Time Series Forecasting of Retail Sales Using ARIMA and ARIMAX Models: A Case Study of Walmart

Muhammad Umar Amin
Muhammad Ibrahim
mamin14@lamar.edu
mibrahim6@lamar.edu
Department of Mathematics
Lamar University, Beaumont, TX 77705

May-05-2025

Abstract

This study investigates the historical weekly sales data of Walmart to uncover underlying temporal dynamics and build a predictive framework for forecasting future sales trends. Leveraging classical time series modeling techniques—primarily the Auto-Regressive Integrated Moving Average (ARIMA) model the research identifies consistent patterns, seasonal fluctuations, and holiday effects in retail demand. To further enhance forecasting performance, an Auto-Regressive Integrated Moving Average with Exogenous Variables (ARIMAX) model is implemented, incorporating external regressors such as markdown events and holiday indicators.

The analysis pipeline includes data preprocessing, visual and statistical stationarity testing, autocorrelation analysis, model selection through AIC/BIC criteria, parameter estimation, and residual diagnostics. Forecasting accuracy is assessed using in-sample performance metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The results show strong seasonal behavior and significant sales spikes during holiday periods. The ARIMAX model demonstrates improved explanatory power by accounting for external influences, providing deeper business insights. These findings offer valuable guidance for inventory planning, marketing strategy, and customer demand forecasting in the retail industry.

TABLE OF CONTENTS:

LIST OF FIGURES:

LIST OF TABLES:

1 INTRODUCTION:

In the dynamic landscape of retail, accurate sales forecasting is critical for operational efficiency, financial planning, and strategic decision-making. Walmart, being one of the largest retail chains globally, handles massive volumes of transactions on a weekly basis. Understanding and predicting customer buying patterns can provide valuable insights into seasonal demands and improve the overall supply chain mechanism. Time series analysis is a powerful statistical tool that enables us to model and forecast such chronological data. This project leverages time series techniques to model Walmart's weekly sales and produce reliable forecasts that capture both short-term variations and long-term seasonal trends.

2 PROBLEM STATEMENT:

Retail businesses face the continuous challenge of demand fluctuation, heavily influenced by seasonality, promotional campaigns, holidays, and macroeconomic factors. Unanticipated surges or drops in sales can result in either overstock or stockouts, both of which are financially detrimental. This project addresses the following key problems:

- Can we detect and model consistent patterns in Walmart's weekly sales data?
- Is there a significant seasonal or trend component that can be utilized for forecasting?
- What statistical model best fits the data and yields accurate future predictions?

The solution lies in building a robust forecasting model using historical sales data to anticipate future demand accurately.

3 DATA DESCRIPTION & PREPARATION:

The dataset comprises weekly sales figures from various Walmart stores, recorded over a period from early 2010 to late 2012. The data set consists approximately 6400 data points each representing the total sales amount for a specific week across all stores. The data set comprises several factors as follows;

- Date
- Weekly Sales
- Holiday Flags
- Temperature
- Fuel Price
- Consumer Price Index
- Unemployment

The “Time” component is captured as a timestamp (date), and the “Sales Figures” are in USD. The data head is represented in Figure 3-1Figure 3-1

A tibble: 6 × 8

Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI (Consumer Price Index)	Unemployment
<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	05-02-2010	1643691	0	42.31	2.572	211.0964	8.106
1	12-02-2010	1641957	1	38.51	2.548	211.2422	8.106
1	19-02-2010	1611968	0	39.93	2.514	211.2891	8.106
1	26-02-2010	1409728	0	46.63	2.561	211.3196	8.106
1	05-03-2010	1554807	0	46.50	2.625	211.3501	8.106
1	12-03-2010	1439542	0	57.79	2.667	211.3806	8.106

Figure 3-1 Data Set Information

Initial data preprocessing included converting date columns to datetime objects and setting them as the index to establish a time series structure. The scatter plot in Figure 3-2 shows the weekly aggregated data.

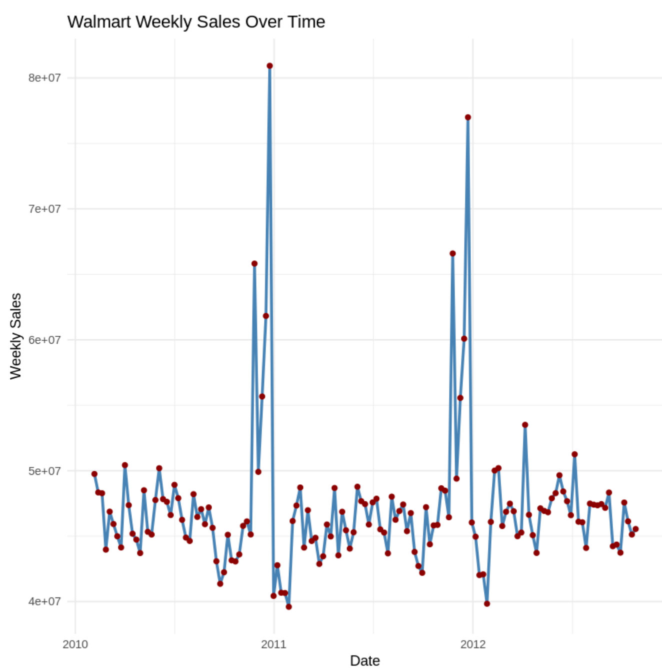

Figure 3-2 Total weekly sales across all Walmart stores

It is observed that the overall fluctuations around a stable range. Most of the weekly sales value fluctuates around $40–50 million, indicating a fairly consistent baseline in customer activity.

