is needed in this dataset, after differencing the variable y by 12 the autocorrelation check for white noise is rejecting the test with null hypothesis is suggesting that it is a not a white noise, so I can fit ARMA model to try to make the residuals to become white noise. After many combination of the ma and ar models, in the end p=1 and q=(2), the residuals are white noise and all the parameters are significant. It is with AIC equal to 411.





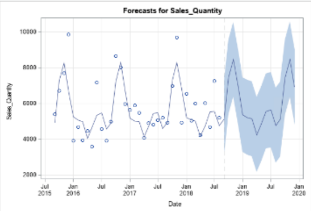
Without differencing I also tried to find the parameters for ARMA model as last dataset. The candidates are (1,1), (0,12), (0,0), (4,1), (3,2), (12,0). After checking they all are not valid models.

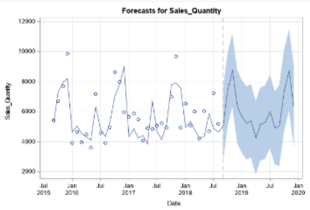


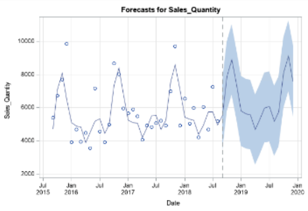


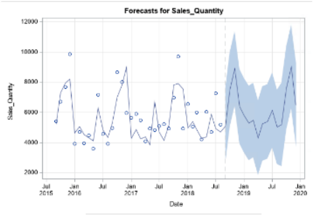
However the plot after differencing 12 seems to still have some trend inside so I tried to differenciate (1 12). Now the plot doesn't seems to contain patterns. But the autocorrelation still suggesting that it is not a white noise so I fit some combinations of (p, q) for the arma models. Finally with p=1 seems to be closely to valid model but the interval for the forecast is too big so I don't take it as a candidate model.

Step 4: try ESM models



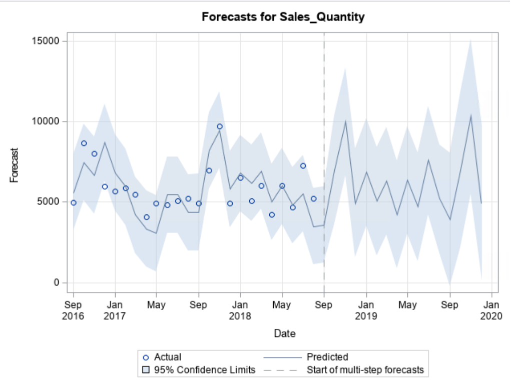


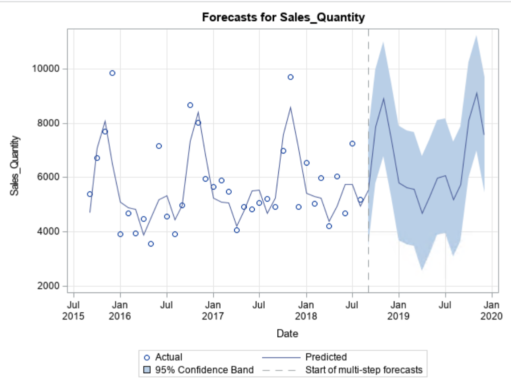




In the model addseasonal (graph 1), multseaonal (graph2), addwinters (graph 3) and winters (graph 4) has smaller interval, with MAPE score: 13.87, 14.93, 12.71 and 16.42, AIC score: 503, 515 and 504.8 separately. Therefore the ESM addseasonal model is the best among all the ESM models for the second product.

