Final Paper

RSM2129 Forecasting Models and Econometric Methods

Course Instructor: Prof. Peter Dungan

Submitted by: Bhanvi Gupta

Topic: Retail sales at a supermart

UNIVERSITY OF TORONTO

The dataset I am analysing was published by Walmart for a prediction competition on Kaggle, a data-science platform. It contains information on weekly retail sales for 45 stores along with other attributes like department of sale, whether there was a holiday or not, fuel price, temperature, unemployment rate, consumer price index (CPI) and markdown on items. Since markdown data is very limited (approximately 50 entries), we will not be studying it's impact on weekly sales. The data was available in the form of 3 different datasets namely stores.csv, test.csv, features.csv. For the purpose of this report, I created another dataset with data belonging to 3 departments of store 1. In total, we have 143 entries of data for analysis.

Here's a gist of variables we will be considering in this report:

1. Weekly_sales: Weekly sales at a store 1 and department 1 in dollar terms.
   I will be dividing it by CPI to get it to real terms while estimating.
2. Weekly_sales01: Weekly sales at store 1 and department 2 in dollar terms.
   I will be dividing it by CPI to get it to real terms while estimating.
3. Weekly_sales02: Weekly sales at store 1 and department 3 in dollar terms.
   I will be dividing it by CPI to get it to real terms while estimating.
4. CPI: Consumer price index
5. IsHoliday - whether the week is a special holiday week. It has data in binary form which will be used like a dummy variable in the analysis
6. Dept - the department number of a store
7. Date – reflects the week
8. Fuel_Price - cost of fuel in the region in dollar terms. I will be dividing it by CPI to get it to real terms while estimating
9. Unemployment - the unemployment rate

The following dates have been configured as true in the IsHoliday field:

Super Bowl:	12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
Labor Day:	10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13
Thanksgiving:	26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13
Christmas:	31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

The data was divided into 3 groups for analysis based on the department number of store 1. To begin the analysis, some exploratory data analysis was performed group-wise.

Group 1: Weekly sales for department 1 with 3 explanatory variables: fuel price, unemployment rate, cpi and temperature

Group 2: Weekly sales for department 2 with 3 explanatory variables: fuel price, unemployment rate, cpi and temperature

Group 3: Weekly sales for department 3 with 3 explanatory variables: fuel price, unemployment rate, cpi and temperature

In this report, we will mostly be talking about the sales in department 1. But some analysis has also been performed on departments 2 and 3 for comparison.

Descriptive Statistics
Department 1:

Date: 04/20/23 Time: 14:31
Sample: 2/05/2010 10/26/2012

	CPI	FUEL_PRICE	TEMPERATURE	UNEMPLOYMENT	WEEKLY_SALES
Mean	215.9969	3.219699	68.30678	7.610420	22513.32
Median	215.4599	3.290000	69.64000	7.787000	18535.48
Maximum	223.4443	3.907000	91.65000	8.106000	57592.12
Minimum	210.3374	2.514000	35.40000	6.573000	14537.37
Std. Dev.	4.350890	0.427313	14.25049	0.383749	9854.349
Skewness	0.265983	-0.151813	-0.402643	-1.046172	1.992864
Kurtosis	1.547967	1.600071	2.114668	3.043197	6.202758
Jarque-Bera	14.24867	12.22644	8.534116	26.09612	155.7728
Probability	0.000805	0.002213	0.014023	0.000002	0.000000
Sum	30887.56	460.4170	9767.870	1088.290	3219405.
Sum Sq. Dev.	2688.095	25.92863	28836.84	20.91137	1.38E+10
Observations	143	143	143	143	143

Some key observations include mean and standard deviation of weekly sales for department 1, values being $22513.32 and $9854.34 respectively.

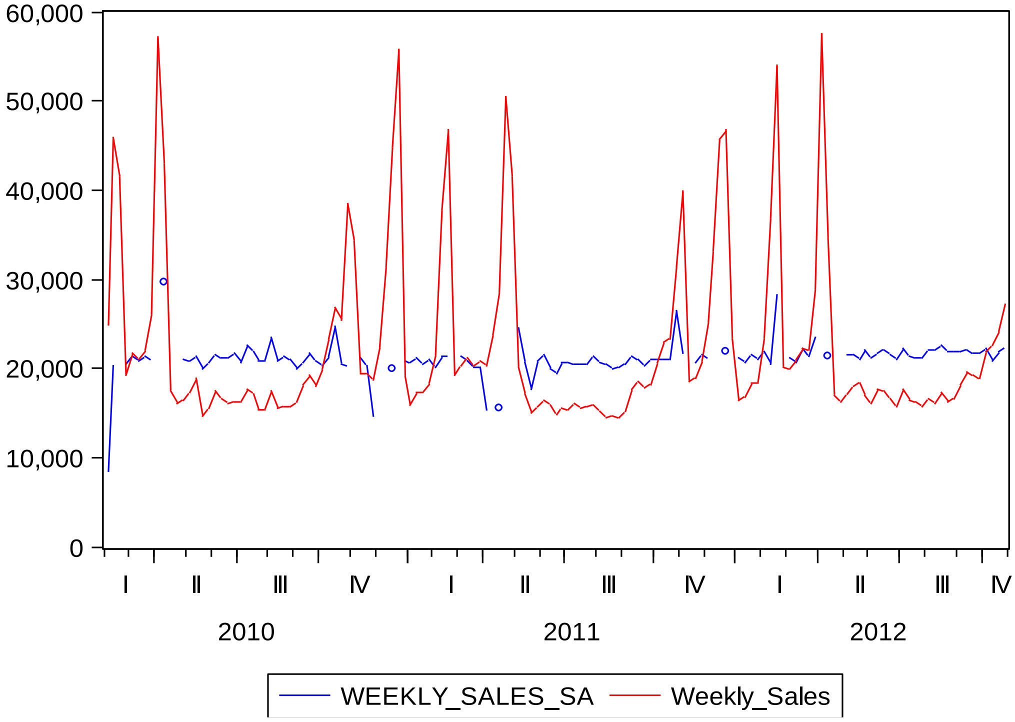
Similarly for group 2 and 3 which have descriptive statistics for department 2 and 3 respectively can be referred in Figure 1 and 2 from the appendix. We can see that the mean weekly sales for department 2 are $46102 and standard deviation is $35819 and mean weekly sales for department 3 are $13150 and standard deviation is $8709. From these stats, it's clear that department 2 has the highest number of sales.

Correlation Matrix:
To rule out multicollinearity, a correlation plot for the relevant variables was plotted From the plot, we can see fuel price, unemployment are moderately correlated to cpi. Since it's not high, we can conclude that there is no multicollinearity.

Covariance Analysis: Ordinary
Date: 04/20/23 Time: 14:35
Sample: 2/05/2010 10/26/2012
Included observations: 143

Correlation t-Statistic Probability	CPI	FUEL_PRICE	TEMPERATURE	UNEMPLOYMENT	WEEKLY_SALES
CPI	1.000000 ----- -----				
FUEL_PRICE	0.755259 13.68299 0.0000	1.000000 ----- -----			
TEMPERATURE	0.118503 1.417135 0.1586	0.228493 2.786927 0.0061	1.000000 ----- -----		
UNEMPLOYMENT	-0.813471 -16.60820 0.0000	-0.513944 -7.114221 0.0000	-0.180695 -2.181544 0.0308	1.000000 ----- -----	
WEEKLY_SALES	-0.043533 -0.517413 0.6057	-0.066396 -0.790150 0.4308	-0.473860 -6.389703 0.0000	0.103399 1.234415 0.2191	1.000000 ----- -----

When we visualized data using weekly_sales, we could see a lot of seasonality from the repeated peaks. These seasonal factors probably correspond to the holidays mentioned in the dataset.



To resolve the seasonality, I invoked the seasonal adjustment method in eViews and got a graph called Weekly_sales_sa that was not so favourable for analysis. It has structural breaks which will cause mushy coefficients in the equation. Here's a linear regression I tried using 2 variables, temperature and fuel price in real terms.
Since the weekly retail sales and fuel prices are nominal variables (in dollar terms), we will be dividing them by CPI to convert them to real terms in our equations.