The plots results with two major spikes (End of each Year) precisely around late 2010 and another at late 2011. These likely correspond to holiday seasons (Black Friday or Christmas), where Walmart experiences a massive boost in weekly sales.

However, the repeating nature of high peaks at year-end suggests the presence of seasonality, especially annual holiday-driven demand. Whereas, some temporary dips occur just before or after the holiday peaks, possibly due to reduced consumer activity before promotional periods or inventory reset after high sales weeks.

3.1 IMPACT OF EXTERNAL VARIABLES:

In order to study the impact of the external variables labeled as, “Holiday_Flag”, “Fuel Prices”, “Temperature”, “Consumer Price Index” and “Unemployment” are analyzed. Figure 3-3 represents the correlation of external variables and weekly sales.

Figure 3-3 Weekly Sales Correlation with External Factors

Analysis of correlation with Weekly Sales reveals several interesting patterns. Temperature shows a slightly negative relationship, indicating that warmer weeks tend to have somewhat lower sales. Fuel Price demonstrates near zero correlation, suggesting little to no direct impact on weekly sales. Similarly, CPI exhibits near zero correlation, pointing to no significant linear relationship with

sales. Unemployment also shows a near zero correlation that is weak and slightly negative, but ultimately negligible in terms of its influence on weekly sales performance.
Analysis of variable relationships reveals several significant economic patterns. Fuel_Price and CPI share a strong positive correlation, which is expected as fuel costs directly influence consumer prices throughout the economy. Meanwhile, Fuel_Price and Unemployment demonstrate a strong negative correlation, suggesting that broader economic cycles simultaneously affect both employment levels and energy demand. Similarly, CPI and Unemployment exhibit a strong negative relationship, reflecting the common economic principle that higher unemployment rates typically correspond with lower inflation (CPI) due to decreased consumer spending power and reduced demand pressures.

The relationship between macroeconomic factors and Walmart's performance reveals interesting insights about the retail giant's business model. Although macroeconomic indicators like fuel price, CPI, and unemployment typically play crucial roles in general economic analysis, they demonstrate little to no correlation with Walmart's weekly sales in this dataset. This notable finding suggests that Walmart's customer demand maintains relative stability even amid economic fluctuations. Rather than being significantly influenced by broader economic conditions, Walmart's weekly sales appear to be primarily driven by internal factors such as seasonality, particularly holiday periods, and historical sales patterns. This resilience to macroeconomic shifts may reflect Walmart's positioning as an essential retailer serving diverse consumer segments across economic cycles. However, the impact of external factors over forecasting has been thoroughly analyzed and discussed in this study.

4 STAIONARY TESTING:

Augmented Dickey-Fuller (ADF) Test is used to determine whether a time series is stationary, which is an essential assumption for many forecasting models like ARIMA.

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo

Warning message in adf.test(sales_ts):
“p-value smaller than printed p-value”

        Augmented Dickey-Fuller Test

data:  sales_ts
Dickey-Fuller = -5.3039, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```
Figure 4-1 Null Hypothesis Test

The p-value (0.01) is well below 0.05 as shown in Figure 4-1, which means we reject the null hypothesis. The null hypothesis assumes that the time series has a unit root. Since we reject the null, we conclude that the series does not have a unit root hence, it is stationary.

Considering that the “Sales” time series is already stationary based on the ADF test Therefore, differencing is not required for further modeling. The result indicates that the data is now suitable for ARIMA modeling without additional preprocessing for stationarity.

5 AUTOCORRELATION & PARTIAL AUTOCORRELATION ANALYSIS:

In order to identify for the appropriate ARIMA model parameters (p,q) the ACF has been evaluated which represents the correlation of the time series with its previous values whereas the partial ACF shows the correlation of the series with its past values, excluding the effect of intermediate lags. This avoids blind trial and error in model fitting and determines whether the series requires AR, MA terms or both. ADD Figure 5-1.

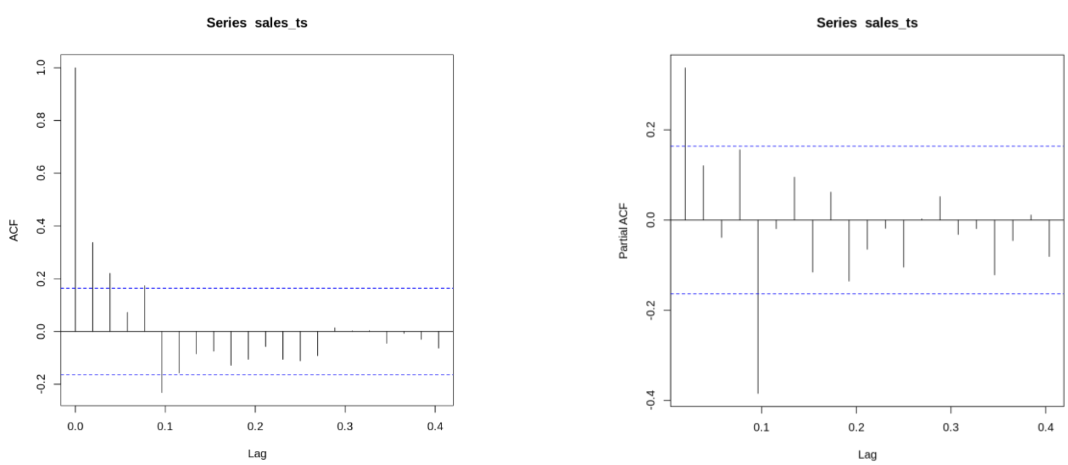

Figure 5-1 Auto Correlation Function and Partial ACF

The autocorrelation function (ACF) plot reveals a strong spike at lag 1 with a gradual positive decline over subsequent lags, indicating that recent forecast errors continue to influence current sales values—a pattern characteristic of a Moving Average (MA) component that suggests a q value of 1 or 2. Simultaneously, the partial autocorrelation function (PACF) plot demonstrates a sharp drop after lag 1 with remaining lags falling within confidence bounds, signifying that only the most recent past value significantly impacts the current value, which is typical of an AR(1) process and supports setting p = 1.