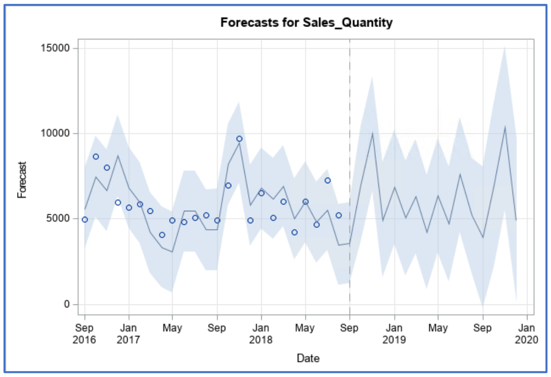
Step 5: Choose the best model





For ARIMA, I've chosen the seasonality differencing model (on the left) along with ARMA p=1 and q=(2),. And the ESM addseasonal mode on the right, ARIMA model forecasting seems to be better fitting the data also with the AIC score smaller. But the deterministic model help as Ill to describe some general behavior monthly.

Quick summary: Sales report

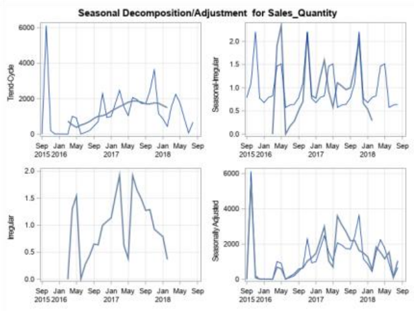


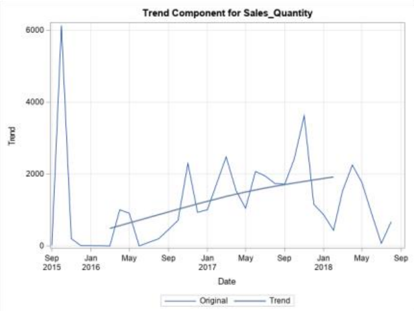
For product 2 ESA154, what I see is that there's always a high sales point in October every years, whereas other months are quite stable randomly changes the sales, so it is important to check the reason on this and to take the action plan accordingly.

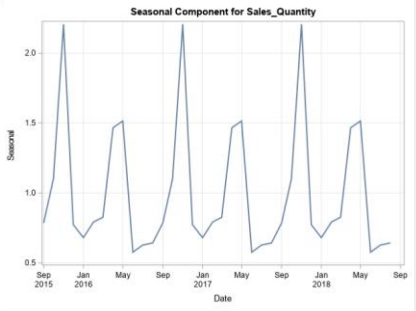
2.4. Product 3: WW01AA

Modeling

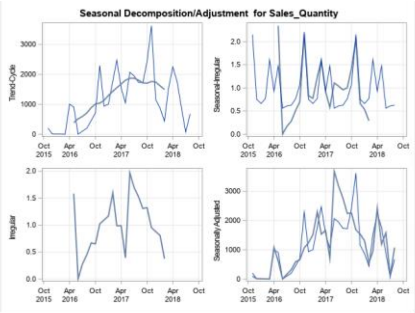
Step 1: time series analyze, identify outliers

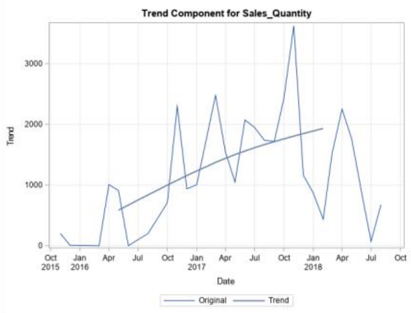


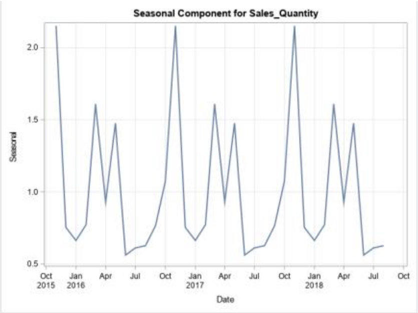




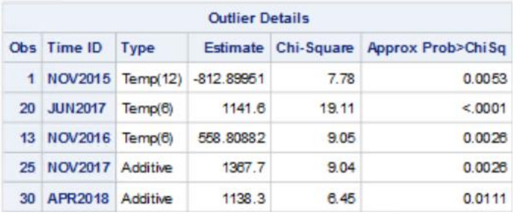
From the decom graph I can see that there exist a seasonality per year and also a linear trend in this model. Take a look at the end of 2015, this data might be consider as outlier. So I remove the date before November 2015. Then the data become as below:







Even after cutting the first two month of data to get a more stationary data it seems to still have some outliers to be process in the upcoming modeling.



