TIME SERIES FORECASTING
WEEKLY SALES OF WALMART STORE

University of Connecticut
Business Analytics & Project Management
OPIM 5671 - Data Mining and Business Intelligence

Group 6:

Kamal Kannan Krishnan
Deexith Reddy
Mukund Ramachandran
Namita Singh
Xiaoyue Wang

TABLE OF CONTENTS

1.0 Executive Summary 3

2.0 Problem Statement 3

3.0 Data Description 4

4.0 Data Preparation and Modification 4

5.0 Methodology and Results 6

- **5.1 Forecasting Total Sales of all departments** 6
- **5.2 Result of Model Comparison (Total Weekly Sales of All Departments)** 9
- **5.3 Forecasting Sales of 5 Individual Departments** 12
  - **5.3.1 Candy and Tobacco Department Sales** 12
  - **5.3.2 Electronics Department Sales** 12
  - **5.3.3 Groceries Department Sales** 13
  - **5.3.4 Gasoline Department Sales** 13
  - **5.3.5 Jewelry Department Sales** 14

6.0 Business Finding and Recommendations 14

7.0 References 16

Appendix 17

1.0 Executive Summary

The project aims to forecast the weekly sales in Walmart stores using time series forecasting methodology with SAS Studio 3.8 software. It uses historical data of weekly sales in Walmart Store Number 34 between February 05, 2010 and November 01, 2012 across 78 different departments to forecast the future weekly sales. The objective of this study is broadly classified into 2 categories-

1) To forecast Total Weekly sales of all departments in store 34
2) To analyze the weekly sales in 5 individual departments, namely Gasoline, Grocery, Candy & tobacco, Electronics, and Jewelry in store 34.

The team built multiple time series models including Exponential Smoothing Methods (ESM) and ARIMAX. Based on comparison of Error metrics, the best model is the Arimax Model with 1st order of seasonal differencing and input variables of Unemployment (with first order autoregression), MarkDown_3 (with a shift of 3). This model is data driven and parsimonious.

2.0 Problem Statement

Retail industry in the USA is faced with growing concern of intense competition and increasingly tighter margins. With advent of technological advancements and consumer awareness, many E-commerce companies have scaled up their market share in this sector. The existing traditional Brick and Mortar stores are met with growing competition from online retail companies which operate on different business models. Overall SWOT analysis of this industry points to the conclusion that it becomes imperative to forecast future Sales to balance the trade and take proactive actions to satisfy the customer requirements. Forecasting sales will help in controlling and managing the following inputs-

1. Operational Efficiency- Supply chain management
2. Inventory control - Implementation of Just in time
3. Marketing & Promotions- Predicting strategies for customer promotions and discounts
4. Financial Planning- Budgeting and Internal Controls

This project attempts to analyse all the macro and micro aspects of the basic challenges given in the data set and come out with strategic insights by in depth data analysis.

3.0 Data Description

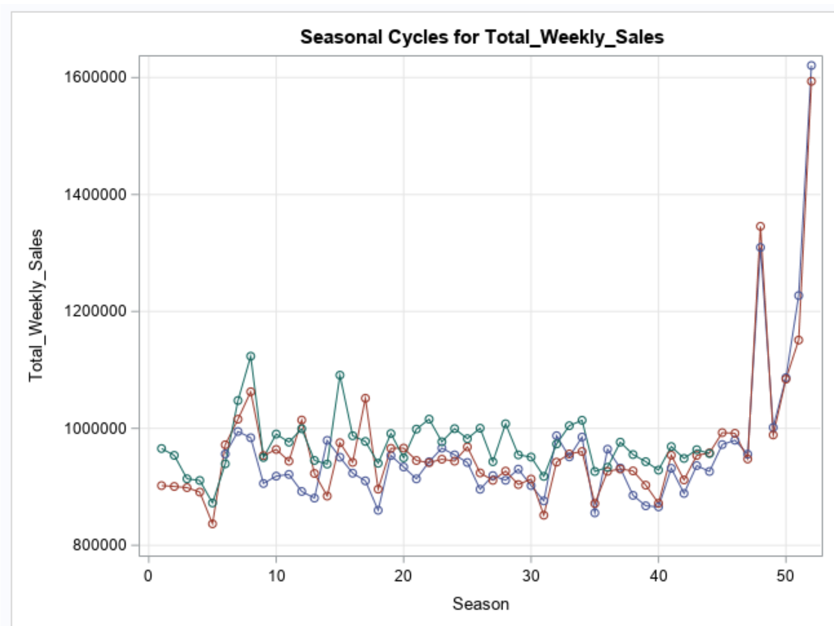
The dataset (train.csv) consists of 421571 rows and 16 columns, which keeps track of Walmart's weekly sales records at 99 departments in 45 stores from February 05, 2010 to November 01, 2012 (143 weeks). The columns are as follows:

- **Time Variable:**
  - **Date -** the end date of weekly cycle from Feb 2010 to Nov 2012.
- **Dependent Variable:**
  - **Weekly Sales -** Sales for a given department in the given store in dollars($).
- **Independent Variables (Used in Modeling):**
  - **Is Holiday -** Whether the week is a holiday week where **1 - Holiday** and **0 - Not a Holiday**. Holidays are Labor day, Valentines day, Super Bowl, Thanksgiving and Christmas.
  - **Temperature (Fahrenheit) -** Temperature in the region where the store is located.
  - **Fuel Price -** Cost of fuel in dollars ($) in the region where the store is located.
  - **Mark Down1 -** 1st round of Promotional Markdown in dollars ($).
  - **MarkDown2 -** 2nd round of Promotional Markdown in dollars ($).
  - **Mark Down3 -** 3rd round of Promotional Markdown in dollars ($).
  - **Mark Down4 -** 4th round of Promotional Markdown in dollars ($).
  - **Mark Down5 -** 5th round of Promotional Markdown in dollars ($).
  - **CPI -** Consumer Price Index.
  - **Unemployment -** Unemployment rate of the store in percentage.
- **Independent Variables (Not used in Modeling)**
  - **Dept -** The department number numbered from **1-99** in Walmart.
  - **Store -** Store numbered from **1-45** in Walmart.
  - **Size -** Sizes in Square Feet for each Store.
  - **Type -** Types of Stores labeled with A, B and C.

4.0 Data Preparation and Modification

To prepare the data, we merged the two datasets train.csv and features.csv by Store and Date to create a merged dataset. Since we are focusing on Store34, we filter and create a new dataset of 10224 rows named Store34_all. As there are missing values in the columns of MarkDown_1-5, which would potentially affect the functioning of the time series models, we replaced all missing values with 0 in the Store_34 dataset. Meanwhile, as the input variables for the time series models are required to be numerical, we built a new column of IsHoliday_numeric as the dummy variable for the original character variable IsHoliday. Here,1 refers to "is a holiday", while 0 refers to "is not a holiday".

Thereafter, we further subset the Store34_all dataset in two ways. On the one hand, we aggregated the weekly sales for each department in the Store34_all dataset which creates 143 rows named Store34 that reflects the total weekly sales in Store34. On the other hand, we also output 5 separate subsets (each with 143 rows) to reflect the weekly sales in each of the 5 aforementioned departments of grocery, jewelry, electronics, gasoline and candy & tobacco, respectively.



During the process of modeling, based on the general rule of no more than ¼ of the data in the holdout sample, we set the last 20 rows (June 25, 2012 to October 26, 2012) of the 143 rows as our holdout sample for validation. The results of our later forecasting also show that the number of rows in our holdout set satisfies the rule that 'at least 4 time points are required for every parameter to be estimated in a model' (Fernandez, et al. 2019, 58), as the number of parameters in our final models are usually less than 5. In addition, as shown in the screenshot above, although the seasonal pattern in each year slightly differs from one another, the patterns in the last 7 weeks of the year (covering the 2 primary holidays of Thanksgiving and Christmas) are very similar to one another in each year. Therefore, in general, more variations among years are reflected in the ordinary weekdays than in the holidays. As our validation set is composed of 1 holiday and 19 workdays, it will keep closer track of the models' performance in predicting the ordinary workdays.

5.0 Methodology and Results

5.1 Forecasting Total Sales of all departments

Our primary forecasting model for the total weekly sales of Store34 is based on the Store34 dataset. The modeling process is as follows.

1) **Decomposition analysis for the target variable - Total_Weekly_Sales**

According to the decomposition analysis (see Appendix 1), the time series Total_Weekly_Sales has a stable seasonality, which annually peaks in February (around Valentine's Day) November (around Thanksgiving) and December (around Christmas), and declines to valleys in April, August and January. At the same time, there is also a weak/slightly increasing trend in the time series. As the trend is not extremely clear, we decided to build two ESM models - an additive seasonal model and an additive winters model.

2) **The ESM models**

Model	MAPE	WMAE
Additive Seasonal	2.3%	21527
Additive Winters	2.017%	17289

As shown in the charts above, both the MAPE and WMAE values of the additive winters model are smaller than that of the additive seasonal model. However, the residuals (especially those of the first 2 lags and lag50 and beyond) for both models are not white noise (See Appendix 2 & 3), indicating that some explanatory variables are needed, in order to fully capture the features of the time series Total_Weekly_Sales, which leads us to the Following Arimax models.

3) Determining order of differencing in the target variable - Total_Weekly_Sales

Order of Differencing	Order of Seasonal Differencing	Standard Deviation
0	0	104630
1	0	124472
0	1	38189
0	2	39564

As shown in the chart above, when adding in a first order differencing, the standard deviation of the time series increases, decreasing its stationarity. Therefore, no order of differencing is required for the Total_Weekly_Sales series. Meanwhile, the standard deviation of the time series significantly decreases from 124472 to 38189 when a first order seasonal differencing is applied, while it slightly increases when a second order differencing is added. Therefore, a first order seasonal differencing would be the best choice for the time series data. However, as a first order seasonal differencing would eliminate 52 rows from the 143 rows of the dataset, which sacrifices more than ⅓ of the entire information thus might potentially affect the actual accuracy of the model. Besides the best scenario of an Arimax model based on the first order seasonal differencing, we also built an Arimax model based on the original time series without any order of differencing to compare with it.

4) Autocorrelation check for the target variable Total_Weekly_Sales

Scenario1: Based on the original time series

The PACF and ACF plots (see Appendix 4 (1)) indicate that the potential choices for the p value includes 1, 4 and 5, while the potential choices for the q value includes 1, 2 and 4. To choose the best Base ARMA model, we tried 56 combinations of p and q values and compared each of their AIC and SBC values (See Appendix 9) to get the best scenario of p=0, q=1,2,4, in which both AIC and SBC are the smallest. This is our base ARMA model.

Scenario2: Based on the time series with first order seasonal differencing

Based on the ACF and PACF plots (See Appendix 4 (2)), both p and q equal to 0. Therefore, the Base ARIMA model is p=0, q=0, sdif=1.

5) Autocorrelation check for all independent variables

According to the autocorrelation checks (See Appendix 5), Temperature, Fuel_Price, CPI, Unemployment, MarkDown_1, MarkDown_2, MarkDown_4, MarkDown_5 all have strong (>0.5) first order autocorrelations, While IsHoliday_numeric has a weak 11th order autocorrelation.