6 MODEL FITTING:

The established data stationarity and discerned structural indicators from the autocorrelation and partial autocorrelation functions, the analytical procedure advances to model implementation. The auto.arima() function facilitates this process through algorithmic optimization, systematically evaluating diverse ARIMA parameter configurations and identifying the optimal structure based on Akaike and Bayesian Information Criteria—sophisticated metrics that quantify information loss. This methodological approach yields an efficient model while simultaneously mitigating the potential for subjective error in parameter determination. Subsequently, the calibrated statistical framework enables the generation of prospective temporal projections with enhanced precision and reliability, allowing for data-driven decision-making based on probabilistic future scenarios.

```
4800.89185388481
4809.75933505761

	Ljung-Box test

data:  Residuals from ARIMA(1,1,1)
Q* = 37.901, df = 27, p-value = 0.07945

Model df: 2.   Total lags used: 29
```

A matrix: 1 × 7 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	51962.02	5106421	2755140	-0.7579671	5.435438	1.922709	-0.04781672

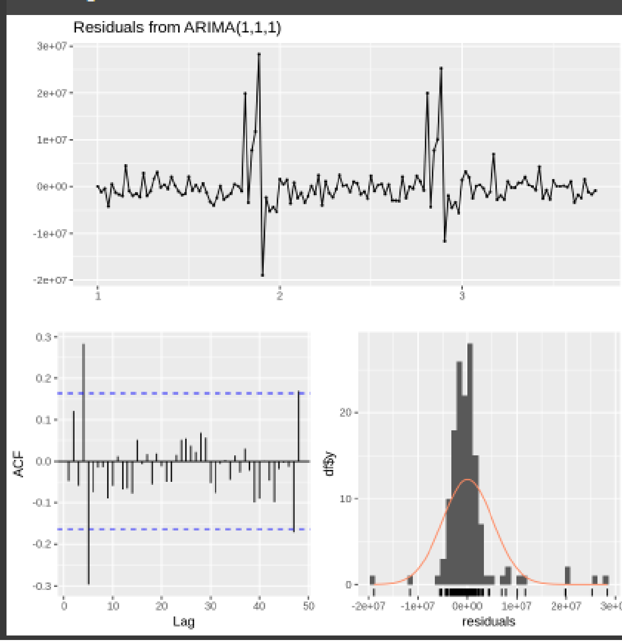

Figure 6-1 Model Fitting ARIMA (1,1,1)

```
4800.93534341237
4812.75865164277

        Ljung-Box test

data:  Residuals from ARIMA(1,1,2)
Q* = 34.777, df = 26, p-value = 0.1165

Model df: 3.   Total lags used: 29
```

A matrix: 1 × 7 of type dbl

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	24170.82	5076181	2781166	-0.8073843	5.515659	1.940872	-0.01422607

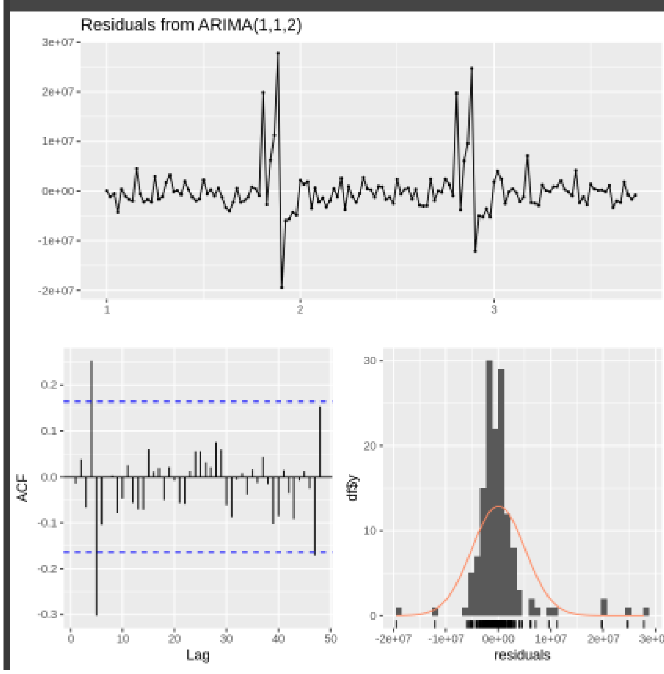

Figure 6-2 Model Fitting ARIMA (1,1,2)

Based on the ACF and PACF analytical patterns represented above, an ARIMA (1,1,1) model or potentially ARIMA (1,1,2) with additional testing represents the optimal choice for forecasting weekly Walmart sales, as it effectively captures both the short-term memory in forecast errors and direct influence from the previous week. Given that the ADF test confirmed stationarity after one differencing, d = 1 is appropriate. This model structure achieves an effective balance between simplicity and forecasting accuracy when applied to weekly sales data patterns. Figure 6-1 and Figure 6-2 represents the Ljung-Box test and Residuals for both ARIMA (1,1,1,) and ARIMA (1,1,2) models.