Dependent Variable: WEEKLY_SALES/CPI
Method: Least Squares
Date: 04/20/23 Time: 14:39
Sample: 2/05/2010 10/26/2012
Included observations: 143

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	127.7581	95.12467	1.343060	0.1814
TEMPERATURE	-1.531123	0.247247	-6.192676	0.0000
FUEL_PRICE/CPI	1120.348	2167.979	0.516771	0.6061
UNEMPLOYMENT	8.470704	9.905351	0.855164	0.3939

R-squared	0.232624	Mean dependent var	104.3109
Adjusted R-squared	0.216061	S.D. dependent var	45.90790
S.E. of regression	40.64699	Akaike info criterion	10.27530
Sum squared resid	229652.7	Schwarz criterion	10.35818
Log likelihood	-730.6839	Hannan-Quinn criter.	10.30898
F-statistic	14.04555	Durbin-Watson stat	1.170184
Prob(F-statistic)	0.000000		

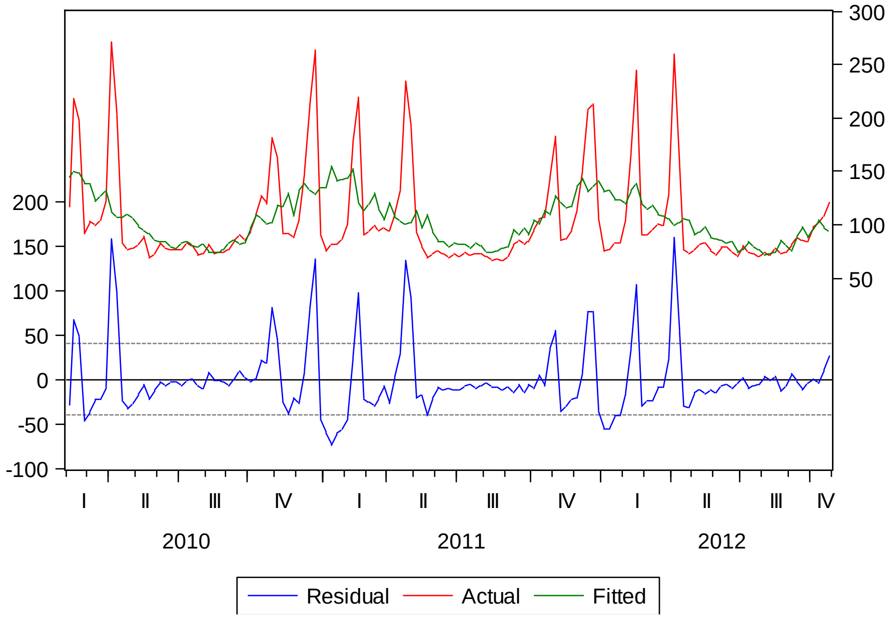
From the results, we can see that only temperature is statistically significant. The coefficient is -1.53, which means that with one unit increase in temperature, there is a 1.53 unit decrease in retail sales for department 1 at store 1.

To resolve this issue, I decided to use the isHoliday variable from my dataset which has binary data and has value 1 in places of expected holidays. So, I didn't have to create another dummy variable for these seasonal changes.

Dependent Variable: WEEKLY SALES/CPI
Method: Least Squares
Date: 04/22/23 Time: 18:47
Sample: 2/05/2010 10/26/2012
Included observations: 143

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	127.9585	95.50097	1.339866	0.1825
UNEMPLOYMENT	8.495972	9.946205	0.854192	0.3945
TEMPERATURE	-1.534708	0.252336	-6.082001	0.0000
FUEL PRICE/CPI	1115.445	2176.675	0.512454	0.6092
ISHOLIDAY	-1.069377	13.67520	-0.078198	0.9378

R-squared	0.232658	Mean dependent var	104.3109
Adjusted R-squared	0.210416	S.D. dependent var	45.90790
S.E. of regression	40.79309	Akaike info criterion	10.28924
Sum squared resid	229642.5	Schwarz criterion	10.39284
Log likelihood	-730.6808	Hannan-Quinn criter.	10.33134
F-statistic	10.46037	Durbin-Watson stat	1.169100
Prob(F-statistic)	0.000000		



From these results, we can see that all variables except temperature still remain statistically insignificant. Also, the fitted graph thus generated looks much better than the seasonally adjusted graph generated by eViews.

After performing a similar activity for department 2 and 3, we can see somewhat similar results. Temperature is statistically significant for department 3 but none of the variables are statistically significant for department 2. (Figure 4 and 3)

After that, to rule out more specification errors, I tried using lagged weekly_sales in the equations for different departments to see if any other variable becomes statistically significant.

Dependent Variable: WEEKLY_SALES/CPI
Method: Least Squares
Date: 04/21/23 Time: 13:54
Sample (adjusted): 2/12/2010 10/26/2012
Included observations: 142 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	88.96883	86.58130	1.027576	0.3060
TEMPERATURE	-0.917659	0.256046	-3.583959	0.0005
WEEKLY_SALES(-1)/CPI(-1)	0.433629	0.076136	5.695450	0.0000
FUEL_PRICE/CPI	277.0495	1975.836	0.140219	0.8887
ISHOLIDAY	-1.903398	12.38076	-0.153738	0.8780
UNEMPLOYMENT	3.802418	9.042311	0.420514	0.6748

R-squared	0.382164	Mean dependent var	104.2140
Adjusted R-squared	0.359449	S.D. dependent var	46.05572
S.E. of regression	36.86043	Akaike info criterion	10.09349
Sum squared resid	184782.0	Schwarz criterion	10.21838
Log likelihood	-710.6377	Hannan-Quinn criter.	10.14424
F-statistic	16.82461	Durbin-Watson stat	1.624259
Prob(F-statistic)	0.000000		

For department 1, we can see that the lagged variable is statistically significant along with the previously significant variable, temperature.

But, when the same regression was used for department 2 and 3, isHoliday also became statistically significant. (Figure 6 and 7). This means there is a lagged behaviour present when it comes to retail sales and holidays by 1 time period for these departments. Now that we know that temperature and isholiday are the only variables with a significant effect along with the lagged variables, we will further dive into the ARIMA methods to implement time- series methods.

ARIMA MODELLING

From our previous findings, we can see a lagged effect of weekly sales present on actual sales. To make a training dataset, I have made another dataset with 100 entries from 2010 and 2011 and reserved 43 entries from 2012 for an ex-post forecast.
But to proceed with these methods, we will first check if the series is stationary.

From the correlogram on the left below for weekly sales of department 1, we can see that initially, the values decline quite quickly. But the observations from 19 to 33 do not decline as quickly as before, which is a sign of non- stationarity.

Date: 04/23/23 Time: 12:12
Sample: 2/05/2010 12/30/2011
Included observations: 100

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.582	0.582	34.846	0.000
		2	0.142	-0.297	36.940	0.000
		3	-0.002	0.105	36.941	0.000
		4	-0.052	-0.096	37.227	0.000
		5	-0.018	0.090	37.260	0.000
		6	0.134	0.168	39.215	0.000
		7	0.317	0.215	50.271	0.000
		8	0.273	-0.077	58.515	0.000
		9	0.137	0.055	60.612	0.000
		10	0.066	0.017	61.099	0.000
		11	-0.066	-0.148	61.597	0.000
		12	-0.143	-0.017	63.966	0.000
		13	-0.166	-0.181	67.192	0.000
		14	-0.159	-0.119	70.179	0.000
		15	-0.126	-0.073	72.076	0.000
		16	-0.129	-0.133	74.096	0.000
		17	-0.070	0.044	74.702	0.000
		18	-0.041	-0.010	74.909	0.000
		19	-0.126	-0.104	76.897	0.000
		20	-0.191	0.011	81.545	0.000
		21	-0.234	-0.061	88.611	0.000
		22	-0.238	-0.034	96.013	0.000
		23	-0.218	-0.025	102.29	0.000
		24	-0.163	-0.050	105.84	0.000
		25	-0.125	-0.111	107.97	0.000
		26	-0.124	-0.002	110.10	0.000
		27	-0.091	-0.024	111.26	0.000
		28	-0.101	-0.078	112.70	0.000
		29	-0.138	-0.023	115.45	0.000
		30	-0.108	-0.002	117.16	0.000
		31	-0.127	-0.136	119.55	0.000
		32	-0.136	-0.009	122.31	0.000
		33	-0.093	-0.055	123.62	0.000
		34	-0.025	-0.021	123.72	0.000
		35	-0.015	-0.102	123.76	0.000
		36	0.047	0.118	124.10	0.000

Date: 04/23/23 Time: 12:16
Sample (adjusted): 2/12/2010 12/30/2011
Included observations: 99 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.020	0.020	0.0394	0.843
		2	-0.352	-0.353	12.818	0.002
		3	-0.108	-0.105	14.030	0.003
		4	-0.100	-0.257	15.086	0.005
		5	-0.140	-0.277	17.182	0.004
		6	-0.038	-0.286	17.334	0.008
		7	0.272	0.018	25.358	0.001
		8	0.101	-0.119	26.478	0.001
		9	-0.074	-0.065	27.083	0.001
		10	0.078	0.092	27.769	0.002
		11	-0.065	-0.039	28.244	0.003
		12	-0.065	0.114	28.726	0.004
		13	-0.036	0.038	28.877	0.007
		14	-0.031	-0.015	28.991	0.010
		15	0.044	0.039	29.225	0.015
		16	-0.074	-0.140	29.889	0.019
		17	0.035	-0.077	30.036	0.026
		18	0.137	0.018	32.347	0.020
		19	-0.023	-0.100	32.415	0.028
		20	-0.027	-0.024	32.505	0.038
		21	-0.047	-0.054	32.784	0.049
		22	-0.029	-0.064	32.895	0.063
		23	-0.042	-0.036	33.124	0.079
		24	0.022	0.017	33.187	0.100
		25	0.044	-0.092	33.446	0.120
		26	-0.039	-0.063	33.656	0.144
		27	0.052	-0.012	34.024	0.165
		28	0.034	-0.064	34.182	0.195
		29	-0.081	-0.079	35.121	0.201
		30	0.059	0.050	35.617	0.221
		31	-0.013	-0.079	35.642	0.259
		32	-0.061	-0.031	36.201	0.279
		33	-0.030	-0.067	36.336	0.316
		34	0.068	0.012	37.054	0.330
		35	-0.063	-0.197	37.677	0.348
		36	-0.100	-0.194	39.271	0.325

After taking the first difference, from the correlogram on the right we can see the ACs decline rather quickly. But, on conducting the ADF tests on both level and 1st difference, we can see that the series is stationary. Figure 5 shows the correlogram for level

Null Hypothesis: D(WEEKLY_SALES) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 5 (Automatic - based on AIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.260194	0.0000
Test critical values:	1% level	-4.059734	
	5% level	-3.458856	
	10% level	-3.155470	

*MacKinnon (1996) one-sided p-values.