6) Cross-correlation check between all independent variables and target variable Total_Weekly_Sales

Scenario1: Based on the original time series

According to the CCF plots (See Appendix 6 (1)), the target variable is weakly (<0.5) correlated with Temperature (shift=0), IsHoliday (shift=4), MarkDown_2 (shift=6), MarkDown_3 (shift=0), MarkDown_5 (shift=3).

Scenario2: Based on the time series with first order seasonal differencing

Based on the CCF plots (See Appendix 6 (2)), the target variable is weakly correlated with Unemployment (shift=0), MarkDown_1 (shift=8), MarkDown_2 (shift=13), MarkDown_3 (shift=3), MarkDown_4 (shift=8), MarkDown_5 (shift=4).

7) Fitting model by backward stepwise variable selection method

Scenario1: Based on the original time series

Inputs	AIC	SBC
(1) Temperature 4 $ (11)IsHoliday_numeric 6 $ (1) MarkDown_2 MarkDown_3 3 $ MarkDown_5	3255	3289
- (1) MarkDown_2	3253	3284

The variable selection gives the results of the final Arimax Model 1 as a model with 0 order of differencing, 0 order of autocorrelation, 1st, 2nd, and 4th order of moving average and inputs variables of Temperature (with first order autoregression), IsHoliday_numeric (with 11th order of autoregression and a shift of 4), MarkDown_2 (with a shift of 6), MarkDown_3, and MarkDown_5 (with a shift of 3).

Scenario2: Based on the time series with first order seasonal differencing

Inputs	AIC	SBC
(1) Unemployment 8 $ (1) MarkDown_1 13 $ (1) MarkDown_2	2179	2209

3 $ MarkDown_3 8 $ (1) MarkDown_4 4 $ (1) MarkDown_5		
- MarkDown_4	2175	2200
- MarkDown_1	2171	2191
- (1) MarkDown_2	2170	2187
- (1) MarkDown_5	2168	2180
- (1) Unemployment	2168	2178
- 13 $ MarkDown_2	2170	2177

The variable selection gives the results of the final Arimax Model 2 as a model with 1st order of seasonal differencing and inputs variables of Unemployment (with first order autoregression), MarkDown_3 (with a shift of 3).

5.2 Result of Model Comparison (Total Weekly Sales of All Departments)

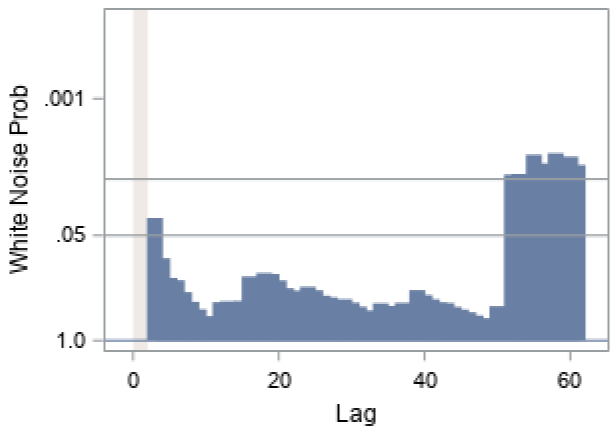
The aforementioned four models on Total_Weekly_Sales of Store 34 are compared by the following measurements of MAPE, WMAE, residuals and forecast plots (See Appendix 11). For the calculation of WMAE, holidays are weighted 5 times as an ordinary workday.

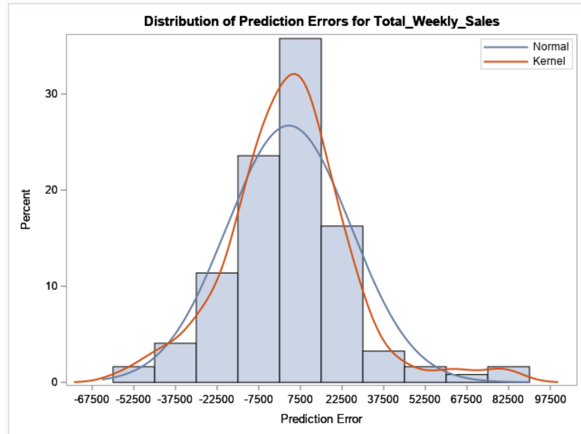
1) MAPE & WMAE

Model	MAPE	WMAE
Additive Seasonal	2.3%	21527
Additive Winters	2.017%	17289
Arimax 1	6.13%	60290
Arimax 2	2.05%	18133

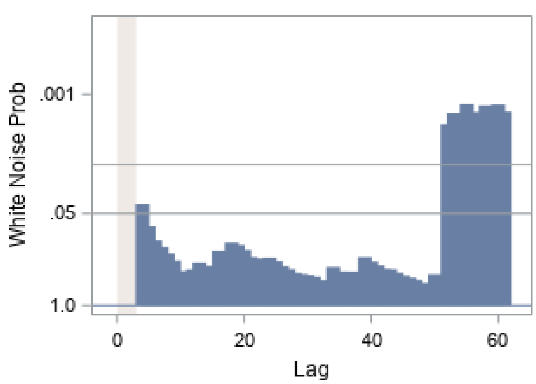
2) Residuals

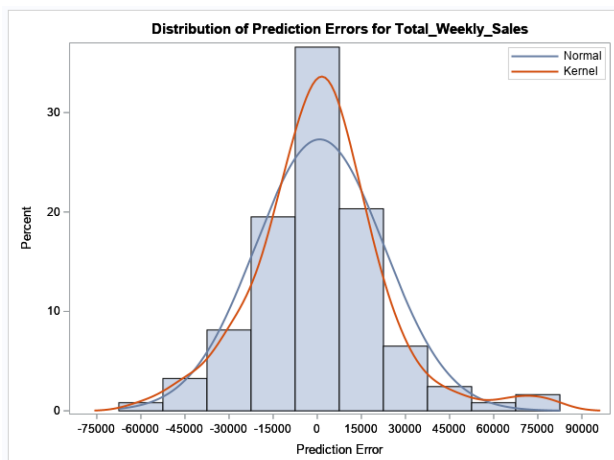
Additive Seasonal:



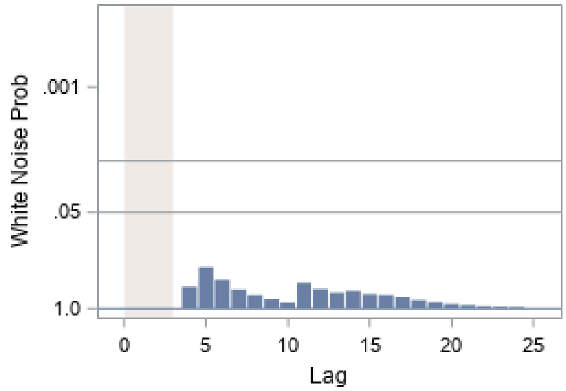


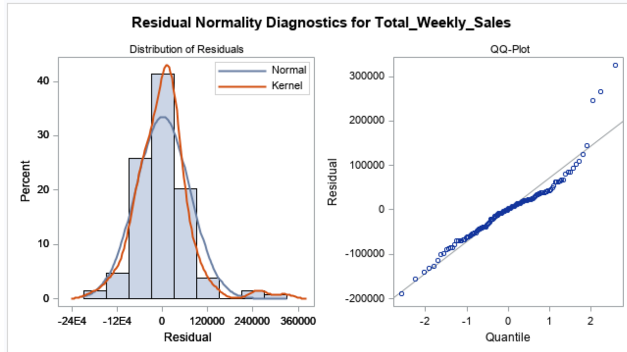
Additive Winters:



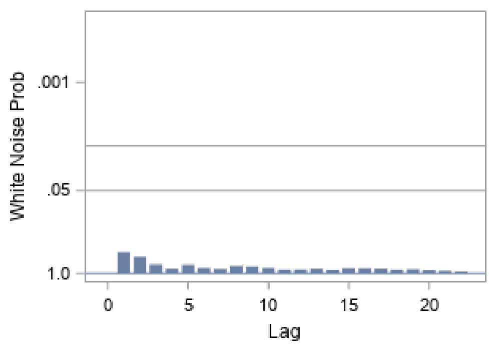


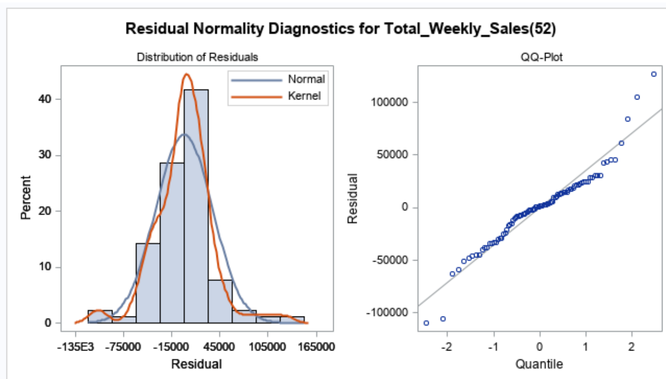
Arimax 1:





Arimax 2:





Based on the comparison above, in general, the additive winters model and the Arimax model 2 did better than the other two models.

In terms of MAPE and WMAE, the additive winters model gives the most robust forecasts. Accordingly, it serves as a sound choice in serving pure forecasting purposes. However, as its residuals are not completely white noise, it does not fully capture the features of the Total_Weekly_Sales series, therefore, lacks the capacity in explaining the causes of sales changes beyond seasonal and trend factors.

Both the Arimax Model 1 and the Arimax Model 2 are successful in fully capturing the major features of the time series. The Arimax Model 1 did a better job than all other models in capturing the sales fluctuation between late June and early July. However, failing in capturing the seasonality factor within the time series, the model made a significant false forecast on a non-existing sales peak, which significantly harms its accuracy.

The Arimax Model 2 is slightly worse than the additive winters model in terms of the MAPE and the WMAE, however, the differences are minor. At the same time, as its residuals are white noise, it successfully captures all major features of the time series by introducing two external/ explanatory elements (the unemployment and MarkDown_3), which makes it a sound choice for examining the external causes (beyond seasonality and trend elements) of the sales changes.

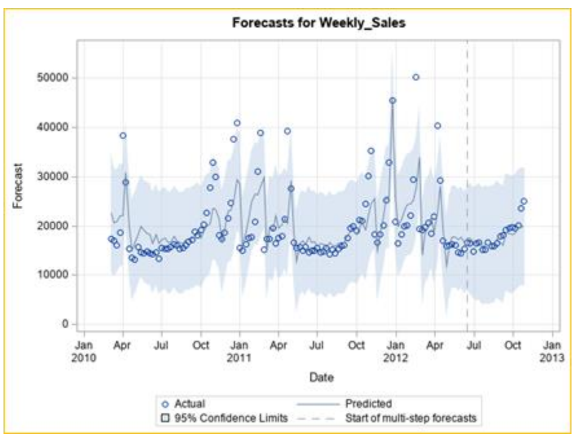
5.3 Forecasting Sales of 5 Individual Departments

The similar approaches aforementioned in step 1) to 7) are also applied to forecast the weekly sales of five (5) individual departments in Store 34. The departments are Candy & tobacco, Electronics, Groceries, Gasoline and Jewelry. The main observations are only stated here for each department.