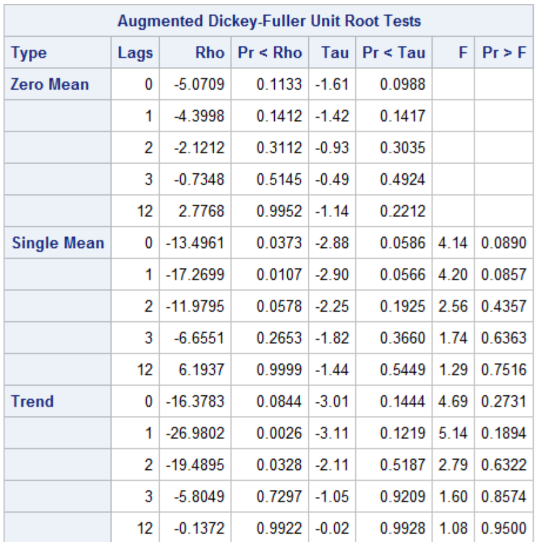
Outlier Details

Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
1	NOV2015	Temp(12)	-812.89951	7.78	0.0053
20	JUN2017	Temp(6)	1141.6	19.11	<.0001
13	NOV2016	Temp(6)	558.80882	9.05	0.0026
25	NOV2017	Additive	1367.7	9.04	0.0026
30	APR2018	Additive	1138.3	6.45	0.0111

Step 2: identify stationarity

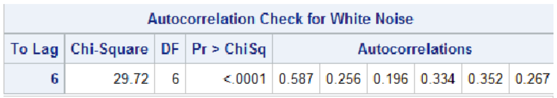
Stochastic approach for seasonality:

The autocorrelation check is saying that the data is not white noise it needs to be process.



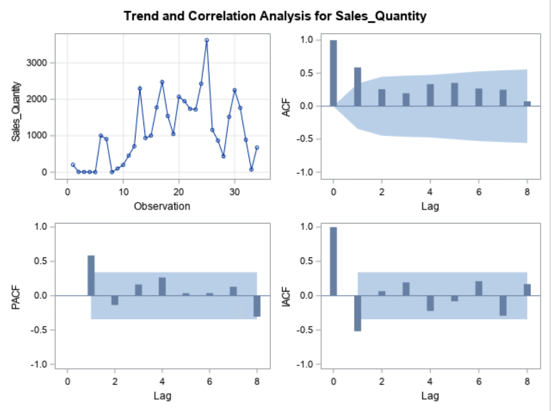
Augmented Dickey-Fuller Unit Root Tests

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-5.0709	0.1133	-1.61	0.0988		
	1	-4.3998	0.1412	-1.42	0.1417		
	2	-2.1212	0.3112	-0.93	0.3035		
	3	-0.7348	0.5145	-0.49	0.4924		
	12	2.7768	0.9952	-1.14	0.2212		
Single Mean	0	-13.4961	0.0373	-2.88	0.0586	4.14	0.0890
	1	-17.2699	0.0107	-2.90	0.0566	4.20	0.0857
	2	-11.9795	0.0578	-2.25	0.1925	2.56	0.4357
	3	-6.6551	0.2653	-1.82	0.3660	1.74	0.6363
	12	6.1937	0.9999	-1.44	0.5449	1.29	0.7516
Trend	0	-16.3783	0.0844	-3.01	0.1444	4.69	0.2731
	1	-26.9802	0.0026	-3.11	0.1219	5.14	0.1894
	2	-19.4895	0.0328	-2.11	0.5187	2.79	0.6322
	3	-5.8049	0.7297	-1.05	0.9209	1.60	0.8574
	12	-0.1372	0.9922	-0.02	0.9928	1.08	0.9500