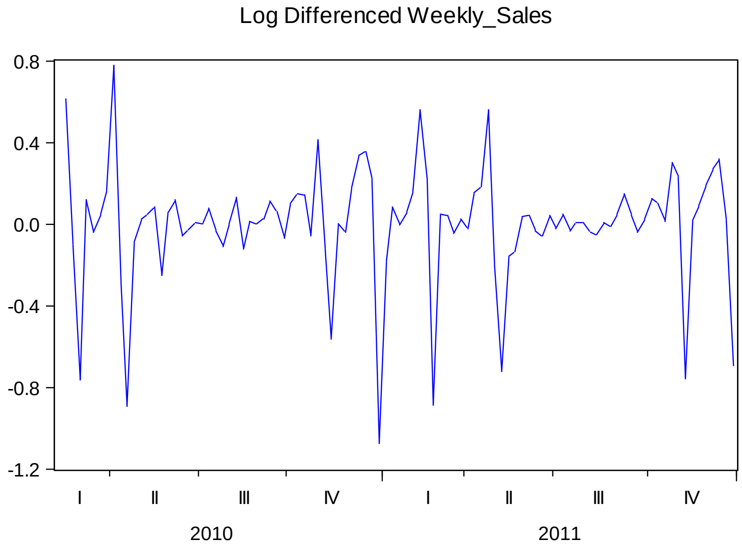
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(WEEKLY_SALES,2)
Method: Least Squares
Date: 04/23/23 Time: 13:38
Sample (adjusted): 3/26/2010 12/30/2011
Included observations: 93 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(WEEKLY_SALES(-1))	-3.138419	0.379945	-8.260194	0.0000
D(WEEKLY_SALES(-1),2)	1.946154	0.322995	6.025343	0.0000
D(WEEKLY_SALES(-2),2)	1.392788	0.275117	5.062520	0.0000
D(WEEKLY_SALES(-3),2)	1.063354	0.216955	4.901262	0.0000
D(WEEKLY_SALES(-4),2)	0.648542	0.151549	4.279429	0.0000
D(WEEKLY_SALES(-5),2)	0.304057	0.102400	2.969297	0.0039
C	-1307.193	1788.712	-0.730801	0.4669
@TREND("2/05/2010")	26.04156	30.19264	0.862513	0.3908

R-squared	0.644899	Mean dependent var	-263.7755
Adjusted R-squared	0.615655	S.D. dependent var	12469.39
S.E. of regression	7730.466	Akaike info criterion	20.82582
Sum squared resid	5.08E+09	Schwarz criterion	21.04368
Log likelihood	-960.4007	Hannan-Quinn criter.	20.91379
F-statistic	22.05265	Durbin-Watson stat	1.957776
Prob(F-statistic)	0.000000		

From the ADF tests, we can see that the absolute ADF value is greater than the critical values which confirms that the series is stationary. We can also see that all the lagged variables are statistically significant.

When we plot the log differenced graph for the ADF, we can see that the retail sales is stationary at first difference as the trend line moves around 0.00



I started off with lag 1 in both AR and MA which yielded in a non-significant MA coefficient. Then, I tried different lags based on the spikes in the AC and PAC values in the correlogram.

Dependent Variable: D(WEEKLY_SALES)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 11:28
Sample: 2/12/2010 12/30/2011
Included observations: 99
Failure to improve objective (non-zero gradients) after 30 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.750068	123.1395	-0.022333	0.9822
AR(1)	0.591676	0.122176	4.842829	0.0000
MA(1)	-1.000000	486.6703	-0.002055	0.9984
SIGMASQ	68853152	7.45E+08	0.092362	0.9266

R-squared	0.208910	Mean dependent var	-15.89515
Adjusted R-squared	0.183929	S.D. dependent var	9376.778
S.E. of regression	8470.669	Akaike info criterion	20.99923
Sum squared resid	6.82E+09	Schwarz criterion	21.10409
Log likelihood	-1035.462	Hannan-Quinn criter.	21.04166
F-statistic	8.362507	Durbin-Watson stat	1.590808
Prob(F-statistic)	0.000054		

Inverted AR Roots	.59
Inverted MA Roots	1.00

Dependent Variable: WEEKLY_SALES/CPI
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 12:44
Sample: 2/05/2010 12/30/2011
Included observations: 100
Convergence achieved after 46 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	108.9683	11.80728	9.228910	0.0000
AR(7)	0.294636	0.061818	4.766203	0.0000
MA(1)	0.636219	0.094124	6.759338	0.0000
SIGMASQ	1265.986	152.7152	8.289854	0.0000

R-squared	0.438816	Mean dependent var	107.5783
Adjusted R-squared	0.421279	S.D. dependent var	47.73581
S.E. of regression	36.31440	Akaike info criterion	10.07328
Sum squared resid	126598.6	Schwarz criterion	10.17748
Log likelihood	-499.6638	Hannan-Quinn criter.	10.11545
F-statistic	25.02231	Durbin-Watson stat	1.746787
Prob(F-statistic)	0.000000		

Inverted AR Roots	.84	.52+.66i	.52-.66i	-.19-.82i
	-.19+.82i	-.76-.36i	-.76+.36i	
Inverted MA Roots	-.64			

Dependent Variable: D(WEEKLY_SALES/CPI)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 13:13
Sample: 2/12/2010 12/30/2011
Included observations: 99
Convergence achieved after 33 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.036074	2.849487	0.012660	0.9899
AR(7)	0.233142	0.070406	3.311383	0.0013
MA(2)	-0.459549	0.092823	-4.950824	0.0000
SIGMASQ	1446.963	150.3351	9.624916	0.0000

R-squared	0.244640	Mean dependent var	-0.118252
Adjusted R-squared	0.220787	S.D. dependent var	43.99023
S.E. of regression	38.83153	Akaike info criterion	10.20465
Sum squared resid	143249.3	Schwarz criterion	10.30950
Log likelihood	-501.1300	Hannan-Quinn criter.	10.24707
F-statistic	10.25595	Durbin-Watson stat	2.178969
Prob(F-statistic)	0.000007		

Inverted AR Roots	.81	.51-.63i	.51+.63i	-.18-.79i
	-.18+.79i	-.73-.35i	-.73+.35i	
Inverted MA Roots	.68	-.68		

From the above results when dependent variable weekely_sales/cpi was used, we can see the combinations which worked out of the numerous ones I tried. On the left we can see the equation with AR(7) and MA(1) values which yielded statistically significant coefficients. I tried AR(7) due to the spike present in AC in the correlogram for the weekly_sales at levels format.
Later, I got another significant result from the first difference of weekly sales/cpi with AR(7) and MA(2). Since the alternating PAC signs are not very clear from the correlogram, this value of MA was achieved on the basis of trial and error.
As we can see from this set of equations, the adjusted R squared is much better in a level form than the first difference form. Therefore, we can rule out the first difference from weekly_sales/cpi specification.

Dependent Variable: WEEKLY_SALES
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 15:26
Sample: 2/05/2010 12/30/2011
Included observations: 100
Convergence achieved after 42 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23287.66	2512.820	9.267537	0.0000
AR(7)	0.294054	0.062642	4.694239	0.0000
MA(1)	0.638346	0.093102	6.856417	0.0000
SIGMASQ	57681460	7047948.	8.184149	0.0000

R-squared	0.440026	Mean dependent var	22979.00
Adjusted R-squared	0.422527	S.D. dependent var	10200.39
S.E. of regression	7751.442	Akaike info criterion	20.80014
Sum squared resid	5.77E+09	Schwarz criterion	20.90435
Log likelihood	-1036.007	Hannan-Quinn criter.	20.84231
F-statistic	25.14554	Durbin-Watson stat	1.742272
Prob(F-statistic)	0.000000		

Inverted AR Roots	.84	.52+.66i	.52-.66i	-.19-.82i
	-.19+.82i	-.76-.36i	-.76+.36i	
Inverted MA Roots	-.64			

Dependent Variable: D(WEEKLY_SALES)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 22:36
Sample: 2/12/2010 12/30/2011
Included observations: 99
Convergence achieved after 30 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.54293	607.6113	0.032164	0.9744
AR(7)	0.234677	0.070837	3.312909	0.0013
MA(2)	-0.460253	0.092964	-4.950866	0.0000
SIGMASQ	65722616	6889319.	9.539784	0.0000

R-squared	0.244879	Mean dependent var	-15.89515
Adjusted R-squared	0.221033	S.D. dependent var	9376.778
S.E. of regression	8275.862	Akaike info criterion	20.92845
Sum squared resid	6.51E+09	Schwarz criterion	21.03330
Log likelihood	-1031.958	Hannan-Quinn criter.	20.97087
F-statistic	10.26920	Durbin-Watson stat	2.174941
Prob(F-statistic)	0.000006		

Inverted AR Roots	.81	.51+.64i	.51-.64i	-.18-.79i
	-.18+.79i	-.73-.35i	-.73+.35i	
Inverted MA Roots	.68	-.68		

On using only weekly_sales as the dependent variable, we can see a similar behaviour as before in which the first difference equations does not fair well in terms of the adjusted R squared but the Aakike and Schwarz values changed significantly.