5.3.1 Candy and Tobacco Department Sales

The sales in this department is high during the four major holidays in the year, which is visible in four peaks in the Trend graph. During analysis and modeling it is noticed that

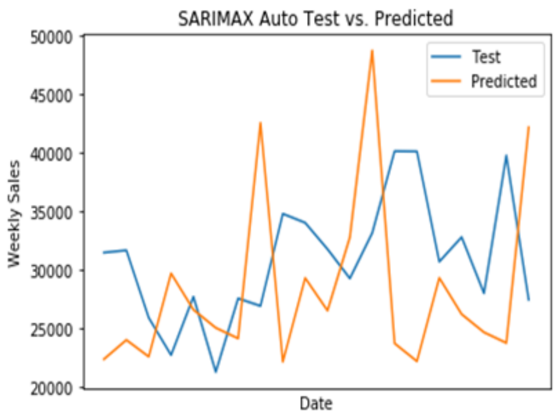
- Sales is auto correlated with lag of 1 and 2
- Temperature is negative correlated
- Markdown 3 is positively correlated after 4 weeks.



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	30500.3	1718.3	17.75	<.0001	0	Weekly_Sales	0
MA1,1	0.53274	0.18959	2.81	0.0050	1	Weekly_Sales	0
AR1,1	0.96392	0.16967	5.68	<.0001	1	Weekly_Sales	0
AR1,2	-0.48045	0.08229	-5.84	<.0001	2	Weekly_Sales	0
NUM1	-183.08655	28.14168	-6.51	<.0001	0	Temperature	0
NUM2	0.39572	0.11283	3.51	0.0005	0	MarkDown_3	4

5.3.2 Electronics Department Sales

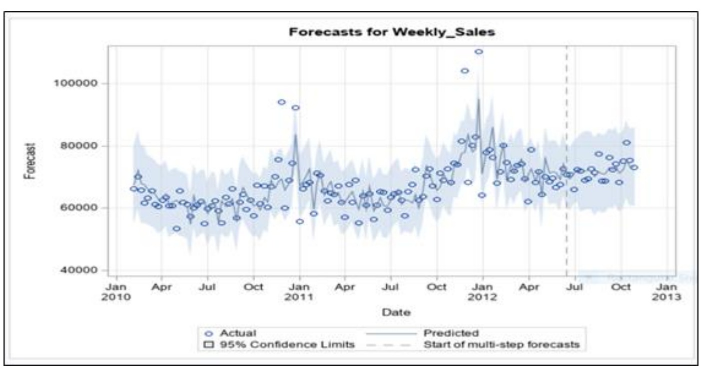
The sales in this department are very high during the end of each year (Christmas period). It is also confirmed in modeling, as we can see the “IsHoliday” variable is significant. It is observed that sales increased by approximately $ 27,500 more in that period for each week.




```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
IsHoliday_x 2.752e+04   4109.990      6.695      0.000    1.95e+04    3.56e+04
ar.L1         -0.8077      0.137     -5.880      0.000      -1.077      -0.538
ar.L2         -0.5918      0.142     -4.173      0.000      -0.870      -0.314
ar.L3         -0.2912      0.129     -2.263      0.024      -0.543      -0.039
ma.S.L12      -0.7835      0.221     -3.551      0.000      -1.216      -0.351
sigma2      5.372e+08      0.211   2.55e+09      0.000    5.37e+08    5.37e+08
===================================================================================
Ljung-Box (Q):                       37.28   Jarque-Bera (JB):               581.37
Prob(Q):                              0.59   Prob(JB):                         0.00
Heteroskedasticity (H):               0.96   Skew:                             1.93
Prob(H) (two-sided):                  0.90   Kurtosis:                        13.58
===================================================================================
```

5.3.3 Groceries Department Sales

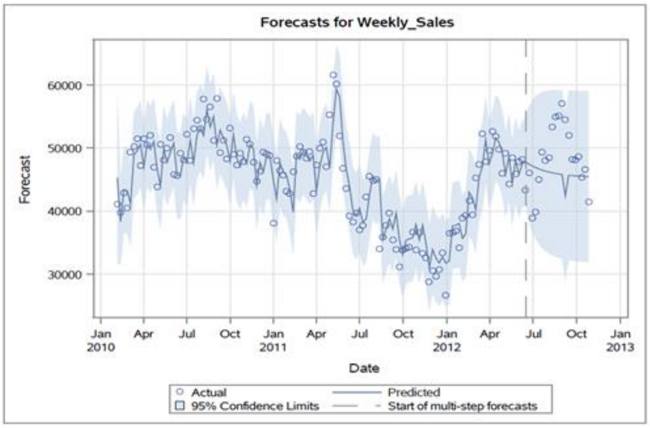
The sales in the Groceries department are seasonal with a slight positive trend. The “Fuel” and “Temperature” are negatively correlated (with immediate effect as lag is 0) to weeky sales.



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-318394.2	85896.8	-3.71	0.0002	0	Weekly_Sales	0
MA1,1	-0.71956	0.07040	-10.22	<.0001	4	Weekly_Sales	0
NUM1	-123.35405	41.51681	-2.97	0.0030	0	Temperature	0
NUM2	-8218.8	2985.7	-2.75	0.0059	0	Fuel_Price	0
NUM3	3254.1	726.57831	4.48	<.0001	0	CPI	0
NUM4	0.58362	0.19428	3.00	0.0027	0	MarkDown_5	0
NUM5	4502.8	1186.7	3.79	0.0001	0	IsHoliday_numeric	0

5.3.4 Gasoline Department Sales

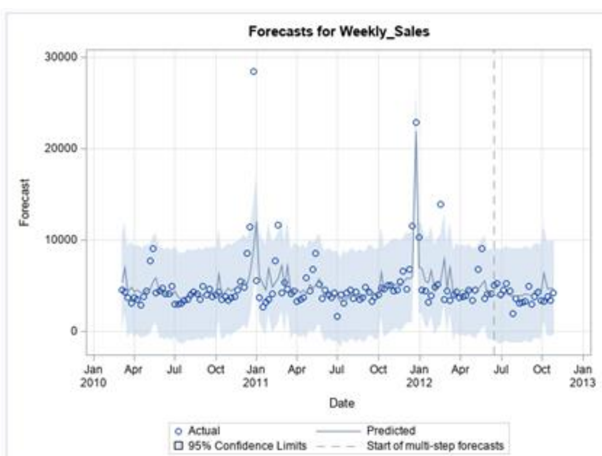
This Store34 is the only store where in Gasoline is sold among the dataset of 45 stores which were available to us. It is noticed that Gasoline sale is less during the Holiday period. It may be due to less travel during this period.



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	45316.6	2040.4	22.21	<.0001	0	Weekly_Sales	0
AR1,1	0.86292	0.04185	20.62	<.0001	1	Weekly_Sales	0
NUM1	-3320.2	835.19075	-3.98	<.0001	0	IsHoliday_numeric	0

5.3.5 Jewelry Department Sales

The sales in the Jewelry department is more during Holidays. Temperature is negatively correlated with sales while Markdown3 has positive correlation with impact showing after 4 weeks.



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	7066.4	1166.0	6.06	<.0001	0	Weekly_Sales	0
AR1,1	0.30164	0.08343	3.62	0.0003	1	Weekly_Sales	0
NUM1	-40.17376	18.90781	-2.12	0.0336	0	Temperature	0
NUM2	0.29521	0.06096	4.84	<.0001	0	MarkDown_3	4
NUM3	1938.4	818.53193	2.37	0.0179	0	IsHoliday_numeric	4

6.0 Business Finding and Recommendations

From the detailed Analysis of Time Series data and further modeling performed we found the following insights in the weekly sales of store 34:

(1) "Markdown" Variable:

- Markdown 3 is negatively impacting the Total Weekly Sales of Store 34.
- Different Markdowns have a positive impact on individual departments. This may be due to selective items being marked down. Department-wise, markdown 3 is a useful tool for boosting sales in the candy & tobacco department and the jewelry department, while markdown 5 is helpful in boosting sales in the grocery department.

(2) "IsHoliday" Variable:

- Is Positively correlated to the Weekly sales of Grocery, Jewelry, Electronics and Candy department.
- Is Negatively correlated to the Gasoline department. It shall be due to many people not travelling by road during holidays or less travel to office.

(3) "Temperature" Variable: Shows Negative correlation to all departments sales. It shall be indirectly related to the seasonality in the sales in that store. If we know the actual location of the store, we can derive more intelligent analysis..

(4) "Unemployment" Variable: have a negative effect on the Total Sales of the store.

(5) "CPI, Fuel Price" Variable: Have no effect on the total sales of store34.

Recommendation:

(1) Instead of indiscriminately applying markdown 3 to all departments of store 34, we recommend Walmart to specify "Markdown 3" to the candy & tobacco department and the jewelry department, in order to boost its weekly sales. Otherwise, the total weekly sales would be negatively affected by the undifferentiated third round of markdown after 3 weeks of its application.

(2) More intelligent analysis can be done when more information is available like:

 (a) Actual "Store Location Information"
 (b) Number of Persons visiting stores each week
 (c) Median income of neighborhood
 (d) Demographic of customers in the neighborhood.

7.0 References

1. **Data Set from Kaggle:**

Walmart Recruiting - Store Sales Forecasting. https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data. Accessed: August 01,2020.

features.csv. https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data?select=features.csv.zip. Accessed: August 01, 2020.

train.csv. https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data?select=train.csv.zip. Accessed: August 01, 2020.

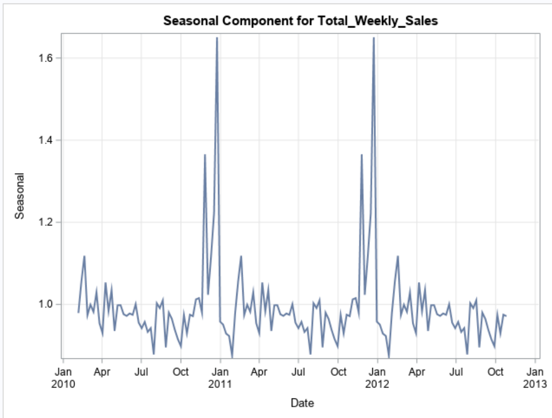
Walmart Department Number. https://blog.8thandwalton.com/wp-content/uploads/2017/10/Walmart-Department-Numbers.pdf?utm_campaign=Lead%20Generation&utm_medium=email&_hsenc=p2ANqtz--Ky_aq4ZWM2S4fzji0o7imvdBOjODgkZBAjcL42C6yefKHbXfrTIt5AOW3q_24qKIl2ckZvVkTs8p08gb6Wke_5VthXg&_hsmi=59683228&utm_content=59683228&utm_source=hs_automation&hsCtaTracking=cdf6bba2-a9f4-4034-8411-9d510d02dc99%7C04f3a820-62fb-4901-9bb1-3b3d0f8161ec. Accessed: August 01,2020.