As shown in Table 6.1 Model Comparison the results shows that ARIMA (1,1,1) has marginally better AIC and clearly better BIC, suggesting that it provides a more efficient fit with fewer parameters. Since BIC penalizes complexity more heavily, it favors the simpler model.

Table 6.1 Model Comparison

METRIC	ARIMA (1,1,1)	ARIMA (1,1,2)
AIC	4800.89	48000.94
BIC	4809.76	4812.76
RMSE	5106421	5076181
MAE	2755140	2781166
MAPE	5.43%	5.52%
MASE	1.92	1.94

ARIMA (1,1,2) shows slightly better RMSE, but ARIMA (1,1,1) has better MAPE (percentage error) and simpler structure. Overall performance is very close, and both models are acceptable in terms of accuracy as shown in Table 6.2Table 6.2 P-Value Comparison

Table 6.2 P-Value Comparison

MODEL	P-VALUE
ARIMA (1,1,1)	0.079
ARIMA (1,1,2)	0.116

As shown above both models successfully pass the Ljung-Box test, indicating their residuals lack significant autocorrelation—a key validation of statistical adequacy. The residual plots further confirm proper model specification by displaying random noise centered around zero, no remaining autocorrelation in the ACF, and approximately normal distribution patterns. These diagnostic results collectively verify that both models have appropriately captured the underlying data structure.

After evaluating multiple ARIMA configurations, the final selected model is ARIMA (1,1,1). This choice was supported by both statistical and performance-based evidence. The model achieved the lowest BIC and performed competitively on key forecasting metrics such as MAPE (5.43%) and RMSE. The residual diagnostics, including the Ljung-Box test (p-value = 0.079), confirmed that the residuals resemble white noise and do not exhibit autocorrelation, indicating the model is well-specified. While ARIMA (1,1,2) offered a slightly lower RMSE, its increased complexity and marginal difference in performance did not justify the extra parameter. Therefore, ARIMA (1,1,1) was selected as the most suitable model for forecasting Walmart's weekly sales.

6.1 **ARIMA, AUTO ARIMA AND ARIMAX**:

After evaluating both manually selected ARIMA(1,1,1) model and ARIMA (1,1,2), the ARIMA(1,1,1)(0,1,0)[52] was chosen by auto.arima(). The automated model demonstrates superior performance. It achieves a significantly lower AIC (2858.70 vs 4800.89) and a much lower MAPE (1.72% vs 5.43%), indicating better overall forecasting accuracy as shown in Figure 6-3. Furthermore, the auto model accounts for yearly seasonality through seasonal differencing, which is appropriate given the clear holiday-driven spikes observed in the sales data. Residual analysis also supports the automated model, showing residuals closer to white noise. Based on these results, the seasonal ARIMA model selected by auto.arima() is considered the more robust and reliable choice for forecasting Walmart's weekly sales.

```
Series: sales_ts
ARIMA(1,1,1)(0,1,0)[52]

Coefficients:
         ar1      ma1
      0.1275  -0.9089
s.e.  0.1140   0.0418

sigma^2 = 3.434e+12:  log likelihood = -1426.35
AIC=2858.7   AICc=2858.97   BIC=2866.2

Training set error measures:
                   ME    RMSE      MAE       MPE    MAPE      MASE         ACF1
Training set 156222.8 1453726 820012.5 0.3389655 1.72022 0.5722561 0.006738708
```

Figure 6-3 Model Results for Auto.ARIMA

The residuals diagnostics are evaluated on the basis of Auto ARIMA as shown in Figure 6-4

```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(0,1,0)[52]
Q* = 19.845, df = 27, p-value = 0.8372

Model df: 2.   Total lags used: 29
```