I found another equation for first difference with AR(7) and MA(8) specification which yielded statistically significant coefficients but the adjusted R squared values were much lower so will be discarding this ARIMA (Figure 14)

Then I combined the significant AR and MA terms to produce the following equation:

Dependent Variable: D(WEEKLY_SALES)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 23:09
Sample: 2/12/2010 12/30/2011
Included observations: 99
Convergence achieved after 29 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	45.89453	770.4967	0.059565	0.9526
AR(7)	0.255022	0.074988	3.400847	0.0010
MA(2)	-0.490829	0.097630	-5.027438	0.0000
MA(8)	0.237987	0.087772	2.711412	0.0080
SIGMASQ	61484086	6406858.	9.596606	0.0000

R-squared	0.293577	Mean dependent var	-15.89515
Adjusted R-squared	0.263517	S.D. dependent var	9376.778
S.E. of regression	8047.019	Akaike info criterion	20.88971
Sum squared resid	6.09E+09	Schwarz criterion	21.02077
Log likelihood	-1029.041	Hannan-Quinn criter.	20.94274
F-statistic	9.766204	Durbin-Watson stat	2.162806
Prob(F-statistic)	0.000001		

Inverted AR Roots	.82	.51+.64i	.51-.64i	-.18+.80i
	-.18-.80i	-.74-.36i	-.74+.36i	
Inverted MA Roots	.84+.27i	.84-.27i	.34+.71i	.34-.71i
	-.34+.71i	-.34-.71i	-.84+.27i	-.84-.27i

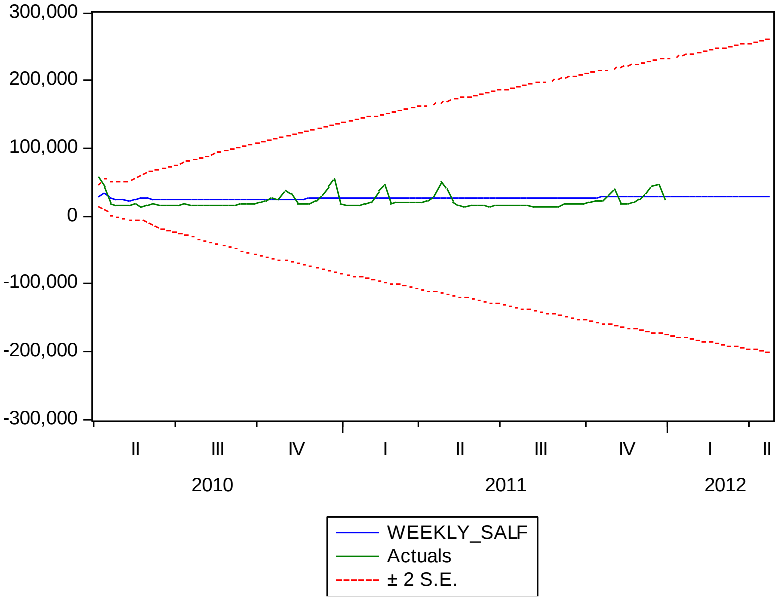
As we can see, the adjusted R squared value is much lesser(0.26) for this equation as compared to the previous one (0.42).
Therefore our final equation for retail sales of store 1 and department 1 is:
Retail_sales = c + AR(7) + MA(1)

I also did the same activity for department 2 and 3 of store 1.
From figure 9,11, AR(3) and MA(3) were found to statistically significant indicators for department 2 and AR(3) and MA(6) were found to statistically significant indicators for department 3 which were configured on the basis of their correlograms containing stationary data points.

For the ex-post forecast, I used 2 equations to compare the efficacy for the equations and to see how closely it is able to predict the actual sales in 2012; the final equation generated D(Retail_sales) = c + AR(7) + MA(2) and D(Retail_sales)= c+ AR(7)+ MA(2) + MA(8)

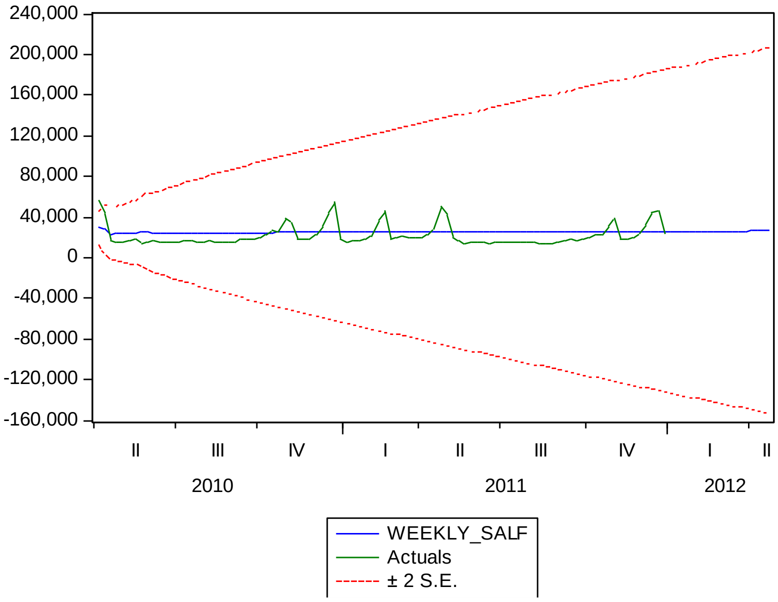
As we have already looked at the other factors to determine the health of these equations, we will now look at how well these two equations performed during forecast. From the theil statistics, we can see that the Theil statistic is slightly better for equation. The Bias 0.17 and for equation 2 is 0.099 and therefore is better for equation 2 but the variance is lower for equation 1. When it comes to covariance, the covariance is closer to 1 for equation 1. Looking at all these factors, neither one of the equations is a clear winner and can be used on a case to case basis.

Equation 1: D(Retail_sales) = c + AR(7) + MA(2)



Forecast: WEEKLY_SALF	
Actual: WEEKLY_SALES	
Forecast sample: 2/05/2010 4/27/2012	
Adjusted sample: 4/02/2010 4/27/2012	
Included observations: 109	
Root Mean Squared Error	10705.09
Mean Absolute Error	9596.762
Mean Abs. Percent Error	46.99436
Theil Inequality Coef.	0.206709
Bias Proportion	0.176299
Variance Proportion	0.661741
Covariance Proportion	0.161961
Theil U2 Coefficient	2.206660
Symmetric MAPE	38.89949

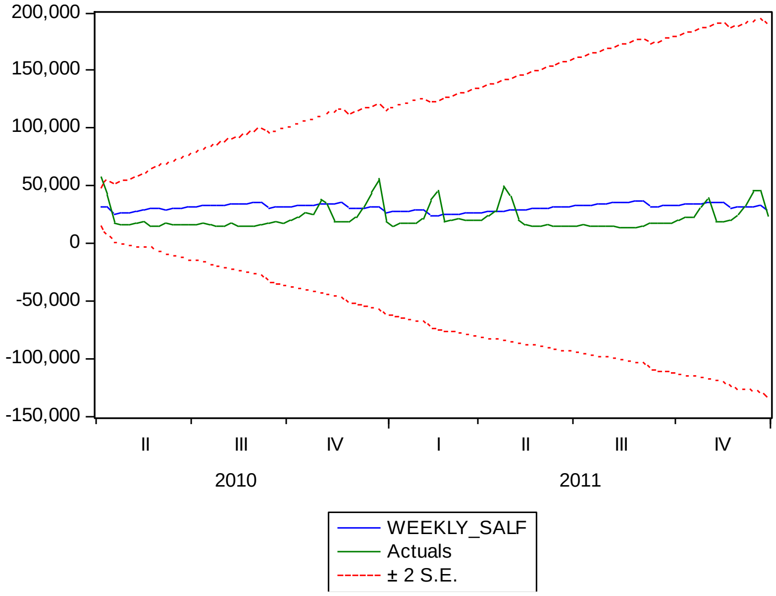
Equation 2: D(Retail_sales)= c+ AR(7)+ MA(2) + MA(8)



Forecast: WEEKLY_SALF	
Actual: WEEKLY_SALES	
Forecast sample: 2/05/2010 4/27/2012	
Adjusted sample: 4/02/2010 4/27/2012	
Included observations: 109	
Root Mean Squared Error	10318.67
Mean Absolute Error	8945.201
Mean Abs. Percent Error	42.45288
Theil Inequality Coef.	0.204246
Bias Proportion	0.099226
Variance Proportion	0.810439
Covariance Proportion	0.090335
Theil U2 Coefficient	2.009527
Symmetric MAPE	36.53154

Later, I again tried to use the isHoliday dummy to see if it becomes statistically significant now that we have our final equations at hand and viola, it did! While the Theil coefficients did not improve much, but the forcast started showing some much needed trend. Since we have only one AR term in our model, we don't have to check for the sum of coefficients to be lesser than 1.

Equation 3: D(weekly_sales)= c + AR(7) + MA(2) + isholiday



Forecast: WEEKLY_SALF	
Actual: WEEKLY_SALES	
Forecast sample: 2/05/2010 4/27/2012	
Adjusted sample: 4/02/2010 12/30/2011	
Included observations: 92	
Root Mean Squared Error	14141.89
Mean Absolute Error	12964.40
Mean Abs. Percent Error	68.28312
Theil Inequality Coef.	0.250497
Bias Proportion	0.409146
Variance Proportion	0.244531
Covariance Proportion	0.346324
Theil U2 Coefficient	3.313967
Symmetric MAPE	49.26352

Using the significant factors from both the types of models, here is an attempt at modelling a mix of multivariate and arima models.