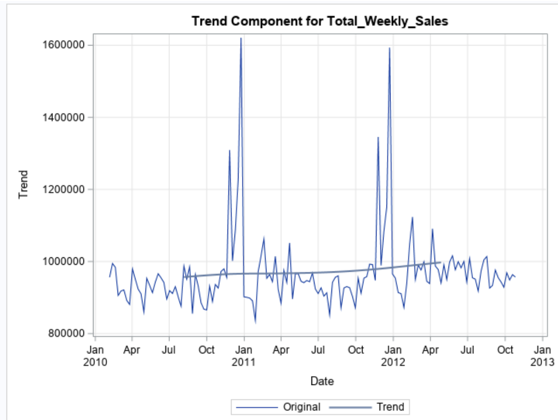
2. **Book:**

Fernandez, George, Marc Huber, Jay Laramore, Danny Modlin, Chip Wells. 2019. *Time Series Modeling Essentials Course Notes*. Cary, NC: SAS Institute Inc.

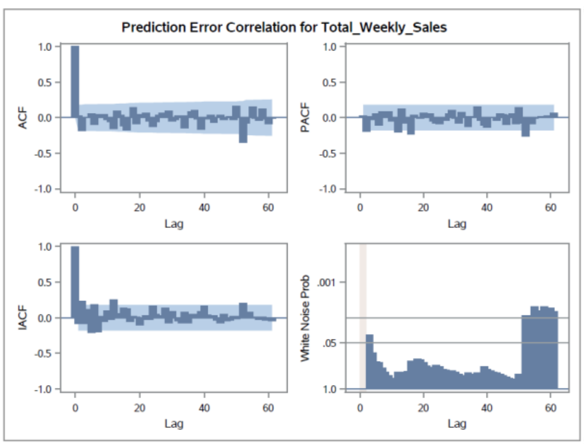
Appendix

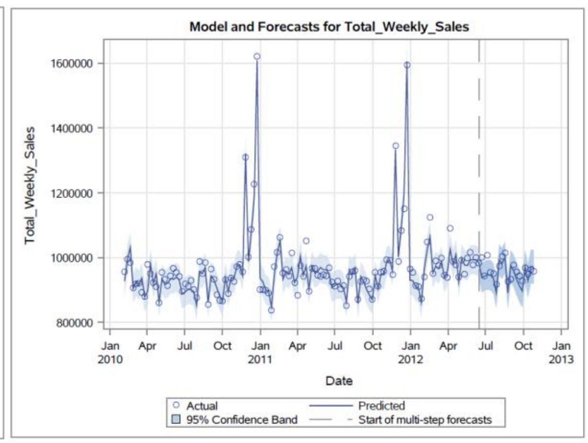
1. Decomposition analysis for the target variable - Total_Weekly_Sales



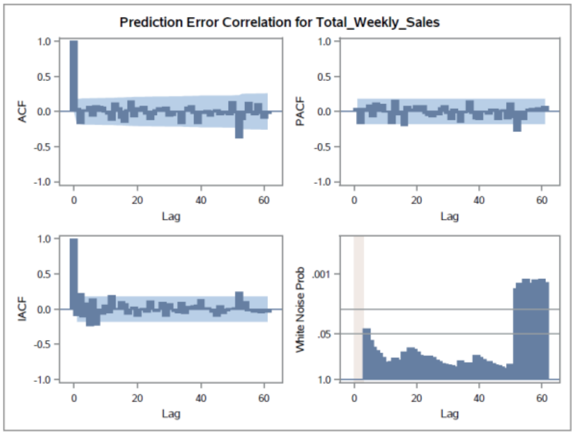


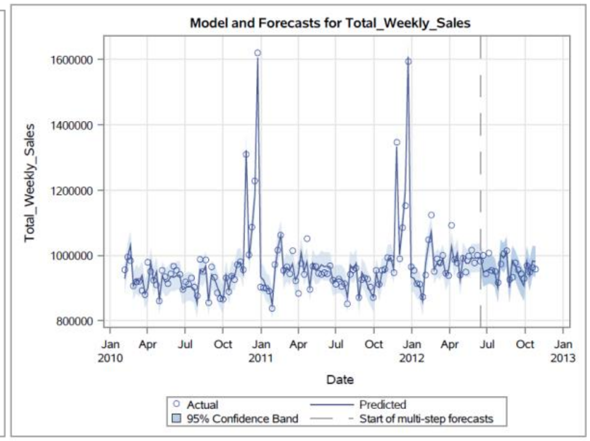
2. Additive seasonal model results





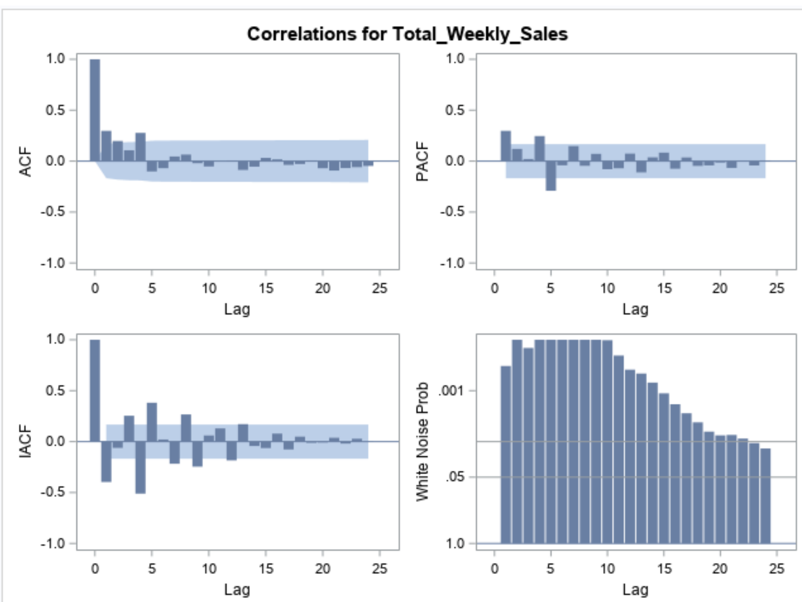
3. Additive winters model results



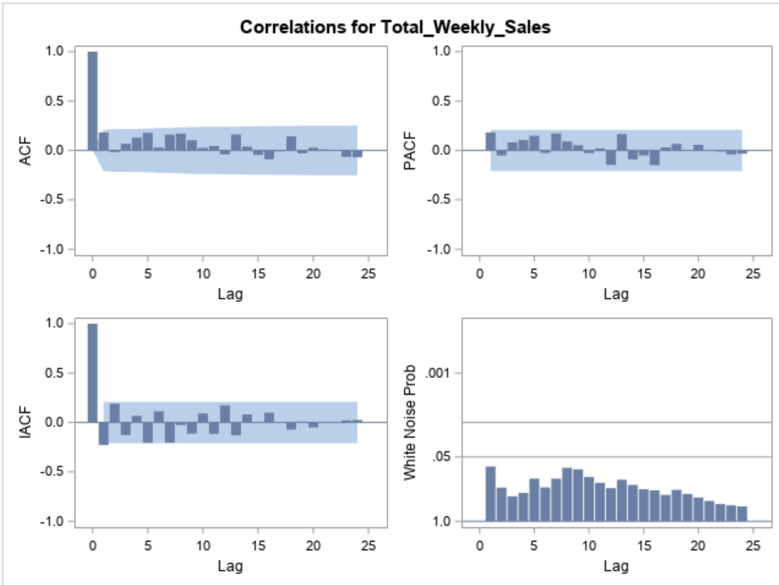


4. Autocorrelation check for the target variable Total_Weekly_Sales

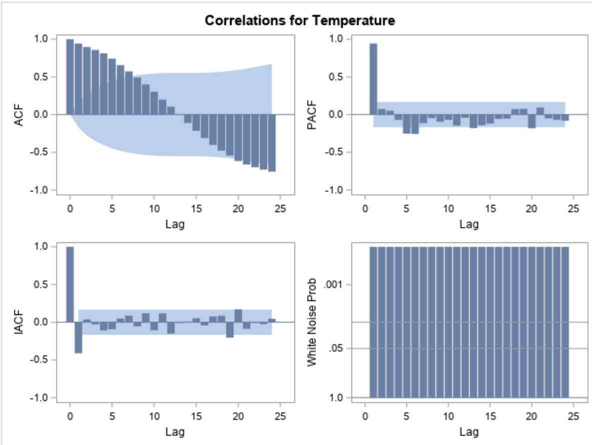
1) **Scenario1**: Based on the original time series

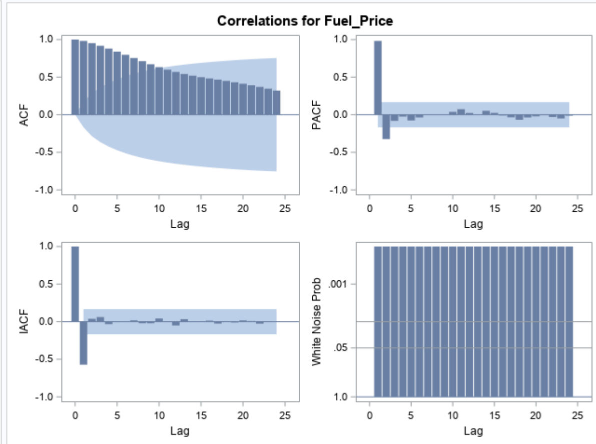


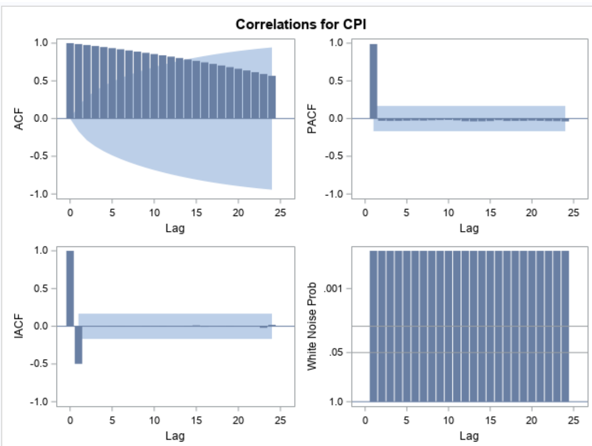
2) **Scenario2**: Based on the time series with first order seasonal differencing