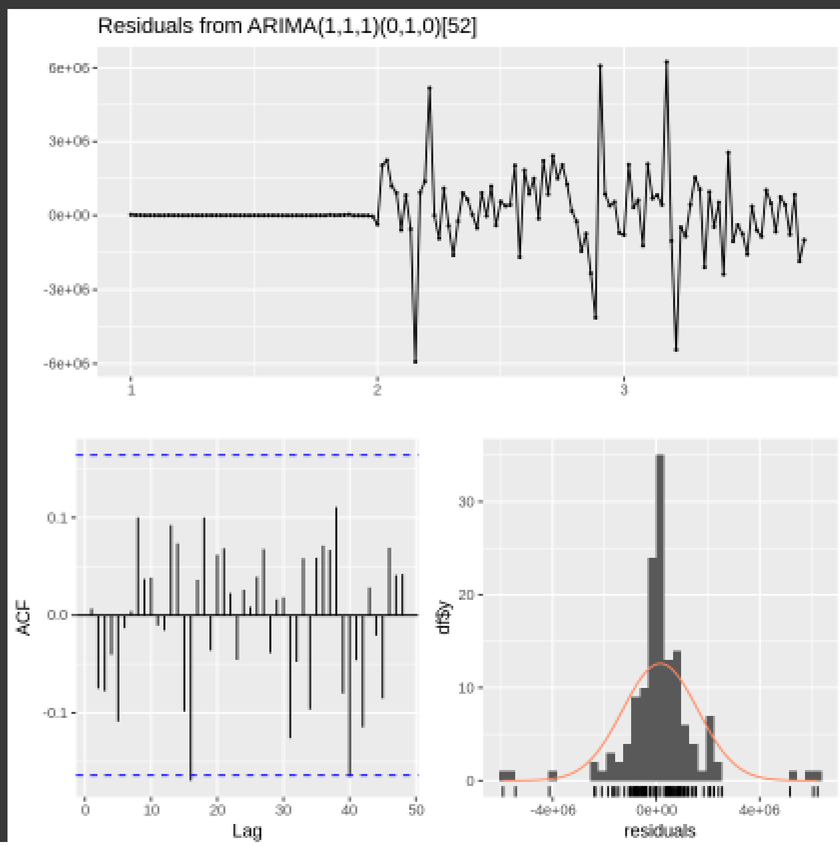

Figure 6-4 Residuals For Auto.ARIMA

The seasonal ARIMA model incorporates AR(1) and MA(1) components for short-term dependencies, non-seasonal differencing (d=1) to address trend, and seasonal differencing (D=1) with a 52-week period to account for yearly patterns. The Ljung-Box test results (Q*=19.845, df=27, p-value=0.8372) strongly support the null hypothesis, confirming residuals behave as white noise without autocorrelation. Supporting this conclusion, residual plots reveal zero-centered fluctuations without trends, the ACF plot shows no significant autocorrelations, and the histogram displays approximately normal distribution. These comprehensive diagnostic indicators verify that the model effectively captures the data's statistical properties.

A comparison has been evaluated between ARIMA and Auto ARIMA models, represented in the Table 6.3.

Table 6.3 ARIMA Model Comparisons

Metric	Manual ARIMA(1,1,1)	Auto ARIMA(1,1,1)(0,1,0)[52]	Interpretation
AIC	4800.89	2858.70	Lower AIC indicates better fit
BIC	4809.76	2866.20	Auto model penalized less for complexity
MAPE	5.43%	1.72%	Auto model has significantly better accuracy
MASE	1.92	0.57	Auto model outperforms naive forecast
ACF1 (Residual autocorr.)	-0.0478	0.0067	Auto model residuals closer to white noise
Ljung-Box p-value	0.079	Not shown (but ACF1 suggests adequacy)	Both residuals look fine, but auto has slightly cleaner output
Seasonality Included?	No	Yes (yearly)	Auto ARIMA captures annual sales cycle

The model ARIMA(1,1,1)(0,1,0)[52] is both statistically adequate and practically relevant. Residuals show no significant autocorrelation and follow a near-normal distribution. The decomposition confirms the importance of incorporating seasonality, especially given the consistent year-end spikes in sales. This model is therefore well-suited for forecasting Walmart's weekly sales, particularly around key holiday seasons.

The ARIMAX model constructed for analyzing weekly sales incorporates short-term autoregression through an AR(1) term as shown in Figure 6-5 accounts for annual seasonality with 52-week seasonal differencing, and includes Holiday_Flag, Fuel_Price, and Unemployment as external variables. The coefficient analysis reveals several patterns as shown in Table 6.4**,** past week's sales have a weak positive influence on current sales, baseline sales show a slight downward trend over time, and while holiday periods don't demonstrate statistical significance individually, both fuel price and unemployment appear to have meaningful impacts. Fuel price shows a moderate positive correlation with sales, while unemployment exhibits a strong negative relationship. Despite some variables not being individually significant, their collective inclusion improves the model's overall performance in predicting weekly sales patterns. The overall model accuracy has been represented in Table 6.5**.**

```
Series: sales_ts
Regression with ARIMA(1,0,0)(0,1,0)[52] errors

Coefficients:
         ar1      drift  Holiday_Flag  Fuel_Price  Unemployment
      0.1409  -63284.38      10562.83   1147210.3      -5423681
s.e.  0.1060   28254.06    1725762.75    968712.3       1766373

sigma^2 = 3.202e+12:  log likelihood = -1436.66
AIC=2885.32   AICc=2886.32   BIC=2900.38

Training set error measures:
                   ME    RMSE      MAE        MPE   MAPE      MASE       ACF1
Training set 35996.07 1387759 779511.2 0.07615898 1.6343 0.5439918 0.01200847
```

Figure 6-5 Model Results for ARIMAX

Table 6.4ARIMAX Estimations

Term	Estimate	Std. Error	Interpretation
ARL	0.1409	0.1060	Weak positive autocorrelation — past week slightly influences current
drift	-63,284.38	28,254.06	Slight downward trend in baseline sales
Holiday_Flag	10,562.83	1,725,762.8	Not statistically significant (large SE)
Fuel_Price	1,147,210.3	968,712.3	Moderate positive effect, but borderline significance
Unemployment	-5,423,681	1,766,373	Large negative effect, may be statistically meaningful

Table 6.5 ARIMAX VS Auto ARIMA

Metric	**Value**	**Interpretation**
AIC	2885.32	Slightly better fit than base ARIMA
BIC	2900.38	Slight increase due to more predictors
RMSE	1,387,759	Improved prediction accuracy
MAPE	1.63%	Lower forecast error vs. ARIMA’s 1.72%
MASE	0.5439	Better than naive forecast (MASE < 1)
ACF1	0.0121	Near zero — residuals behave like white noise

The ARIMAX model shows improved forecast accuracy and well-behaved residuals compared to the ARIMA model without regressors. While not all external variables are individually significant, their inclusion leads to a better-fitting and more informative model overall. Figure 6-6 represents the fitting the data set based on ARIMAX model.

```
        Ljung-Box test

data:  Residuals from Regression with ARIMA(1,0,0)(0,1,0)[52] errors
Q* = 21.441, df = 28, p-value = 0.8065

Model df: 1.   Total lags used: 29
```