Dependent Variable: D(WEEKLY_SALES/CPI)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/25/23 Time: 02:01
Sample: 2/12/2010 12/30/2011
Included observations: 99
Convergence achieved after 33 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	104.1593	22.34099	4.662251	0.0000
ISHOLIDAY	-27.40924	10.04170	-2.729541	0.0076
TEMPERATURE	-0.796264	0.221272	-3.598577	0.0005
WEEKLY_SALES(-1)/CPI	-0.448966	0.101317	-4.431279	0.0000
AR(7)	0.237470	0.076861	3.089611	0.0027
MA(2)	-0.458335	0.117144	-3.912586	0.0002
SIGMASQ	1072.952	115.1057	9.321442	0.0000

R-squared	0.439886	Mean dependent var	-0.118252
Adjusted R-squared	0.403356	S.D. dependent var	43.99023
S.E. of regression	33.97925	Akaike info criterion	9.966323
Sum squared resid	106222.2	Schwarz criterion	10.14982
Log likelihood	-486.3330	Hannan-Quinn criter.	10.04056
F-statistic	12.04203	Durbin-Watson stat	1.723918
Prob(F-statistic)	0.000000		

Inverted AR Roots	.81	.51-.64i	.51+.64i	-.18-.79i
	-.18+.79i	-.73-.35i	-.73+.35i	
Inverted MA Roots	.68	-.68		

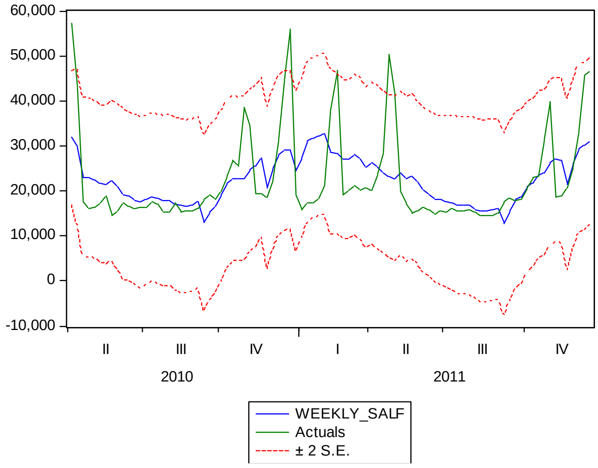
Dependent Variable: D(WEEKLY_SALES/CPI)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 13:13
Sample: 2/12/2010 12/30/2011
Included observations: 99
Convergence achieved after 33 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.036074	2.849487	0.012660	0.9899
AR(7)	0.233142	0.070406	3.311383	0.0013
MA(2)	-0.459549	0.092823	-4.950824	0.0000
SIGMASQ	1446.963	150.3351	9.624916	0.0000

R-squared	0.244640	Mean dependent var	-0.118252
Adjusted R-squared	0.220787	S.D. dependent var	43.99023
S.E. of regression	38.83153	Akaike info criterion	10.20465
Sum squared resid	143249.3	Schwarz criterion	10.30950
Log likelihood	-501.1300	Hannan-Quinn criter.	10.24707
F-statistic	10.25595	Durbin-Watson stat	2.178969
Prob(F-statistic)	0.000007		

Inverted AR Roots	.81	.51-.63i	.51+.63i	-.18-.79i
	-.18+.79i	-.73-.35i	-.73+.35i	
Inverted MA Roots	.68	-.68		

Here, I am comparing the first difference of dependent variable weekly_sales/cpi ARIMA and multivariate models. From the results above, we can see a lot of explanatory power has been adding more X variables in the multivariate model. While there us not a significant difference in the Akakie and Shwartz values for the 2 models, we can definitely better adjusted R squared values and a smaller standard error in the multivariate equation. The forecast for this models has a more trended graph, since it is taking other factors into account and quite intuitively has a vey low bias as per the forecast details. (Figure 15)



Forecast: WEEKLY_SALF	
Actual: WEEKLY_SALES	
Forecast sample: 2/05/2010 4/27/2012	
Adjusted sample: 4/02/2010 12/23/2011	
Included observations: 91	
Root Mean Squared Error	8368.012
Mean Absolute Error	5820.615
Mean Abs. Percent Error	23.61902
Theil Inequality Coef.	0.175654
Bias Proportion	0.000424
Variance Proportion	0.381237
Covariance Proportion	0.618338
Theil U2 Coefficient	1.359352
Symmetric MAPE	22.61041

Running tests on this final equation to check it's fitness. Since we have used ARIMA methods, it is only possible to have the following issues:

1. Heteroskedasticity: The first test we will check if the variance is consistent in the equation or not. I have used the White test to check it. The null hypothesis of this test is that the data is homoscedastic and the alternate hypothesis is that there is heteroskedasticity present in the data. From figure 16, we can see that the probability is 0, which is lesser than 0.05, therefore, we can say that heterskedasticity is present. (Figure 16)

To correct for it, I should have used the white correction or the newey west correction which would have reduced the standard error but unfortunately, I could not find the correction option in the student version.

2. Structural breaks: By looking at the peaks in actual weekly sales graph, we can check if these structural breaks persist in the forecast model. For this I have used the Chow breakpoint test. I set the breakpoint at 12/24/2010 which was the highest peak present in the plot. The null hypothesis in the chow test is that there are no structural breaks and the alternative hypothesis is that there is a structural break. On running the test for the equation, the p value was found to be greater than 0.05. Therefore, we can say that there are no structural breaks present. (Figure 17)

In conclusion, while I may not be sure if the equation is perfect for a real forecast, it sure covers a lot of breath in terms of the kind of variables used and tests conducted. It would have been helpful if there was more explanatory data available in the dataset about retail_sales such as number of visits by customers and consumer confidence index. I am particularly happy about using both estimation techniques (ARIMA and multivariate regression) in conjunction to get a consolidated equation because, in my opinion that's how a true equation would be made for a real life scenario since it is hard to rely only on one of the techniques.

APPENDIX

FIGURE1

Date: 04/22/23 Time: 00:43
Sample: 2/05/2010 10/26/2012

	UNEMPLOYMENT	TEMPERATURE	CPI	FUEL_PRICE	WEEKLY ...
Mean	7.610420	68.30678	215.9969	3.219699	46102.09
Median	7.787000	69.64000	215.4599	3.290000	45561.85
Maximum	8.106000	91.65000	223.4443	3.907000	65615.36
Minimum	6.573000	35.40000	210.3374	2.514000	35819.83
Std. Dev.	0.383749	14.25049	4.350890	0.427313	3440.673
Skewness	-1.046172	-0.402643	0.265983	-0.151813	1.721073
Kurtosis	3.043197	2.114668	1.547967	1.600071	10.79032
Jarque-Bera	26.09612	8.534116	14.24867	12.22644	432.2020
Probability	0.000002	0.014023	0.000805	0.002213	0.000000
Sum	1088.290	9767.870	30887.56	460.4170	6592599.
Sum Sq. Dev.	20.91137	28836.84	2688.095	25.92863	1.68E+09
Observations	143	143	143	143	143

FIGURE 2

Date: 04/20/23 Time: 14:53
Sample: 2/05/2010 10/26/2012

	WEEKLY_SALES02	UNEMPLOYMENT	TEMPERATURE	FUEL_PRICE
Mean	13150.48	7.610420	68.30678	3.219699
Median	10366.85	7.787000	69.64000	3.290000
Maximum	51159.17	8.106000	91.65000	3.907000
Minimum	6165.730	6.573000	35.40000	2.514000
Std. Dev.	8708.979	0.383749	14.25049	0.427313
Skewness	2.801496	-1.046172	-0.402643	-0.151813
Kurtosis	10.72404	3.043197	2.114668	1.600071
Jarque-Bera	542.5320	26.09612	8.534116	12.22644
Probability	0.000000	0.000002	0.014023	0.002213
Sum	1880518.	1088.290	9767.870	460.4170
Sum Sq. Dev.	1.08E+10	20.91137	28836.84	25.92863
Observations	143	143	143	143

FIGURE 3

Dependent Variable: WEEKLY_SALES01/CPI
Method: Least Squares
Date: 04/20/23 Time: 14:49
Sample: 2/05/2010 10/26/2012
Included observations: 143

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	189.8342	37.03962	5.125166	0.0000
TEMPERATURE	0.003634	0.096273	0.037751	0.9699
UNEMPLOYMENT	5.055332	3.856943	1.310709	0.1921
FUEL_PRICE/CPI	-1012.623	844.1672	-1.199552	0.2324

R-squared	0.038946	Mean dependent var	213.4856
Adjusted R-squared	0.018204	S.D. dependent var	15.97317
S.E. of regression	15.82712	Akaike info criterion	8.388900
Sum squared resid	34819.17	Schwarz criterion	8.471777
Log likelihood	-595.8063	Hannan-Quinn criter.	8.422577
F-statistic	1.877631	Durbin-Watson stat	1.785755
Prob(F-statistic)	0.136178		

FIGURE 4

Dependent Variable: WEEKLY_SALES02/CPI
Method: Least Squares
Date: 04/20/23 Time: 14:55
Sample: 2/05/2010 10/26/2012
Included observations: 143

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	65.01343	89.07242	0.729894	0.4667
TEMPERATURE	0.987940	0.231516	4.267259	0.0000
UNEMPLOYMENT	-4.745532	9.275129	-0.511641	0.6097
FUEL_PRICE/CPI	-2387.555	2030.043	-1.176110	0.2416

R-squared	0.120368	Mean dependent var	60.84879
Adjusted R-squared	0.101383	S.D. dependent var	40.15052
S.E. of regression	38.06085	Akaike info criterion	10.14382
Sum squared resid	201359.3	Schwarz criterion	10.22670
Log likelihood	-721.2833	Hannan-Quinn criter.	10.17750
F-statistic	6.340200	Durbin-Watson stat	0.473092
Prob(F-statistic)	0.000463		