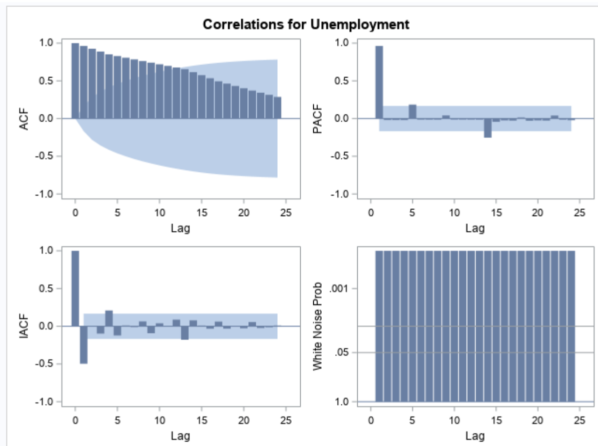


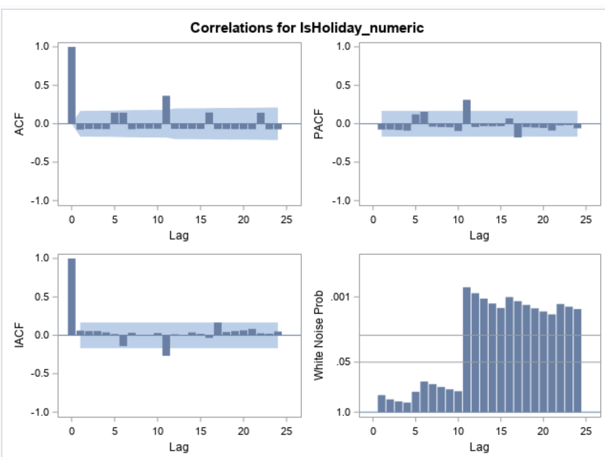
5. Autocorrelation check for all independent variables;

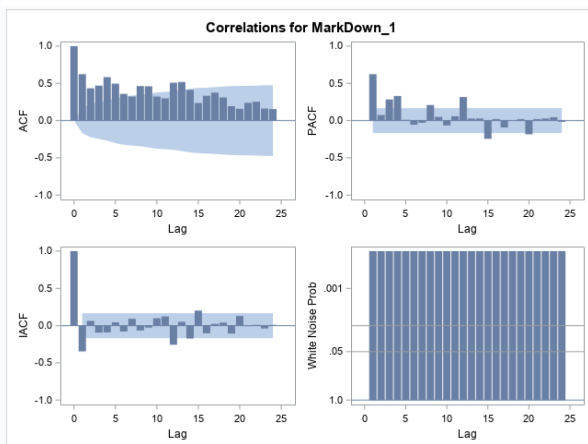


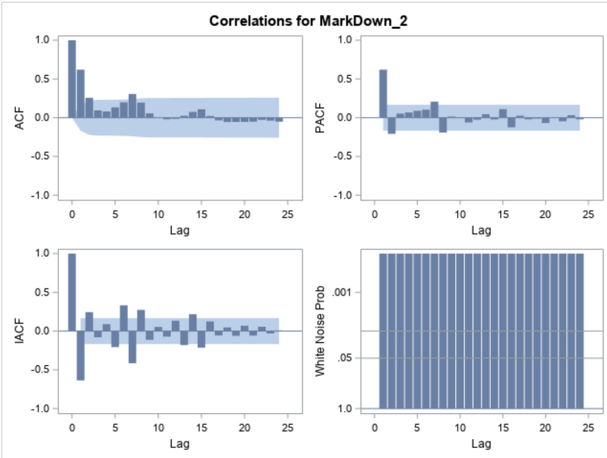


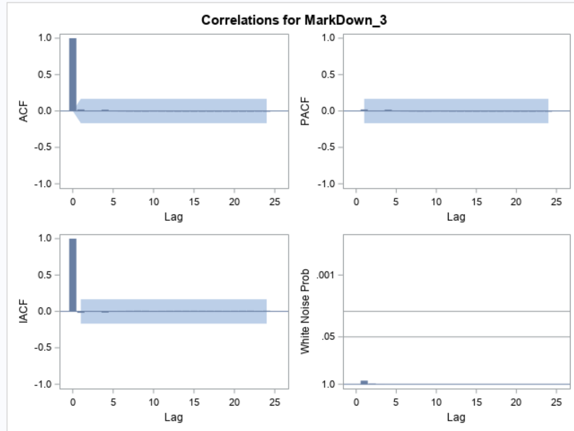


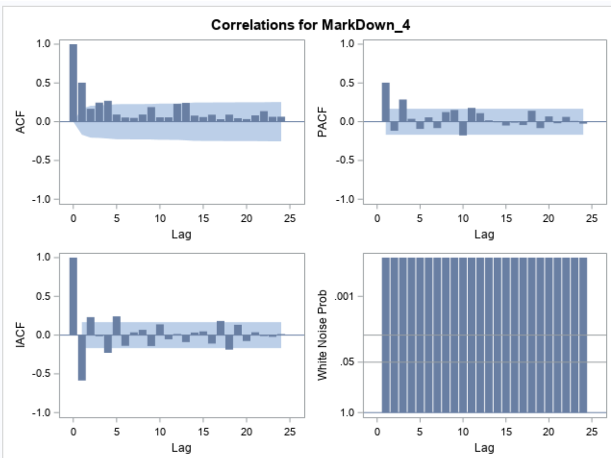


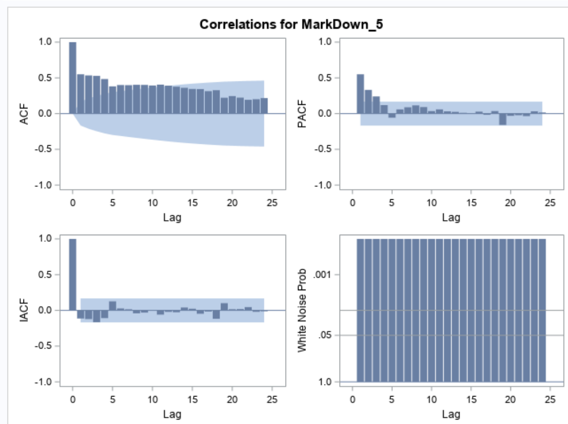





Correlations for MarkDown_2
ACF
PACF
IACF
White Noise Prob
Lag

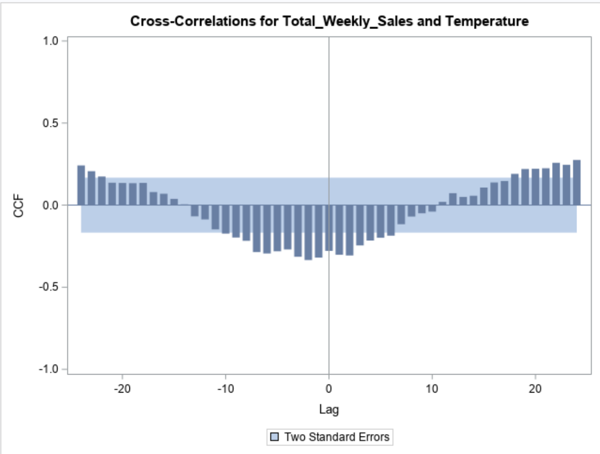

Correlations for MarkDown_3
ACF
PACF
IACF
White Noise Prob
Lag

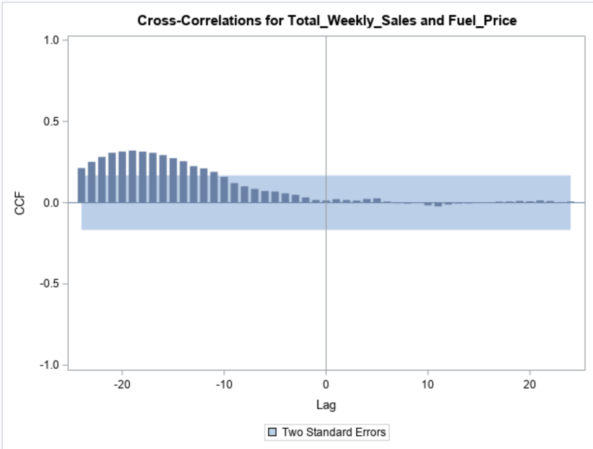

Correlations for MarkDown_4
ACF
PACF
IACF
White Noise Prob
Lag

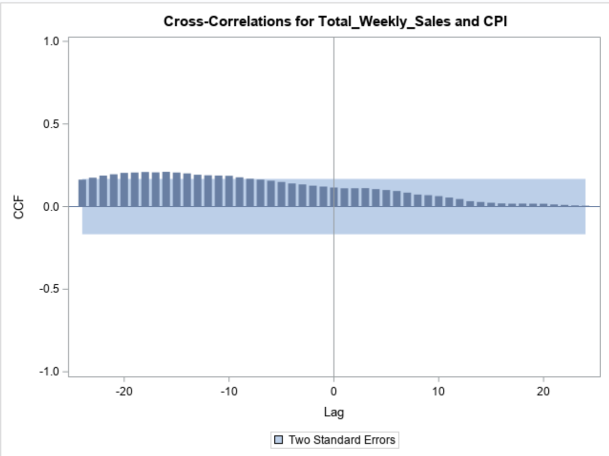

Correlations for MarkDown_5
ACF
PACF
IACF
White Noise Prob
Lag

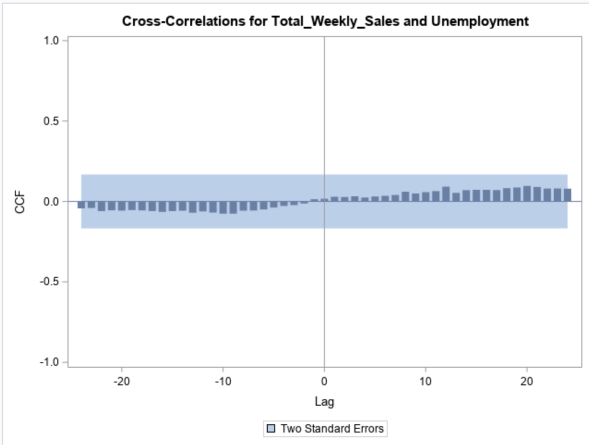
6. Cross-correlation check between all independent variables and target variable Total_Weekly_Sales

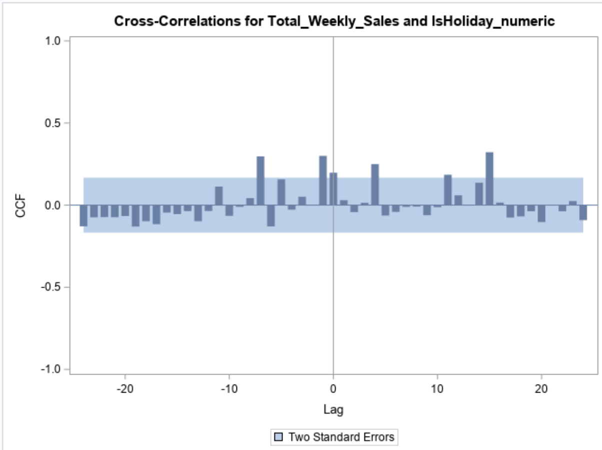
1) **Scenario1**: Based on the original time series