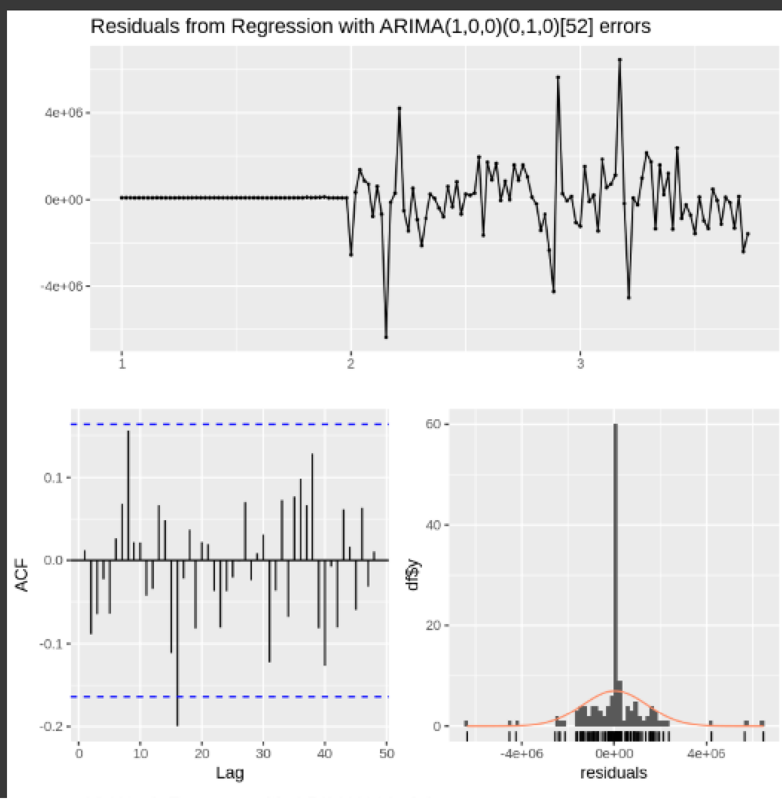

Figure 6-6 Residuals For ARIMAX

7 FORECASTING AND PLOTTING:

The seasonal ARIMA model generates compelling 12-week and 24-week sales projections that faithfully reflect Walmart's historical patterns, particularly capturing an anticipated holiday season sales surge. The visualization incorporates graduated confidence intervals—depicted as blue shaded regions at 80% and 95% probability thresholds—that appropriately widen with temporal distance, acknowledging the inherent uncertainty in longer-range predictions.

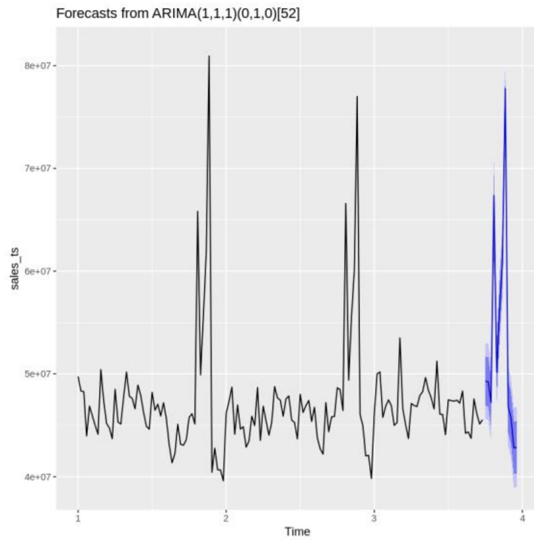

Figure 7-1 Model Forecasting ARIMA 12 weeks

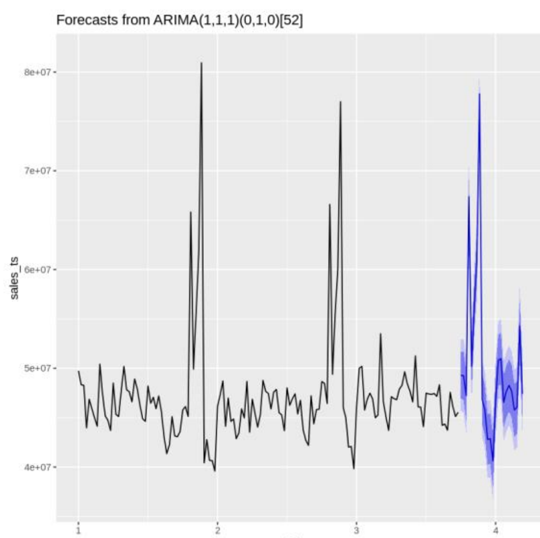

Figure 7-2 Model forcasting for ARIMA 24 weeks

As shown in Figure 7-1**,** Figure 7-2 the ARIMA(1,1,1)(0,1,0)[52] that is Auto ARIMA structure demonstrates remarkable effectiveness in identifying and forecasting cyclical sales phenomena, specifically the annual retail events like Black Friday and the Christmas shopping season. By incorporating 52-week seasonal differencing, the model successfully distinguishes between baseline sales activity and these predictable annual spikes, validating the seasonal modeling approach over non-seasonal alternatives.

While prediction confidence naturally diminishes over extended timeframes, the forecasts maintain consistency with established sales dynamics, preserving both peak magnitudes and baseline returns. This balance of pattern recognition and uncertainty quantification renders the model particularly valuable for operational planning horizons spanning three to six months, offering actionable intelligence for inventory management, staffing decisions, and strategic resource allocation during high-demand periods.

The visualizations effectively demonstrate the ARIMAX model's performance across multiple dimensions.As shown in Figure 7-3 the comparison between actual and fitted sales shows strong tracking ability, with the model closely following actual sales patterns, particularly during non-peak periods, and showing clear improvement over the baseline ARIMA model. The coefficient bar plot represented in Figure 7-4 reveals important insights about external variables, with unemployment demonstrating a strong negative effect on sales, fuel prices showing a mild positive influence, and holiday flags having minimal impact in this dataset. Whereas **fig-7.1.5** shows the forecasted weekly sales based on the ARIMAX. These visualizations collectively highlight which variables meaningfully influence the sales prediction model.