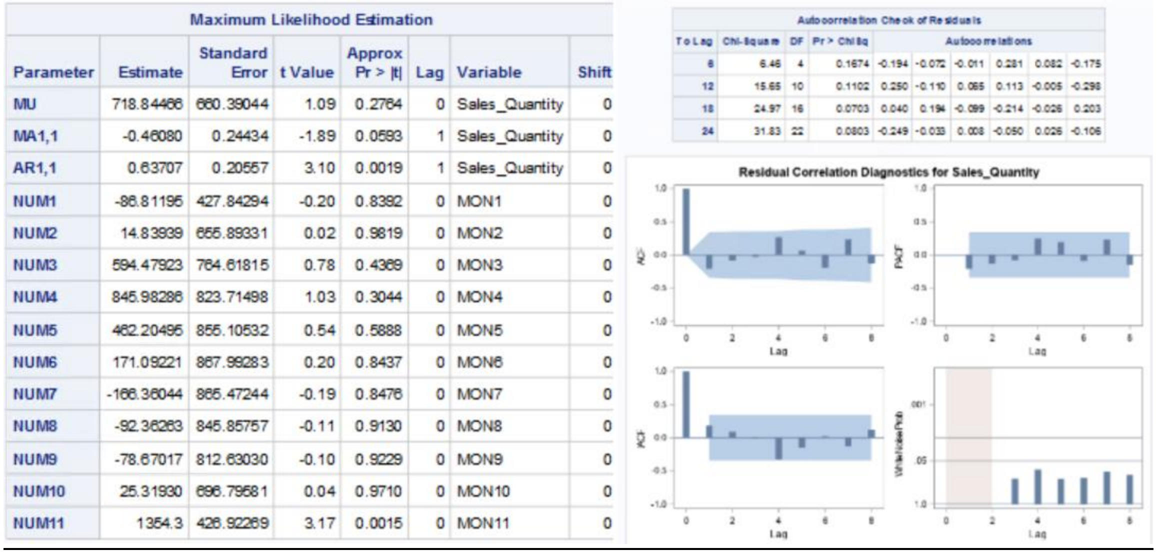


Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	29.72	6	<.0001	0.587	0.256	0.196	0.334	0.352	0.267



Deterministic approach by create dummy



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	718.84466	660.39044	1.09	0.2764	0	Sales_Quantity	0
MA1,1	-0.46080	0.24434	-1.89	0.0593	1	Sales_Quantity	0
AR1,1	0.63707	0.20557	3.10	0.0019	1	Sales_Quantity	0
NUM1	-86.81195	427.84294	-0.20	0.8392	0	MON1	0
NUM2	14.83939	655.89331	0.02	0.9819	0	MON2	0
NUM3	594.47923	764.61815	0.78	0.4369	0	MON3	0
NUM4	845.98286	823.71498	1.03	0.3044	0	MON4	0
NUM5	462.20495	855.10532	0.54	0.5888	0	MON5	0
NUM6	171.09221	867.99283	0.20	0.8437	0	MON6	0
NUM7	-166.36044	865.47244	-0.19	0.8476	0	MON7	0
NUM8	-92.36263	845.85757	-0.11	0.9130	0	MON8	0
NUM9	-78.67017	812.63030	-0.10	0.9229	0	MON9	0
NUM10	25.31930	696.79581	0.04	0.9710	0	MON10	0
NUM11	1354.3	426.92269	3.17	0.0015	0	MON11	0

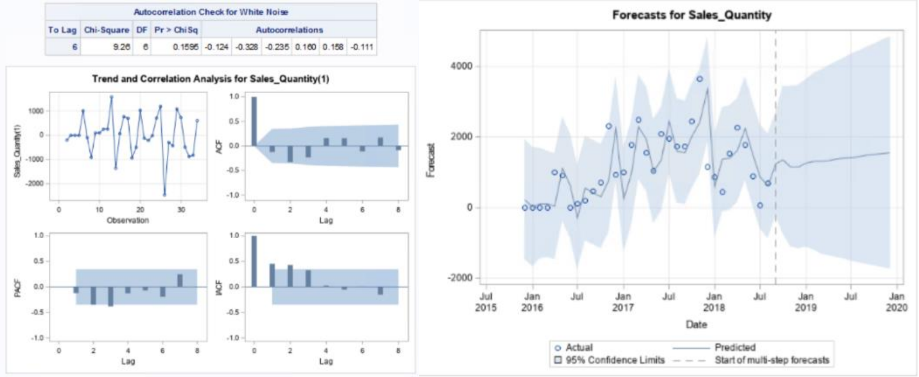
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	6.46	4	0.1674	-0.194	-0.072	-0.011	0.281	0.082	-0.175
12	15.65	10	0.1102	0.250	-0.110	0.065	0.113	-0.005	-0.298
18	24.97	16	0.0703	0.040	0.194	-0.099	-0.214	-0.026	0.203
24	31.83	22	0.0803	-0.249	-0.033	0.008	-0.050	0.026	-0.106