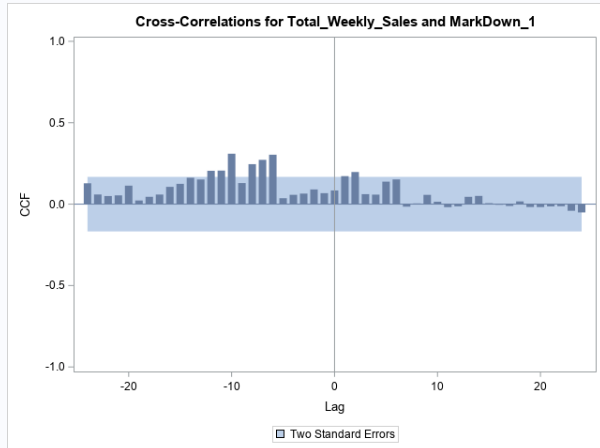


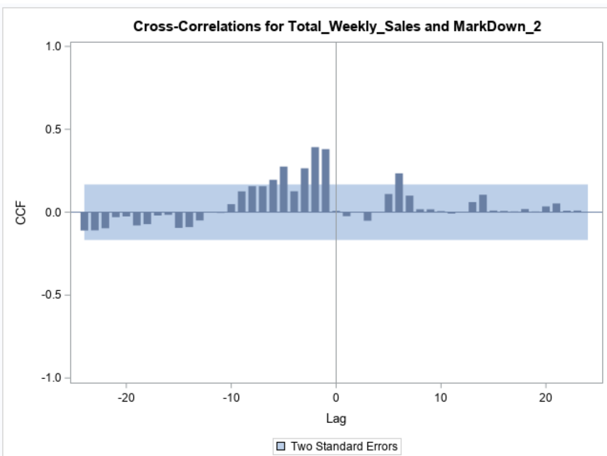


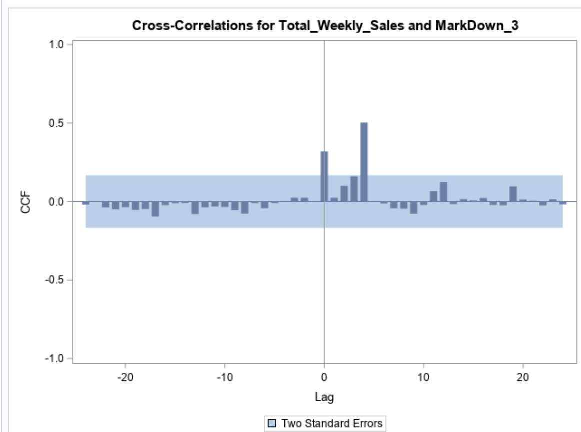


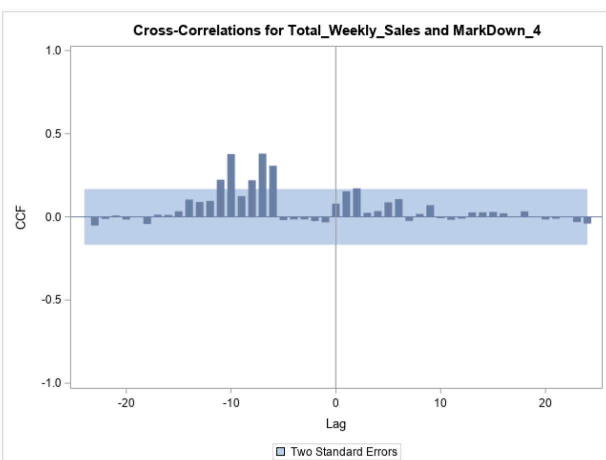


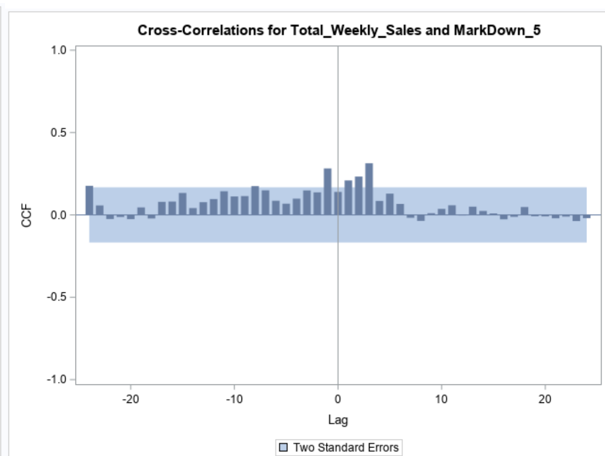

Cross-Correlations for Total_Weekly_Sales and IsHoliday_numeric
CCF
Lag
Two Standard Errors


Cross-Correlations for Total_Weekly_Sales and MarkDown_1
CCF
Lag
Two Standard Errors

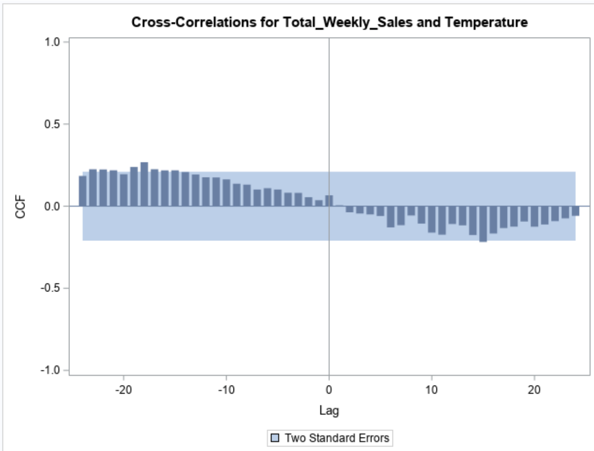

Cross-Correlations for Total_Weekly_Sales and MarkDown_2
CCF
Lag
Two Standard Errors

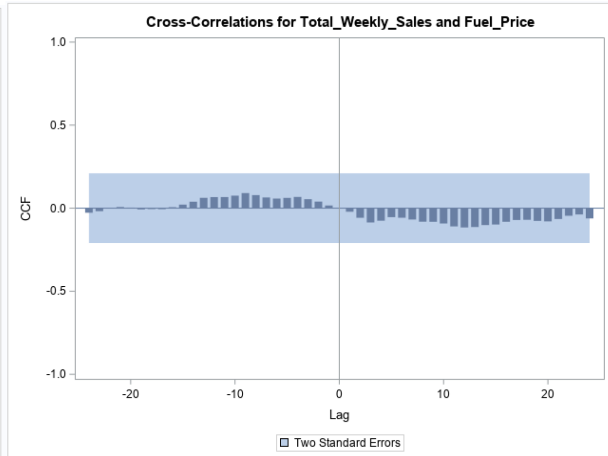

Cross-Correlations for Total_Weekly_Sales and MarkDown_3
CCF
Lag
Two Standard Errors

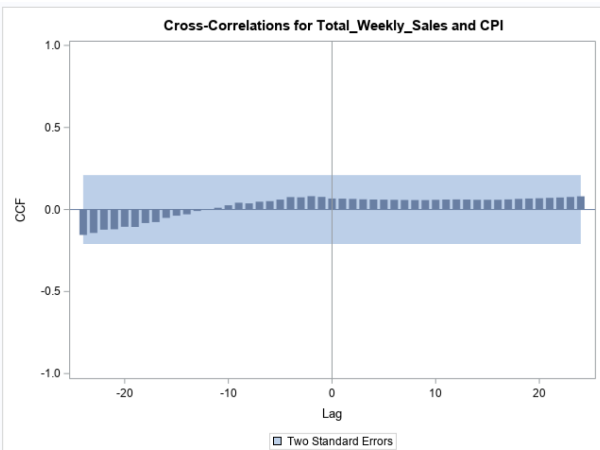

Cross-Correlations for Total_Weekly_Sales and MarkDown_4
CCF
Lag
Two Standard Errors

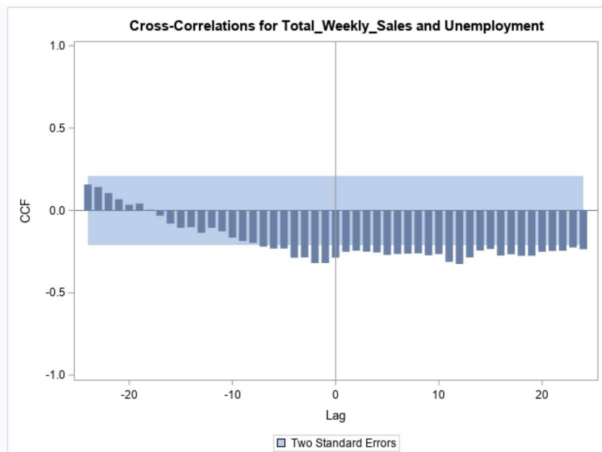

Cross-Correlations for Total_Weekly_Sales and MarkDown_5
CCF
Lag
Two Standard Errors

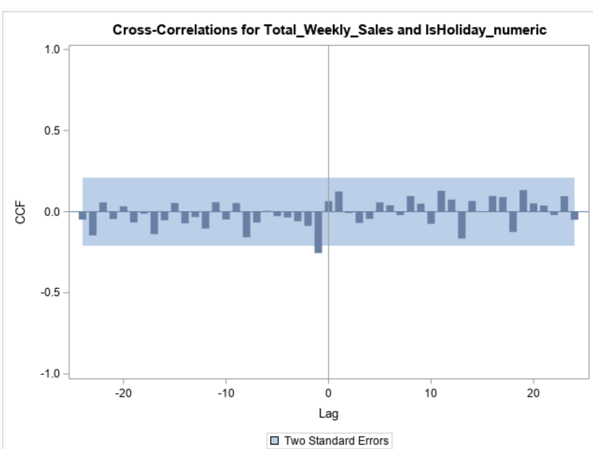
2) **Scenario2**: Based on the time series with first order seasonal differencing