FIGURE 5

Dependent Variable: WEEKLY_SALES01/CPI
Method: Least Squares
Date: 04/21/23 Time: 13:53
Sample (adjusted): 2/12/2010 10/26/2012
Included observations: 142 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	158.4440	40.08673	3.952530	0.0001
TEMPERATURE	-0.014413	0.097294	-0.148137	0.8825
WEEKLY_SALES01(-1)/CPI(-1)	0.177620	0.091674	1.937514	0.0548
FUEL_PRICE/CPI	-873.0102	834.5036	-1.046143	0.2974
ISHOLIDAY	-11.78398	5.727507	-2.057436	0.0416
UNEMPLOYMENT	4.172212	3.824806	1.090830	0.2773

R-squared	0.072028	Mean dependent var	213.3008
Adjusted R-squared	0.037912	S.D. dependent var	15.87558
S.E. of regression	15.57174	Akaike info criterion	8.370127
Sum squared resid	32977.14	Schwarz criterion	8.495021
Log likelihood	-588.2790	Hannan-Quinn criter.	8.420879
F-statistic	2.111234	Durbin-Watson stat	2.015849
Prob(F-statistic)	0.067744		

FIGURE 6

Dependent Variable: WEEKLY_SALES01/CPI
Method: Least Squares
Date: 04/21/23 Time: 13:53
Sample (adjusted): 2/12/2010 10/26/2012
Included observations: 142 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	158.4440	40.08673	3.952530	0.0001
TEMPERATURE	-0.014413	0.097294	-0.148137	0.8825
WEEKLY_SALES01(-1)/CPI(-1)	0.177620	0.091674	1.937514	0.0548
FUEL_PRICE/CPI	-873.0102	834.5036	-1.046143	0.2974
ISHOLIDAY	-11.78398	5.727507	-2.057436	0.0416
UNEMPLOYMENT	4.172212	3.824806	1.090830	0.2773

R-squared	0.072028	Mean dependent var	213.3008
Adjusted R-squared	0.037912	S.D. dependent var	15.87558
S.E. of regression	15.57174	Akaike info criterion	8.370127
Sum squared resid	32977.14	Schwarz criterion	8.495021
Log likelihood	-588.2790	Hannan-Quinn criter.	8.420879
F-statistic	2.111234	Durbin-Watson stat	2.015849
Prob(F-statistic)	0.067744		

FIGURE 7

Dependent Variable: WEEKLY_SALES02/CPI
Method: Least Squares
Date: 04/23/23 Time: 00:20
Sample (adjusted): 2/12/2010 10/26/2012
Included observations: 142 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.90054	55.86823	0.338306	0.7357
UNEMPLOYMENT	-0.733103	5.824582	-0.125864	0.9000
TEMPERATURE	0.202420	0.158643	1.275946	0.2041
FUEL_PRICE/CPI	-882.9622	1277.984	-0.690903	0.4908
ISHOLIDAY	-21.76536	8.172871	-2.663123	0.0087
WEEKLY_SALES02(-1)/CPI(-1)	0.792380	0.053748	14.74257	0.0000

R-squared	0.662685	Mean dependent var	60.81892
Adjusted R-squared	0.650284	S.D. dependent var	40.29105
S.E. of regression	23.82684	Akaike info criterion	9.220838
Sum squared resid	77209.72	Schwarz criterion	9.345732
Log likelihood	-648.6795	Hannan-Quinn criter.	9.271590
F-statistic	53.43679	Durbin-Watson stat	1.800085
Prob(F-statistic)	0.000000		

FIGURE 8

Null Hypothesis: WEEKLY_SALES has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on AIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.428270	0.0000
Test critical values:	1% level	-3.498439	
	5% level	-2.891234	
	10% level	-2.582678	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(WEEKLY_SALES)
Method: Least Squares
Date: 04/23/23 Time: 13:52
Sample (adjusted): 2/19/2010 12/30/2011
Included observations: 98 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
WEEKLY_SALES(-1)	-0.556569	0.086581	-6.428270	0.0000
D(WEEKLY_SALES(-1))	0.319563	0.097846	3.265966	0.0015
C	12473.42	2126.143	5.866690	0.0000

R-squared	0.303446	Mean dependent var	-231.5164
Adjusted R-squared	0.288782	S.D. dependent var	9174.976
S.E. of regression	7737.602	Akaike info criterion	20.77571
Sum squared resid	5.69E+09	Schwarz criterion	20.85484
Log likelihood	-1015.010	Hannan-Quinn criter.	20.80771
F-statistic	20.69285	Durbin-Watson stat	1.929904
Prob(F-statistic)	0.000000		

FIGURE 9 weekly_sales01

Date: 04/24/23 Time: 13:35
Sample: 2/05/2010 12/30/2011
Included observations: 100

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.114	0.114	1.3351	0.248
		2	0.075	0.063	1.9223	0.382
		3	-0.131	-0.148	3.7135	0.294
		4	0.035	0.063	3.8417	0.428
		5	-0.135	-0.131	5.8010	0.326
		6	-0.115	-0.116	7.2379	0.299
		7	-0.085	-0.026	8.0344	0.330
		8	-0.055	-0.071	8.3695	0.398
		9	-0.007	-0.004	8.3743	0.497
		10	-0.006	-0.018	8.3783	0.592
		11	0.089	0.056	9.2769	0.596
		12	-0.059	-0.101	9.6746	0.644
		13	-0.029	-0.055	9.7743	0.712
		14	-0.090	-0.071	10.728	0.707
		15	-0.046	-0.077	10.985	0.754
		16	0.050	0.087	11.287	0.791
		17	0.145	0.124	13.880	0.676
		18	-0.000	-0.072	13.880	0.737
		19	-0.069	-0.093	14.483	0.755
		20	-0.027	-0.023	14.578	0.800
		21	-0.138	-0.187	17.042	0.709
		22	-0.014	0.026	17.068	0.760
		23	-0.112	-0.063	18.736	0.716
		24	0.018	-0.016	18.782	0.764
		25	-0.005	0.023	18.786	0.807
		26	0.060	-0.029	19.281	0.824
		27	-0.005	-0.087	19.285	0.859
		28	0.064	-0.009	19.859	0.870
		29	-0.023	-0.048	19.934	0.895
		30	-0.000	-0.012	19.934	0.918
		31	-0.048	-0.022	20.281	0.930
		32	-0.036	-0.029	20.475	0.943
		33	0.052	0.002	20.888	0.950
		34	-0.022	-0.090	20.962	0.961
		35	0.093	0.071	22.312	0.952
		36	0.055	0.036	22.802	0.957

Date: 04/24/23 Time: 13:36
Sample (adjusted): 2/12/2010 12/30/2011
Included observations: 99 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.449	-0.449	20.533	0.000
		2	0.074	-0.159	21.096	0.000
		3	-0.206	-0.307	25.519	0.000
		4	0.179	-0.074	28.905	0.000
		5	-0.101	-0.103	29.982	0.000
		6	-0.006	-0.164	29.986	0.000
		7	0.004	-0.086	29.988	0.000
		8	-0.019	-0.141	30.029	0.000
		9	0.023	-0.107	30.089	0.000
		10	-0.050	-0.148	30.367	0.001
		11	0.140	0.023	32.595	0.001
		12	-0.097	-0.037	33.680	0.001
		13	0.043	-0.026	33.899	0.001
		14	-0.053	-0.022	34.235	0.002
		15	-0.032	-0.158	34.356	0.003
		16	0.000	-0.140	34.356	0.005
		17	0.135	0.065	36.592	0.004
		18	-0.042	0.044	36.812	0.006
		19	-0.059	-0.024	37.248	0.007
		20	0.084	0.122	38.147	0.008
		21	-0.134	-0.107	40.437	0.007
		22	0.121	0.001	42.344	0.006
		23	-0.124	-0.051	44.362	0.005
		24	0.085	-0.071	45.330	0.005
		25	-0.052	-0.011	45.690	0.007
		26	0.066	0.026	46.295	0.008
		27	-0.071	-0.052	46.986	0.010
		28	0.080	-0.015	47.895	0.011
		29	-0.059	-0.054	48.385	0.013
		30	0.043	-0.026	48.652	0.017
		31	-0.030	-0.023	48.781	0.022
		32	-0.043	-0.045	49.057	0.027
		33	0.097	0.058	50.483	0.026
		34	-0.109	-0.099	52.294	0.023
		35	0.079	-0.034	53.266	0.025
		36	0.003	0.050	53.268	0.032

Dependent Variable: WEEKLY SALES01
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 15:01
Sample: 2/05/2010 12/30/2011
Included observations: 100
Convergence achieved after 46 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	45751.93	131.1147	348.9458	0.0000
AR(3)	0.584838	0.144405	4.049982	0.0001
MA(3)	-0.915936	0.072886	-12.56667	0.0000
SIGMASQ	11385556	1322680.	8.607945	0.0000

R-squared	0.133100	Mean dependent var	45899.56
Adjusted R-squared	0.106009	S.D. dependent var	3642.294
S.E. of regression	3443.828	Akaike info criterion	19.18700
Sum squared resid	1.14E+09	Schwarz criterion	19.29120
Log likelihood	-955.3498	Hannan-Quinn criter.	19.22917
F-statistic	4.913138	Durbin-Watson stat	1.806943
Prob(F-statistic)	0.003217		

Inverted AR Roots	.84	-.42+.72i	-.42-.72i
Inverted MA Roots	.97	-.49+.84i	-.49-.84i