Residual Correlation Diagnostics for Sales_Quantity

In this way after fitting the seasonality of the month and then take ARMA(1,1) the data become white noise for residuals, it is also important to see from the Maximum likelihood estimation that for this product November has an significant positive impact on the sales quantity. Finally this model is with AIC score equal to 550.

Step 3: Identify parameters for the model and validate the model

Now this model simply just do 1 differencing it become white noise. I try to tune the model from 3 different method ESACF, MINIC and SCAN options in proc arima. There are some candidate (p,q) to be try, if I can improve the performance by fitting more parameters.



Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.26	6	0.1595	-0.124	-0.328	-0.235	0.160	0.158	-0.111

Trend and Correlation Analysis for Sales_Quantity(1)

Forecasts for Sales_Quantity

possible candidates:

- MINIC: (2,3) ; p=(2) q=(3) AIC=534
- ESACF: (0, 0), (1, 0), (2, 0), (3, 0) ; p=(2) AIC=536
- Scan: (5,2)

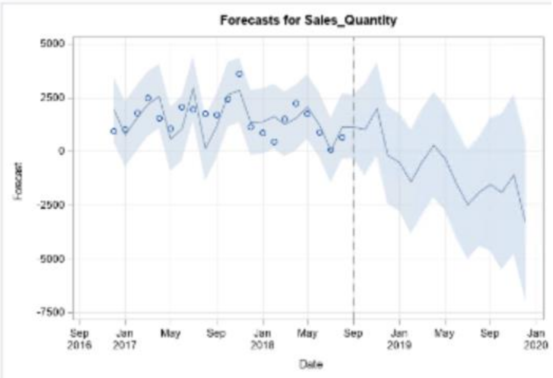
Although the final model it fit well the data, the forecast is not good. Therefore I try other differencing: 2. Below are the possible candidates:

- MINIC: (3,1) ;

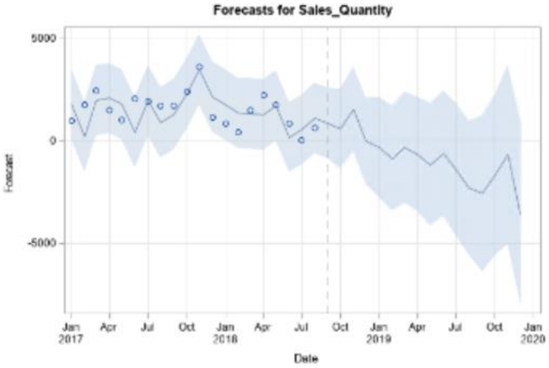
- ESACF: (0, 2), (1, 2), (2, 0), (3, 0) ; AR2 AIC=523
- Scan: (0,2), (2,1)

The forecast is still the same most likely due to the seasonality not identified. After adding the seasonality to the model the forecast seems more reasonable so I fit the ARMA model to it. After trying many combinations.