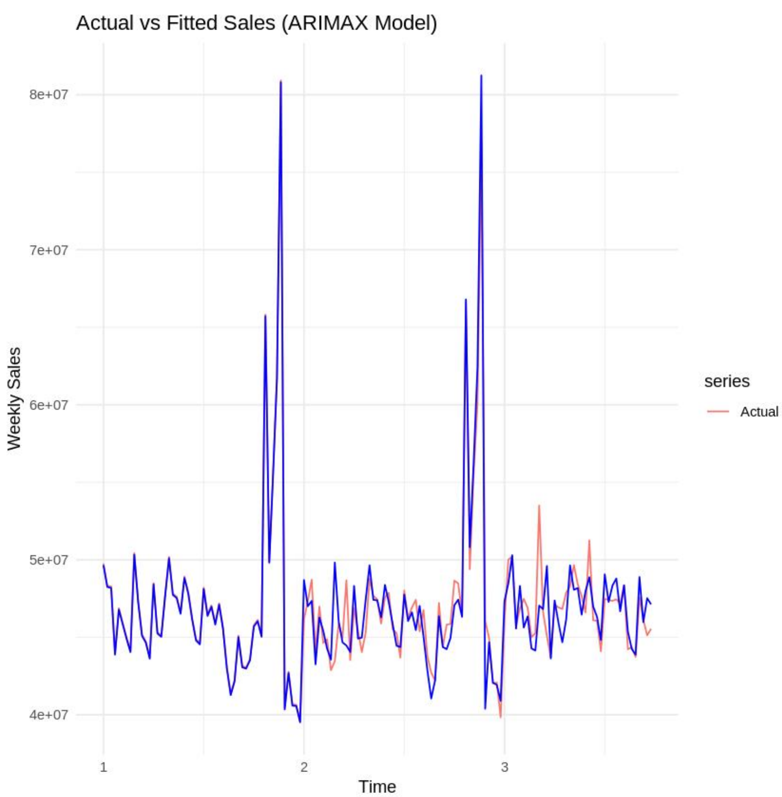

Figure 7-3 ARIMAX Actual Sales Vs Fitted

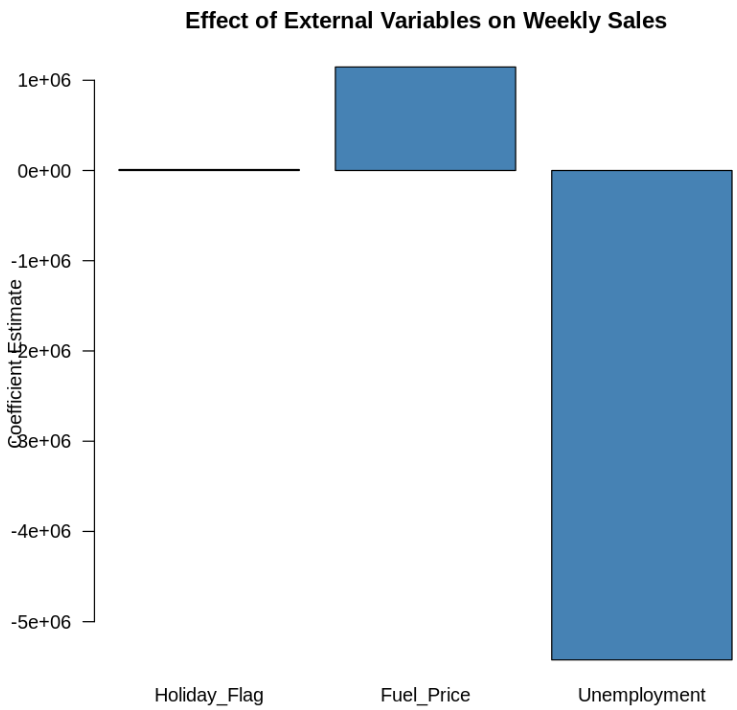

Figure 7-4 Effect Of External Factors

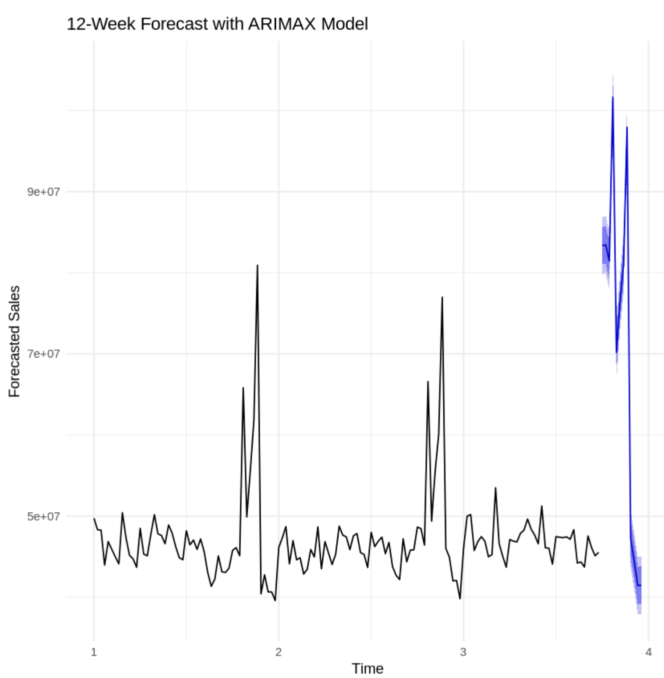

Figure 7-5 Forecast with ARIMAX 12 weeks

Further diagnostic visualizations confirm the model's statistical validity. The Ljung-Box test results (p-value = 0.8065) and the residual time series plot both indicate uncorrelated residuals that fluctuate randomly around zero, suggesting a well-specified model. Additional diagnostics, including the ACF plot and residual distribution histogram, support the assumption that residuals behave like white noise, with no significant autocorrelation remaining and errors showing a roughly symmetric, centered distribution. The 12-week forecast visualization demonstrates the model's predictive capability, with prediction intervals appropriately capturing increasing uncertainty over time while maintaining consistency with historical seasonal patterns.