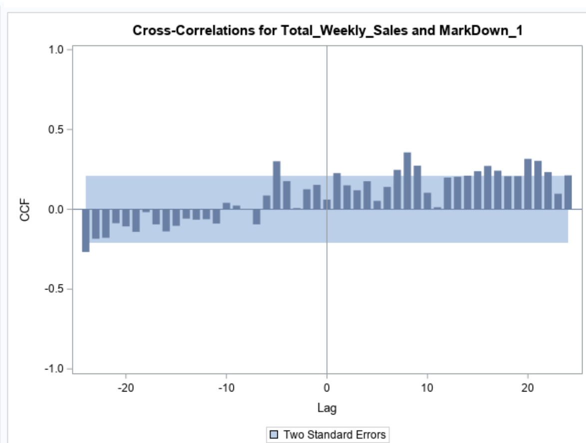


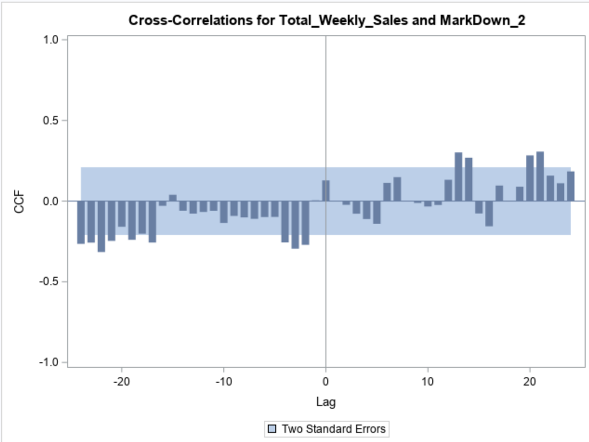


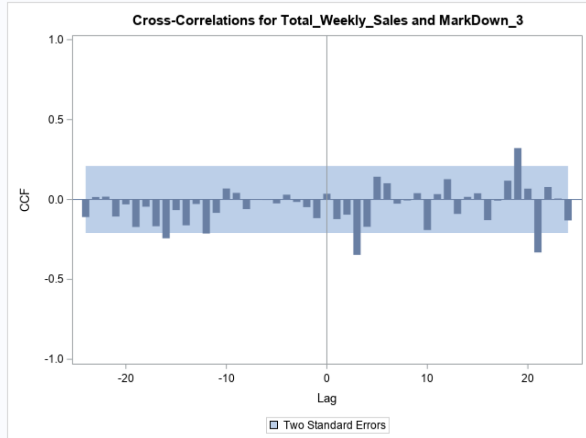


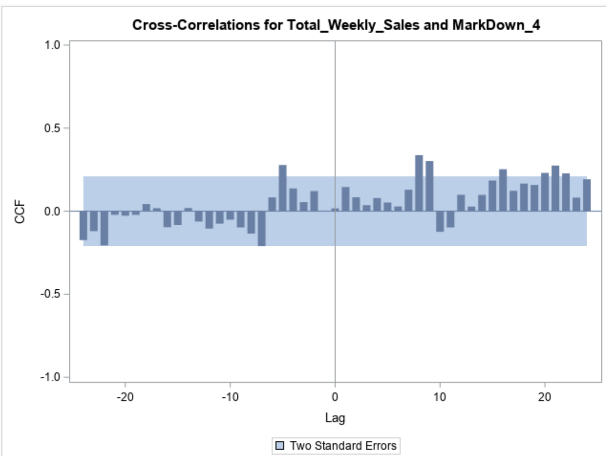


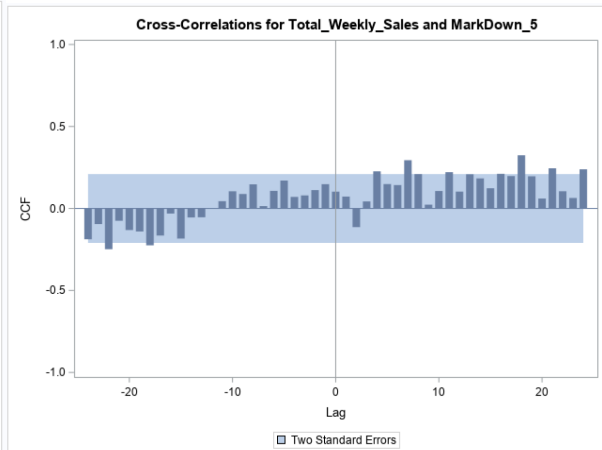












7. Estimates for Arimax Model 1

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	989618.3	59500.8	16.63	<.0001	0	Total_Weekly_Sales	0
MA1,1	-0.59426	0.20349	-2.92	0.0035	1	Total_Weekly_Sales	0
MA1,2	-0.39345	0.17152	-2.29	0.0218	2	Total_Weekly_Sales	0
MA1,3	-0.58382	0.19427	-3.01	0.0027	4	Total_Weekly_Sales	0
NUM1	2453.8	743.90964	3.30	0.0010	0	Temperature	0
NUM1,1	3325.1	750.89664	4.43	<.0001	1	Temperature	0
NUM2	85525.3	15420.5	5.55	<.0001	0	IsHoliday_numeric	4
NUM1,1	25821.7	13078.9	1.97	0.0483	11	IsHoliday_numeric	4
NUM3	11.48906	2.60629	4.41	<.0001	0	MarkDown_2	6
NUM4	6.27550	1.41729	4.43	<.0001	0	MarkDown_3	0
NUM5	12.96701	3.08155	4.21	<.0001	0	MarkDown_5	3

8. Estimates for Arimax Model 2

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	190441.0	68332.6	2.79	0.0053	0	Total_Weekly_Sales	0
NUM1	-16634.9	6841.6	-2.43	0.0150	0	Unemployment	0
NUM2	-2.32316	0.85140	-2.73	0.0064	0	MarkDown_3	3

9. Base ARMA model selection steps

p	q	AIC	SBC
1	1	3700	3709
1	2 (insignificant)	3702	3711
1	0	3701	3707
1	4	3686	3695
1(insignificant)	1,4 (q=1 insignificant)	3686	3698
0	4	3703	3709
1	2,4 (q=2 insignificant)	3687	3698
1	1,2,4	All variables are insignificant	
4	1	3691	3700
4	2	3699	3708
4	4	All variables are insignificant	
4 (insignificant)	1,4	All variables are insignificant	
0	1,4	3685	3694
4 (insignificant)	2,4 (q=4 insignificant)	3701	3713

0	2(insignificant)	3710	3716
0	0	3712	3715
4 (insignificant)	1,2,4	3678	3693
0	1,2,4	3677	3688
5 (insignificant)	1	3702	3711
0	1	3704	3710
5	2	All variables are insignificant	
5	4	3699	3708
5 (insignificant)	1,4	3685	3697
5	2,4	3694	3704
5 (insignificant)	1,2,4	3677	3688
1,4 (p=1 insignificant)	1 (insignificant)	3693	3703
1,4	2 (insignificant)	3693	3705
1,4	0	3693	3702
1,4	4	3685	3697
1,4	1,4 (q=1 insignificant)	3688	3670
1,4	2,4	3686	3701
1,4 (all insignificant)	1,2,4	3680	3698
1,5	1	3696	3708
1,5	2 (insignificant)	3698	3710

1,5	0	3697	3706
1,5	4	3681	3693
1,5	1,4	Do not converge	
1,5	2,4 (q=2 insignificant)	3681	3696
1,5 (all insignificant)	1,2,4 (q=1 insignificant)	3679	3697
0	2,4	3699	3708
4,5	1	3688	3700
4,5	2	3695	3707
4,5 (p=4 insignificant)	4 (insignificant)	3701	3713
5 (insignificant)	0	3713	3718
4,5 (insignificant)	1,4	3687	3702
4,5 (p=4 insignificant)	2,4 (q=4 insignificant)	3696	3711
4,5 (insignificant)	1,2,4	3679	3697
1,4,5	1 (q=1 insignificant)	3684	3699
1,4,5	0	3683	3694
1,4,5	2(q=2 insignificant)	3682	3697
1,4,5(p=4 insignificant)	4(q=4 insignificant)	3683	3698
1,4,5 (all insignificant)	1,4 (q=1 insignificant)	3685	3703
1,4,5 (p=4 insignificant)	2,4 (all insigificant)	3683	3701

1,5	0	3697	3706
1,4,5 (all insignificant)	1,2,4	3681	3702

10. Estimates for the department-wise Arimax Models

1) Grocery

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-318394.2	85896.8	-3.71	0.0002	0	Weekly_Sales	0
MA1,1	-0.71956	0.07040	-10.22	<.0001	4	Weekly_Sales	0
NUM1	-123.35405	41.51681	-2.97	0.0030	0	Temperature	0
NUM2	-8218.8	2985.7	-2.75	0.0059	0	Fuel_Price	0
NUM3	3254.1	726.57831	4.48	<.0001	0	CPI	0
NUM4	0.58362	0.19428	3.00	0.0027	0	MarkDown_5	0
NUM5	4502.8	1186.7	3.79	0.0001	0	IsHoliday_numeric	0

2) Gasoline

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	45316.6	2040.4	22.21	<.0001	0	Weekly_Sales	0
AR1,1	0.86292	0.04185	20.62	<.0001	1	Weekly_Sales	0
NUM1	-3320.2	835.19075	-3.98	<.0001	0	IsHoliday_numeric	0

3) Candy & Tobacco

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	30500.3	1718.3	17.75	<.0001	0	Weekly_Sales	0
MA1,1	0.53274	0.18959	2.81	0.0050	1	Weekly_Sales	0
AR1,1	0.96392	0.16967	5.68	<.0001	1	Weekly_Sales	0
AR1,2	-0.48045	0.08229	-5.84	<.0001	2	Weekly_Sales	0
NUM1	-183.08655	28.14168	-6.51	<.0001	0	Temperature	0
NUM2	0.39572	0.11283	3.51	0.0005	0	MarkDown_3	4

4) Jewelry

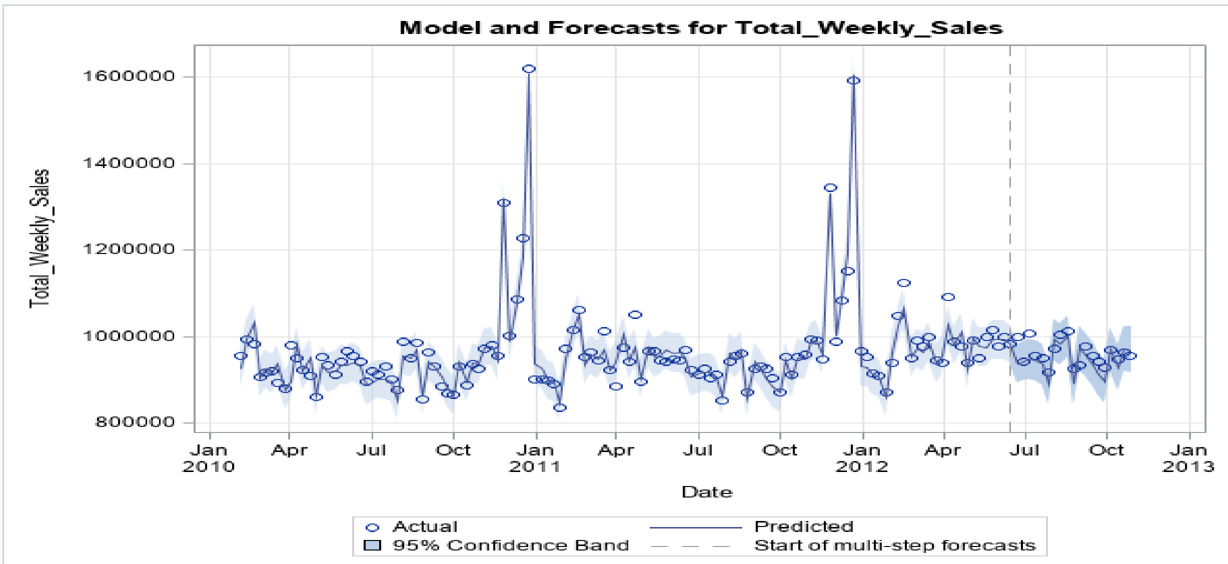
Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	7066.4	1166.0	6.06	<.0001	0	Weekly_Sales	0
AR1,1	0.30164	0.08343	3.62	0.0003	1	Weekly_Sales	0
NUM1	-40.17376	18.90781	-2.12	0.0336	0	Temperature	0
NUM2	0.29521	0.06096	4.84	<.0001	0	MarkDown_3	4
NUM3	1938.4	818.53193	2.37	0.0179	0	IsHoliday_numeric	4