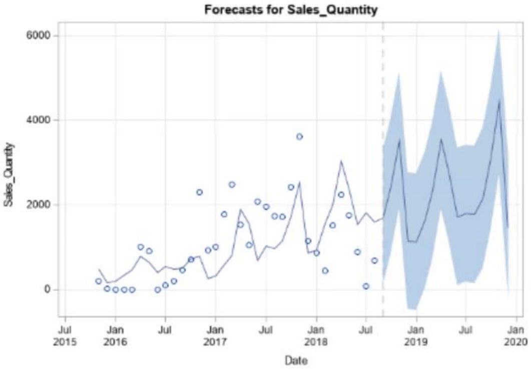
- For differencing (1 12): The best model is ARMA (2, 3) with AIC=341



- For differencing (1 12): there are 2 candidates models ARMA p=(2 12) q=1 with AIC=331 and ARMA p=(2) q=1 with AIC=335 therefore ARMA p=(2 12) q=1 is better than differencing one .

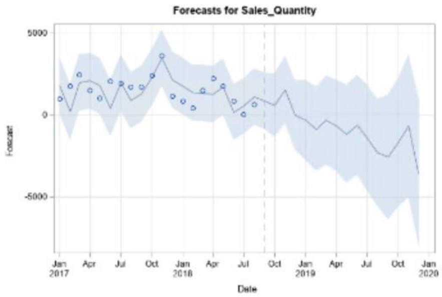


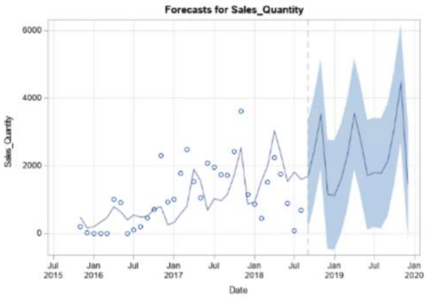
Step 4: try ESM models



From the interval of forecast only winters model has a smaller interval with AIC score equal to 459.

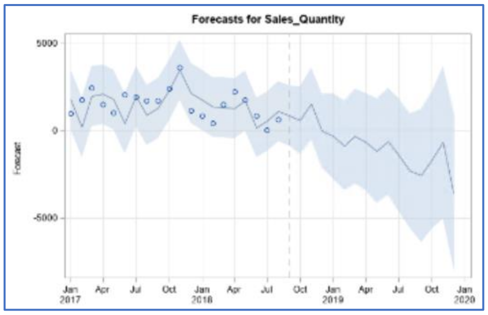
Step 5: Choose the best model





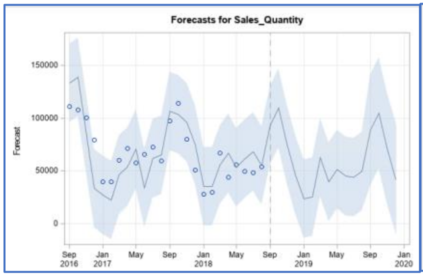
For ARIMA, I've chosen the seasonality differencing model (on the left) along with the parameters differencing (1 12) p=(2 12) q=1 with AIC=331, it is better then the ESM winters model (on the right). But the deterministic model help as Ill to describe some general behavior monthly.

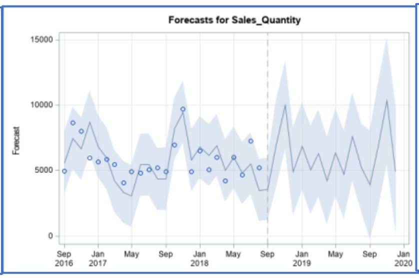
Quick summary: Sales report

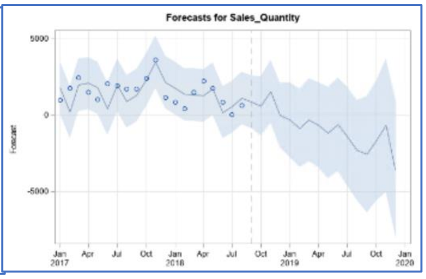


For this product there are 3 missing data in 2016: Jan2016, Jul2016 and Sep2016. First need to identify the reason for it. I can see the trend is going down and that from the deterministic approach I've identified that the sales in November (end of the year) increases.

2.5. Summary for 3 products







Due to the deadline the models for 3 products are not perfect as shown in the graph above but they all capture some sales evolutions. Especially there are some common points out of 3 products that is that the sales tend to increase in the end of the year and decrease in the beginning of the year. With this insight the sales department can take more actions with some promotion plan in general.