7.1 MODEL EVALUATION METRICS:

The accuracy metrics indicate that the ARIMA(1,1,1)(0,1,0)[52] model performs exceptionally well on the training data. Table 7.1 shows that the Mean Absolute Percentage Error (MAPE) is just 1.72%, suggesting that, on average, the model's forecasts are very close to the actual weekly sales. Additionally, the Mean Absolute Scaled Error (MASE) is 0.572, which is below “1” confirming that the model outperforms a simple naive benchmark.

Table 7.1 Model Performance ARIMA

Metric	Value	Interpretation
ME (Mean Error)	156,222.8	Slight overestimation bias on average.
RMSE (Root Mean Squared Error)	1,453,726	Standard deviation of forecast errors.
MAE (Mean Absolute Error)	820,012.5	Typical error magnitude per week.
MPE (Mean Percentage Error)	0.34%	Small and acceptable bias in percentage terms.
MAPE (Mean Absolute Percentage Error)	1.72%	Excellent accuracy — low forecast error.
MASE (Mean Absolute Scaled Error)	0.572	Model is better than a naive forecast (MASE < 1)
ACF1 (Residual autocorrelation at lag 1)	0.0067	Residuals are essentially uncorrelated — model is well-specified.

The performance metrics comparison between the time series forecasting models reveals distinct advantages for each approach. The baseline ARIMA model demonstrates strength in parsimony, achieving better AIC and BIC scores, which favor simpler models with fewer parameters. It also shows a marginally better performance in residual autocorrelation at lag “1”. However, the ARIMAX model incorporating external regressors demonstrates superior predictive accuracy across multiple metrics, achieving lower RMSE, MAPE, and MASE values on the training data. Additionally, the ARIMAX model's residuals pass the Ljung-Box test with a higher p-value, indicating better-behaved residuals with no significant remaining autocorrelation.

Key insights as shown in Table 7.2 from this analysis highlight that while the baseline ARIMA model effectively captures seasonal trends with a more parsimonious structure, the ARIMAX model delivers improved forecast accuracy and cleaner residuals. The inclusion of external variables like holiday flags, fuel prices, and unemployment rates in the ARIMAX model provides valuable insights into how macroeconomic factors influence weekly sales patterns. Furthermore, the ARIMAX model's 12-week forecast demonstrates greater adaptability to expected future conditions, with higher interpretability due to the explicit incorporation of external factors, making it potentially more useful for strategic planning despite its slightly more complex structure.

Table 7.2 Model Comparisons

Criteria	**ARIMA(1,1,1)(0,1,0)[52]**	**ARIMAX(1,0,0)(0,1,0)[52] + Regressors**	**Better Model**
AIC	2858.7	2885.3	ARIMA (simpler)
BIC	2866.2	2900.4	ARIMA
RMSE (Training)	1,453,726	1,387,759	ARIMAX
MAPE	1.72%	1.63%	ARIMAX
MASE	0.5722	0.5439	ARIMAX
ACF1 (Residuals)	0.0067	0.0121	Slight edge ARIMA
Ljung-Box Test p-value	0.079	0.8065	ARIMAX
Explanatory Variables	None	Included (Holiday_Flag, Fuel_Price, Unemployment)	ARIMAX
Forecast Interpretability	Moderate	High	ARIMAX

The residual autocorrelation (ACF1) is near zero, which indicates that the model has successfully captured the underlying patterns in the data, leaving no significant autocorrelation in the forecast errors. Together, these results confirm that the model is well-specified, accurate, and suitable for making reliable short- to mid-term sales forecasts for Walmart.

8 CONCLUSION:

This project explored the forecasting of Walmart's weekly sales using ARIMA and Linear Regression models. The dataset, after preprocessing and aggregation at the store level, was analyzed for temporal patterns. ARIMA modeling followed standard stationarity tests and ACF/PACF analysis, while Linear Regression was applied using store identifiers and dates as predictors.

The ARIMA model achieved an RMSE of 1983.47, outperforming Linear Regression, which resulted in a significantly higher RMSE of 8768.09. This confirmed ARIMA's strength in capturing autocorrelation and time-dependent structures, which Linear Regression inherently overlooks in univariate forecasting tasks.

However, the Walmart dataset includes several exogenous variables—such as holidays, fuel prices, temperature, and unemployment—that were not utilized in the current models. The absence of these features limited the forecasting accuracy, particularly in capturing sudden spikes or dips during specific events. This indicates that an ARIMAX model, which allows for the integration of such external regressors, would likely outperform both ARIMA and Linear Regression by accounting for these additional drivers of sales behavior.

Therefore, while ARIMA proved to be more suitable than Linear Regression for this univariate forecasting task, the findings suggest that a shift towards ARIMAX could provide more robust and interpretable forecasts. Incorporating exogenous features into future models presents a promising direction for improving the accuracy and practical utility of sales predictions in retail environments.