FIGURE 10- weekly_sales02

Date: 04/24/23 Time: 13:37
Sample: 2/05/2010 12/30/2011
Included observations: 100

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.791	0.791	64.525	0.000
		2	0.552	-0.198	96.290	0.000
		3	0.353	-0.040	109.39	0.000
		4	0.144	-0.189	111.59	0.000
		5	-0.001	0.015	111.59	0.000
		6	-0.114	-0.100	112.99	0.000
		7	-0.193	-0.034	117.08	0.000
		8	-0.218	-0.004	122.33	0.000
		9	-0.232	-0.067	128.35	0.000
		10	-0.221	-0.008	133.90	0.000
		11	-0.215	-0.095	139.21	0.000
		12	-0.205	-0.020	144.10	0.000
		13	-0.195	-0.074	148.54	0.000
		14	-0.179	-0.014	152.32	0.000
		15	-0.141	0.007	154.69	0.000
		16	-0.105	-0.033	156.03	0.000
		17	-0.095	-0.085	157.13	0.000
		18	-0.116	-0.131	158.79	0.000
		19	-0.069	0.163	159.39	0.000
		20	-0.041	-0.117	159.60	0.000
		21	-0.041	-0.035	159.81	0.000
		22	-0.039	-0.082	160.01	0.000
		23	-0.038	0.011	160.20	0.000
		24	-0.015	0.017	160.23	0.000
		25	0.005	-0.055	160.23	0.000
		26	-0.004	-0.057	160.24	0.000
		27	0.008	0.003	160.25	0.000
		28	0.006	-0.038	160.25	0.000
		29	0.005	-0.023	160.25	0.000
		30	0.003	-0.043	160.26	0.000
		31	-0.004	-0.029	160.26	0.000
		32	0.000	-0.013	160.26	0.000
		33	-0.019	-0.075	160.31	0.000
		34	-0.052	-0.081	160.73	0.000
		35	-0.044	0.045	161.04	0.000
		36	-0.043	-0.101	161.33	0.000

Date: 04/24/23 Time: 13:37
Sample (adjusted): 2/12/2010 12/30/2011
Included observations: 99 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.081	0.081	0.6630	0.416
		2	-0.096	-0.103	1.6088	0.447
		3	0.022	0.040	1.6605	0.646
		4	-0.158	-0.177	4.2881	0.368
		5	-0.079	-0.042	4.9445	0.423
		6	-0.079	-0.111	5.6126	0.468
		7	-0.132	-0.125	7.5132	0.377
		8	-0.027	-0.058	7.5953	0.474
		9	-0.058	-0.110	7.9681	0.537
		10	0.012	-0.019	7.9836	0.630
		11	-0.010	-0.099	7.9953	0.714
		12	-0.001	-0.038	7.9954	0.785
		13	-0.013	-0.098	8.0150	0.843
		14	-0.053	-0.106	8.3459	0.871
		15	0.007	-0.059	8.3510	0.909
		16	0.065	-0.007	8.8533	0.919
		17	0.085	0.035	9.7312	0.915
		18	-0.122	-0.211	11.575	0.868
		19	0.035	0.030	11.732	0.897
		20	0.055	-0.046	12.112	0.912
		21	-0.004	-0.001	12.114	0.936
		22	-0.013	-0.092	12.134	0.955
		23	-0.063	-0.089	12.653	0.959
		24	0.002	-0.013	12.654	0.971
		25	0.060	-0.011	13.146	0.975
		26	-0.051	-0.076	13.507	0.979
		27	0.031	-0.027	13.638	0.985
		28	-0.006	-0.050	13.644	0.990
		29	-0.001	-0.027	13.644	0.993
		30	0.023	-0.026	13.723	0.995
		31	-0.026	-0.052	13.819	0.997
		32	0.056	0.017	14.295	0.997
		33	0.038	-0.001	14.511	0.998
		34	-0.101	-0.111	16.076	0.996
		35	0.017	0.023	16.119	0.997
		36	0.037	-0.032	16.338	0.998

FIGURE 11

Dependent Variable: WEEKLY SALES02
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 14:51
Sample: 2/05/2010 12/30/2011
Included observations: 100
Convergence achieved after 22 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12789.84	1692.419	7.557135	0.0000
AR(3)	0.389278	0.084978	4.580920	0.0000
MA(6)	-0.289424	0.131239	-2.205317	0.0298
SIGMASQ	56959323	7159340.	7.955946	0.0000

R-squared	0.176762	Mean dependent var	12758.94
Adjusted R-squared	0.151036	S.D. dependent var	8359.920
S.E. of regression	7702.768	Akaike info criterion	20.78321
Sum squared resid	5.70E+09	Schwarz criterion	20.88742
Log likelihood	-1035.161	Hannan-Quinn criter.	20.82539
F-statistic	6.870897	Durbin-Watson stat	0.594712
Prob(F-statistic)	0.000305		

Inverted AR Roots	.73	-.37+.63i	-.37-.63i	
Inverted MA Roots	.81	.41-.70i	.41+.70i	-.41-.70i
	-.41+.70i	-.81		

FIGURE 12

Dependent Variable: WEEKLY SALES02/CPI
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 14:54
Sample: 2/05/2010 12/30/2011
Included observations: 100
Convergence achieved after 21 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	70.79296	32.20769	2.198014	0.0304
FUEL PRICE/CPI	-756.2596	2113.141	-0.357884	0.7212
AR(3)	0.388272	0.084699	4.584148	0.0000
MA(6)	-0.294211	0.132244	-2.224750	0.0285
SIGMASQ	1246.689	157.8517	7.897846	0.0000

R-squared	0.177773	Mean dependent var	59.74808
Adjusted R-squared	0.143153	S.D. dependent var	39.13501
S.E. of regression	36.22573	Akaike info criterion	10.07374
Sum squared resid	124668.9	Schwarz criterion	10.20399
Log likelihood	-498.6868	Hannan-Quinn criter.	10.12645
F-statistic	5.134963	Durbin-Watson stat	0.596257
Prob(F-statistic)	0.000866		

Inverted AR Roots	.73	-.36+.63i	-.36-.63i	
Inverted MA Roots	.82	.41-.71i	.41+.71i	-.41-.71i
	-.41+.71i	-.82		

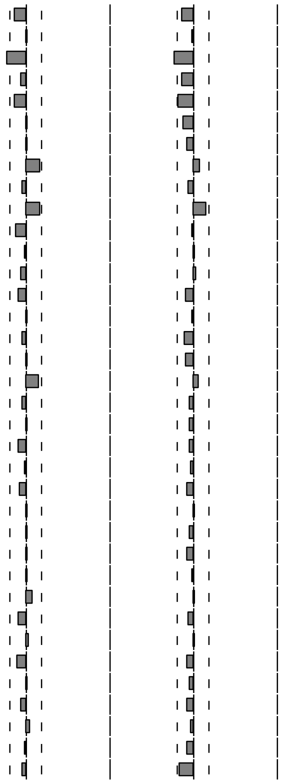
Correlogram of residuals- q stat for ar(7) ma(1) on weekly sales

Date: 04/24/23 Time: 17:26
Sample: 2/05/2010 12/30/2011
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.127	0.127	1.6675	
		2	0.117	0.102	3.0924	
		3	-0.071	-0.100	3.6166	0.057
		4	0.003	0.011	3.6175	0.164
		5	-0.036	-0.019	3.7602	0.289
		6	0.053	0.053	4.0606	0.398
		7	0.041	0.037	4.2442	0.515
		8	0.198	0.178	8.5907	0.198
		9	0.029	-0.017	8.6857	0.276
		10	0.148	0.117	11.165	0.193
		11	-0.104	-0.115	12.410	0.191
		12	-0.045	-0.049	12.642	0.244
		13	-0.103	-0.050	13.896	0.239
		14	-0.142	-0.162	16.272	0.179
		15	-0.063	-0.028	16.752	0.211
		16	-0.088	-0.124	17.682	0.222
		17	-0.050	-0.047	17.993	0.263
		18	0.076	0.069	18.715	0.284
		19	-0.081	-0.070	19.532	0.299
		20	-0.061	-0.049	20.003	0.333
		21	-0.156	-0.064	23.152	0.231
		22	-0.099	-0.021	24.433	0.224
		23	-0.145	-0.096	27.232	0.163
		24	-0.066	0.006	27.822	0.182
		25	-0.060	-0.061	28.319	0.204
		26	-0.056	-0.082	28.745	0.230
		27	-0.033	-0.009	28.901	0.268
		28	0.014	-0.023	28.931	0.314
		29	-0.114	-0.077	30.800	0.279
		30	-0.033	-0.029	30.961	0.319
		31	-0.126	-0.086	33.308	0.265
		32	-0.025	-0.020	33.406	0.305
		33	-0.057	-0.056	33.903	0.329
		34	0.041	0.016	34.164	0.364
		35	-0.002	-0.049	34.165	0.412
		36	-0.013	-0.092	34.190	0.459

Correlogram of residuals ar(7) ma(2)

Date: 04/24/23 Time: 17:29
Sample (adjusted): 2/12/2010 12/30/2011
Q-statistic probabilities adjusted for 2 ARMA terms