5) Electronics

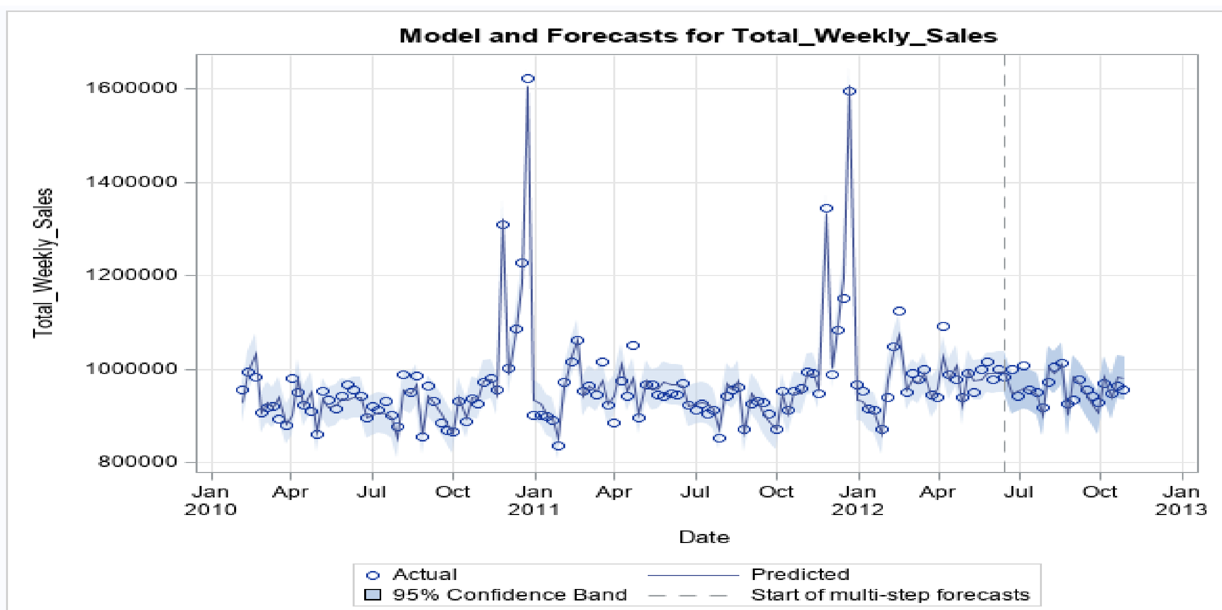
```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
IsHoliday_x  2.752e+04   4109.990      6.695      0.000    1.95e+04    3.56e+04
ar.L1          -0.8077      0.137     -5.880      0.000      -1.077      -0.538
ar.L2          -0.5918      0.142     -4.173      0.000      -0.870      -0.314
ar.L3          -0.2912      0.129     -2.263      0.024      -0.543      -0.039
ma.S.L12       -0.7835      0.221     -3.551      0.000      -1.216      -0.351
sigma2       5.372e+08      0.211   2.55e+09      0.000    5.37e+08    5.37e+08
==================================================================================
Ljung-Box (Q):                       37.28   Jarque-Bera (JB):               581.37
Prob(Q):                              0.59   Prob(JB):                         0.00
Heteroskedasticity (H):               0.96   Skew:                             1.93
Prob(H) (two-sided):                  0.90   Kurtosis:                        13.58
==================================================================================
```

11. Forecast plots for the 4 models of total weekly sales in store 34

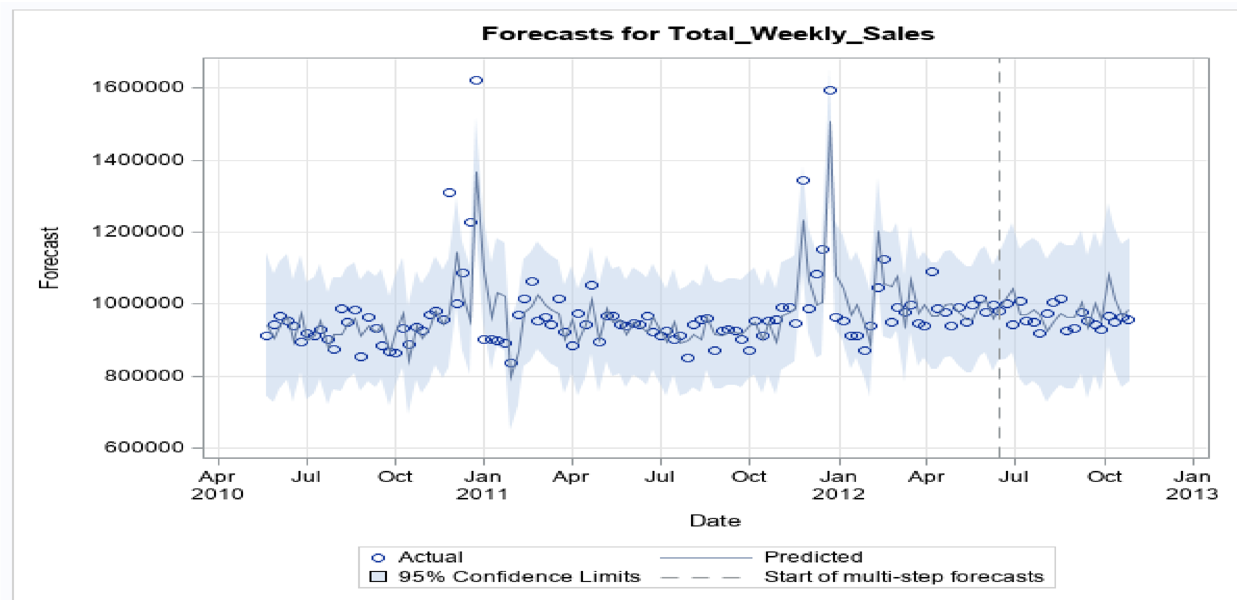
Additive Seasonal:



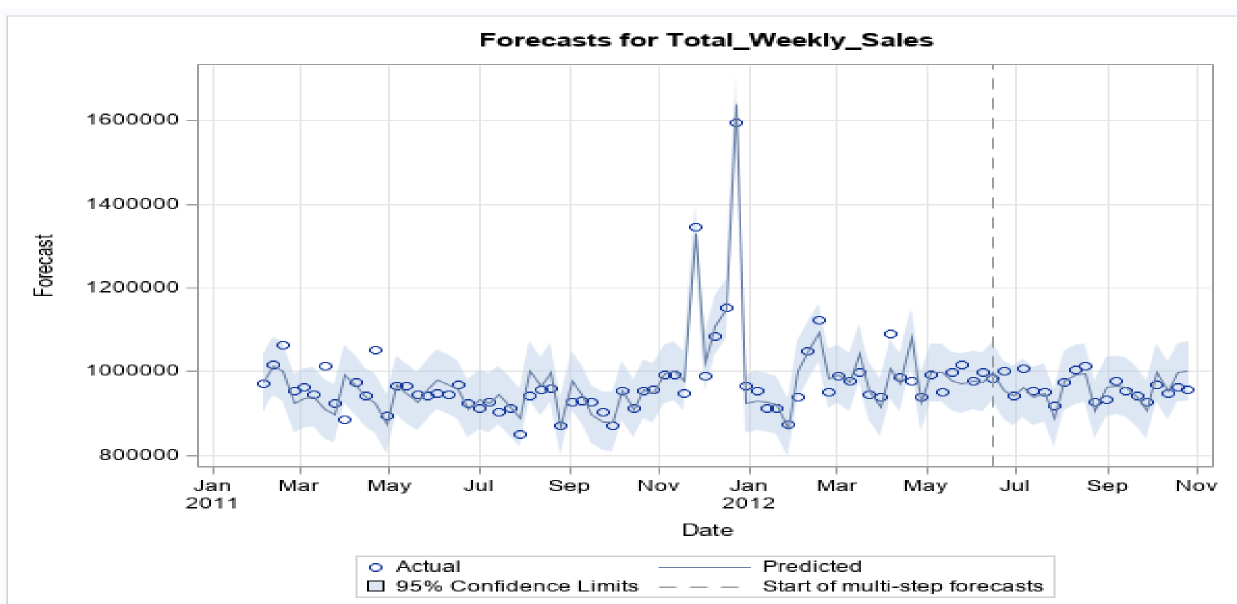
Additive Winters:



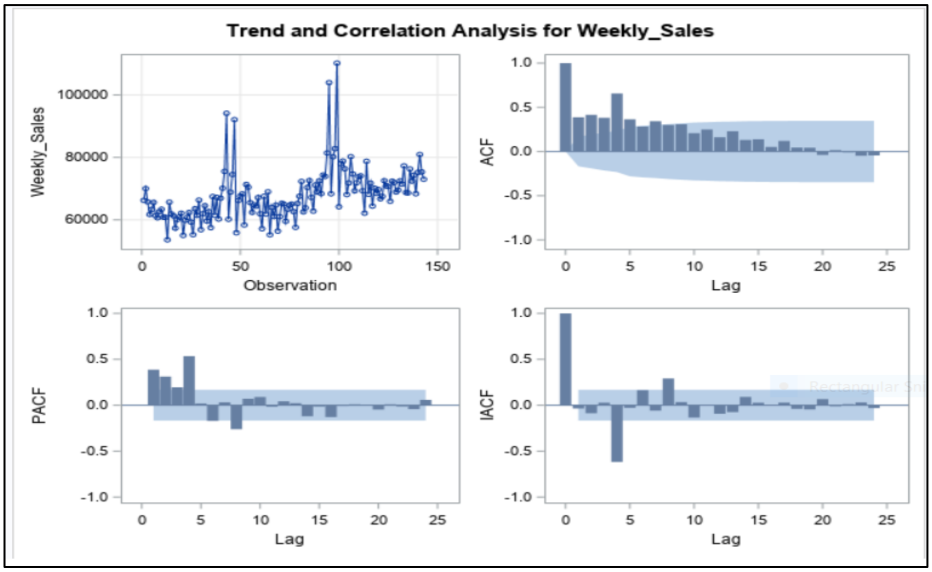
Arimax 1:



Arimax 2:



12. Grocery Department



Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	-318394.2	85896.8	-3.71	0.0002	0	Weekly_Sales	0
MA1,1	-0.71956	0.07040	-10.22	<.0001	4	Weekly_Sales	0
NUM1	-123.35405	41.51681	-2.97	0.0030	0	Temperature	0
NUM2	-8218.8	2985.7	-2.75	0.0059	0	Fuel_Price	0
NUM3	3254.1	726.57831	4.48	<.0001	0	CPI	0
NUM4	0.58362	0.19428	3.00	0.0027	0	MarkDown_5	0
NUM5	4502.8	1186.7	3.79	0.0001	0	IsHoliday_numeric	0

```
proc arima data=Work.preProcessedData plots
     (only)=(series(acf corr crosscorr) residual(corr normal wn)
        forecast(forecast forecastonly) ) out=work.out;
    identify var=Weekly_Sales crosscorr=(Temperature Fuel_Price CPI MarkDown_1
    MarkDown_2 MarkDown_3 MarkDown_4 MarkDown_5 IsHoliday_numeric)
        outcov=work.outcov;
    estimate q=(4) input=(Temperature Fuel_Price CPI
    MarkDown_5 IsHoliday_numeric) method=ML
        outest=work.outest outstat=work.outstat;
    forecast lead=20 back=20 alpha=0.05 id=Date interval=week.6 printall;
    run;
```