Autocorrelation / Partial Correlation		AC	PAC	Q-Stat	Prob
	1	-0.126	-0.126	1.6190	
	2	0.003	-0.013	1.6201	
	3	-0.224	-0.228	6.8263	0.009
	4	-0.064	-0.131	7.2628	0.026
	5	-0.132	-0.184	9.1096	0.028
	6	0.009	-0.114	9.1178	0.058
	7	0.012	-0.081	9.1342	0.104
	8	0.180	0.086	12.712	0.048
	9	-0.036	-0.053	12.852	0.076
	10	0.175	0.160	16.275	0.039
	11	-0.122	-0.016	17.954	0.036
	12	-0.005	0.016	17.957	0.056
	13	-0.065	0.048	18.443	0.072
	14	-0.085	-0.091	19.284	0.082
	15	0.004	-0.000	19.286	0.115
	16	-0.034	-0.099	19.427	0.149
	17	0.004	-0.086	19.429	0.195
	18	0.159	0.074	22.554	0.126
	19	-0.037	-0.045	22.723	0.159
	20	0.023	-0.036	22.790	0.199
	21	-0.098	-0.046	24.023	0.195
	22	-0.009	-0.018	24.035	0.241
	23	-0.080	-0.082	24.878	0.252
	24	0.018	0.008	24.921	0.301
	25	0.007	-0.046	24.928	0.354
	26	0.007	-0.075	24.936	0.409
	27	0.019	-0.005	24.989	0.463
	28	0.084	0.006	25.991	0.464
	29	-0.088	-0.061	27.093	0.459
	30	0.038	0.008	27.301	0.502
	31	-0.105	-0.076	28.931	0.469
	32	0.010	-0.037	28.947	0.520
	33	-0.058	-0.069	29.461	0.545
	34	0.054	-0.021	29.912	0.573
	35	-0.013	-0.066	29.937	0.620
	36	-0.045	-0.164	30.259	0.652

FIGURE 14

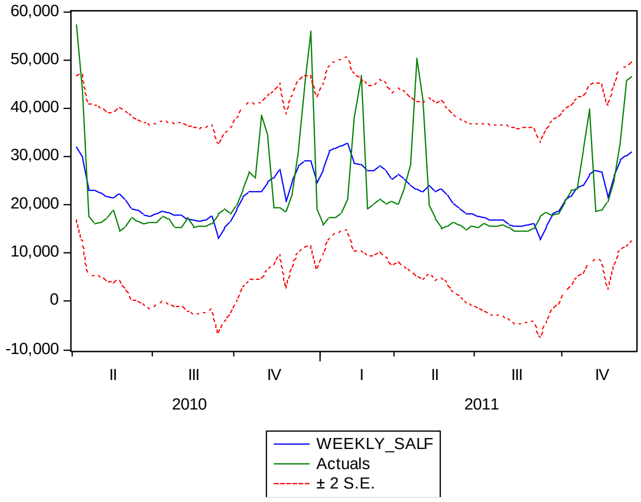
Dependent Variable: D((WEEKLY SALES/CPI))
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/24/23 Time: 13:22
Sample: 2/12/2010 12/30/2011
Included observations: 99
Convergence achieved after 33 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.333778	7.106224	0.046970	0.9626
AR(7)	0.321021	0.062664	5.122884	0.0000
MA(8)	0.205971	0.086146	2.390949	0.0188
SIGMASQ	1692.088	174.0597	9.721307	0.0000

R-squared	0.116677	Mean dependent var	-0.118252
Adjusted R-squared	0.088782	S.D. dependent var	43.99023
S.E. of regression	41.99207	Akaike info criterion	10.36360
Sum squared resid	167516.7	Schwarz criterion	10.46845
Log likelihood	-508.9980	Hannan-Quinn criter.	10.40602
F-statistic	4.182806	Durbin-Watson stat	1.921950
Prob(F-statistic)	0.007925		

Inverted AR Roots	.85	.53+.66i	.53-.66i	-.19-.83i
	-.19+.83i	-.77-.37i	-.77+.37i	
Inverted MA Roots	.76+.31i	.76-.31i	.31-.76i	.31+.76i
	-.31-.76i	-.31+.76i	-.76-.31i	-.76+.31i

FIGURE 15



Forecast: WEEKLY_SALF
Actual: WEEKLY_SALES
Forecast sample: 2/05/2010 4/27/2012
Adjusted sample: 4/02/2010 12/23/2011
Included observations: 91

Root Mean Squared Error	8368.012
Mean Absolute Error	5820.615
Mean Abs. Percent Error	23.61902
Theil Inequality Coef.	0.175654
Bias Proportion	0.000424
Variance Proportion	0.381237
Covariance Proportion	0.618338
Theil U2 Coefficient	1.359352
Symmetric MAPE	22.61041

FIGURE 16

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	7.76E+20	Prob. F(44,53)	0.0000
Obs*R-squared	98.00000	Prob. Chi-Square(44)	0.0000
Scaled explained SS	205.2560	Prob. Chi-Square(44)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/25/23 Time: 02:27
Sample: 2/12/2010 12/23/2011
Included observations: 98

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1082.791	1.53E-06	7.06E+08	0.0000
GRADF 01^2	-1.077258	0.155676	-6.919859	0.0000
GRADF 01*GRADF 02	-0.005753	0.036787	-0.156385	0.8763
GRADF 01*GRADF 03	0.001631	0.000589	2.770146	0.0077
GRADF 01*GRADF 04	0.000382	0.000299	1.277305	0.2071
GRADF 01*GRADF 05	54.47546	7.541245	7.223669	0.0000
GRADF 01*GRADF 06	0.000466	0.000175	2.661439	0.0103
GRADF 01*GRADF 07	0.000783	0.000267	2.930684	0.0050
GRADF 01*GRADF 08	0.191594	0.291082	0.658213	0.5133
GRADF 01	0.000394	0.000129	3.062081	0.0034
GRADF 02^2	0.001727	0.002603	0.663495	0.5099
GRADF 02*GRADF 03	-0.000108	5.91E-05	-1.835903	0.0720
GRADF 02*GRADF 04	-4.78E-05	2.41E-05	-1.979092	0.0530
GRADF 02*GRADF 05	0.546997	0.962978	0.568027	0.5724
GRADF 02*GRADF 06	2.30E-05	3.34E-05	0.687332	0.4949
GRADF 02*GRADF 07	2.78E-05	3.86E-05	0.722003	0.4735
GRADF 02*GRADF 08	-0.181573	0.057291	-3.169323	0.0025
GRADF 02	-0.000108	3.04E-05	-3.553896	0.0008
GRADF 03^2	-2.42E-06	6.71E-07	-3.615345	0.0007
GRADF 03*GRADF 04	-8.08E-07	4.35E-07	-1.857511	0.0688
GRADF 03*GRADF 05	-0.033077	0.013860	-2.386499	0.0206
GRADF 03*GRADF 06	-9.77E-07	4.38E-07	-2.230942	0.0299
GRADF 03*GRADF 07	-7.46E-07	4.78E-07	-1.560709	0.1245
GRADF 03*GRADF 08	1.70E-05	0.000422	0.040174	0.9681
GRADF 03	-4.67E-08	2.35E-07	-0.198459	0.8434
GRADF 04^2	-3.87E-07	1.40E-07	-2.757851	0.0080
GRADF 04*GRADF 05	-0.006476	0.006986	-0.926936	0.3582
GRADF 04*GRADF 06	-2.74E-08	1.77E-07	-0.154680	0.8777
GRADF 04*GRADF 07	2.42E-07	1.30E-07	1.864268	0.0678
GRADF 04*GRADF 08	0.000869	0.000279	3.121201	0.0029
GRADF 04	6.93E-07	1.25E-07	5.559997	0.0000
GRADF 05^2	-702.4445	97.03271	-7.239255	0.0000
GRADF 05*GRADF 06	-0.011077	0.004168	-2.657640	0.0104
GRADF 05*GRADF 07	-0.021107	0.007153	-2.950763	0.0047
GRADF 05*GRADF 08	-9.300625	7.514166	-1.237746	0.2213
GRADF 05	-0.013277	0.003489	-3.805870	0.0004
GRADF 06^2	4.31E-08	4.30E-08	1.003260	0.3203
GRADF 06*GRADF 07	2.74E-07	9.37E-08	2.925400	0.0051
GRADF 06*GRADF 08	0.000516	0.000212	2.438384	0.0181
GRADF 06	-2.34E-07	1.47E-07	-1.596407	0.1163
GRADF 07^2	1.64E-07	1.13E-07	1.452369	0.1523
GRADF 07*GRADF 08	-0.000106	0.000386	-0.275086	0.7843
GRADF 07	-2.42E-07	1.16E-07	-2.082090	0.0422
GRADF 08^2	0.832460	0.259040	3.213634	0.0022
GRADF 08	2344874.	0.003348	7.00E+08	0.0000

R-squared	1.000000	Mean dependent var	1082.791
Adjusted R-squared	1.000000	S.D. dependent var	2425.523
S.E. of regression	1.29E-07	Akaike info criterion	-28.58079
Sum squared resid	8.86E-13	Schwarz criterion	-27.39381
Log likelihood	1445.459	Hannan-Quinn criter.	-28.10068
F-statistic	7.76E+20	Durbin-Watson stat	1.606977
Prob(F-statistic)	0.000000		

FIGURE 17

Chow Breakpoint Test: 12/24/2010
Null Hypothesis: No breaks at specified breakpoints

Equation Sample: 2/12/2010 12/23/2011

F-statistic	1.162552	Prob. F(8,82)	0.3318
Log likelihood ratio	11.36252	Prob. Chi-Square(8)	0.1820
Wald Statistic	7.628229	Prob. Chi-Square(8)	0.4706

WARNING: the MA backcasts differ for the original and test equation.
Under the null hypothesis, the impact of this difference vanishes asymptotically.