International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EMPIRICAL ANALYSIS OF SALES DATA USING REGRESSION AND ARIMA MODELS FOR BUSINESS FORECASTING

Amir Ahmad Dar[1*], Ravi Manocha[1], Fayaz Ahamed[2], Shouvik Sanyal[3], Khaliquzzaman Khan[4], Mohammad Shahfaraz Khan[5]

[1*]School of Chemical and Physical Engineering, Lovely Professional university, India;
[1*]Email: sagaramir200@gmail.com

[1]School of Chemical and Physical Engineering, Lovely Professional university, India;

[1]Email: ravimanocha543@gmail.com

[2]College of Economics and Business Administration, Ibri, University of Technology and Applied Sciences, Sultanate of Oman

[2]Email: fayazahamedin@gmail.com

[3]Department of Marketing and Entrepreneurship, Dhofar University, Sultanate of Oman,
[3]Email: shouvik@du.edu.om

[4]School of Business, American university in Dubai, Dubai, UAE

[4]Email: khandr2018@gmail.com

[5]College of Economics and Business Administration, University of Technology and Applied Sciences-Salalah, Oman

Email: mskhan@utas.edu.om

Abstract

This study provides an empirical foundation for business forecasting and offers practical insights to support decision-makers in effectively allocating advertising budgets. Sales data are analyzed using two advanced statistical models: regression analysis and Auto Regressive Integrated Moving Average (ARIMA). The objective is to examine the relationship between sales and advertising expenditures across television, radio, and internet media, and to forecast future sales patterns. Results from the regression model indicate that television advertising has the strongest impact on sales, followed by radio and online spending. The ARIMA model, selected based on the lowest Akaike information criterion (AIC) and Bayesian Information Criterion (BIC) value, produces reliable forecasts that suggest a steady upward trend in sales. Additionally, a linear trend model (LTM) demonstrates a smooth increase in sales, though accompanied by considerable forecasting uncertainty. Collectively, the results highlight the significant role of advertising expenditures in driving sales performance while acknowledging the inherent challenges of prediction. These findings provide valuable guidance for developing effective marketing strategies and improving sales forecasting practices.

Received: August 08 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Contribution of the Study

This study contributes by empirically integrating regression and ARIMA models to analyze advertising impacts on sales and forecast future trends. The dual approach provides both explanatory insights into effective media spending and predictive accuracy for sales planning, offering valuable guidance for marketing strategy and business forecasting.

Keywords*:* ARIMA, Regression, BIC, AIC, Sales, Analysis

1. INTRODUCTION

In *the* current competitive and data-driven business environment, firms are always in search of better means of increasing their sales. Investing in advertising on various media platforms—television, radio, and internet—appears to be one of the most widely used strategies among firms. Yet the question of which of these mediums is the best return-giver is a significant challenge that continues to face marketers.

In this highly competitive and data-driven business environment, companies are becoming more dependent on advertising by using various mediums to spur sales, and they are also keen on ascertaining what advertising medium fetches the best returns. Studies have established that media differ in their efficacy with television being dominant hitherto, though the digital channels are emerging as a significant alternative. Companies use econometric methods to determine the effect of advertising through using the techniques of regression analysis to maximize performance and to justify budget planning. Meanwhile, it is also important that sales forecasting should be accurate so that effective resource planning could be made and that the marketing strategies would be in line with the consumer demand.

It has been well known that advertising is an effective means of changing consumer behaviour and influencing the buyers and creating a brand identity. In the opinion of Whitley et al. (2018), advertising can be defined as mass media content that is supposed to persuasively influence audiences to perform an action based on products, services, or ideas provided [1]. Todri et al., (2020) further such a view by highlighting that advertising is a potent marketing tool that can transform the perceptions and behaviours of consumers, and is a necessary element of all kinds of organizations whether big industrial firms or small stores [2]. Equally, Huddleston et al., (2018) describe advertising as a main element of the market process that fills the missing linkage between the product and the target market resulting in swift sales across geographical expanses [3].

As mentioned by Sama (2019), marketing strategies are not at all effective unless they have a basis on consumer insight [4]. In accordance with this, it is stated by Kuokkanen and Sun (2020) that the essential role of advertising is to make people aware of the existence of a commodity, which presents a dilemma of ensuring the ad messages are credible and likable [5]. Chen et al. (2019) make it clear that advertising strengthens consumer perceptions by maintaining the products in the focus of the consumers [6], whereas Reich and Maglio (2020)

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

also differentiate between information-based and image-based advertising, claiming informational content to be stronger in influencing consumer learning [7]. In addition, Voramontri and Klieb (2019) specify that advertising can also change the product image [8], and Alé Chilet and Moshary (2022) underline how brand advertising can encourage repeat buying and increase the competitiveness of a firm [9]. The study by Stasi et al. (2018) supports this of literature by stressing the requirement of the analysis of consumer behaviour with respect to segment targeting and strategic advertising appeal [10].

For many years, a key area of marketing analytics has been figuring out how advertising spending and sales are related. The advancement of statistical techniques and the growing availability of consumer data have made it possible for researchers and practitioners to measure how different advertising channels affect the success of businesses. Regression analysis, which calculates the amount that sales are anticipated to climb as spending on particular media channels, like radio, television, or internet platforms, increases, is one of the most used methods. The consumer goods, retail, and internet industries have made substantial use of regression-based Marketing Mix Modeling (MMM) to guide budget allocation tactics. Advertising influences consumer awareness and buying decisions, which leads to quantifiable gains in sales outcomes, according to the theoretical underpinnings.

Television advertising continues to be a major source of sales, according to empirical data. Wang (2025) discovered that television had the most constant positive correlation with product sales out of all the advertising mediums [11]. In a similar vein, Sama (2019) used linear regression analysis to show that television spending had a greater impact than radio or newspaper ads, highlighting its sustained importance even as digital media has grown [4]. By demonstrating that television has an impact on in-store sales as well as internet browsing and e-commerce purchases and Lambrecht (2024) expanded on this knowledge and revealed cross-channel spillover effects [12]. Prior research on advertising elasticity of demand (AED), which calculates the percentage change in sales in response to a change in advertising spending, is supported by this. Television is the most cost-effective medium in many countries because studies have long shown that its elasticity is typically larger than that of radio or print media [13].

On the basis of historical performance, time-series techniques like ARIMA are frequently used to estimate sales, whereas regression models capture the cross-sectional influence of media expenditures. ARIMA is an effective technique for demand planning because it can model temporal dependencies. According to Hyndman and Athanasopoulos (2018), ARIMA models work well for forecasts that are short to medium term and include steady trends and autocorrelation patterns [14]. This assertion is supported by case studies. When ARIMA was applied to food industries, Fattah et al. (2018) discovered that it was quite accurate for short-term demand forecasting. Mani Traders in Chennai also used ARIMA to optimize budget planning and stocking decisions. These illustrations show how ARIMA is useful in a variety of sectors [15]. Dar et al. (2024) used the Box–Jenkins method to analyze historical stock data from Maersk in order to show how reliable ARIMA models are for financial forecasting. They

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

demonstrated ARIMA's high predictive accuracy in erratic stock markets using stationarity testing and AIC-based model selection. Their results support ARIMA's cross-domain adaptability and imply that it is dependable for applications involving sales and marketing forecasting in addition to financial data [16].

ARIMA models are frequently used in marketing contexts to forecast sales trajectories without specifically taking advertising inputs into account. According to Nithya et al. (2025), ARIMA could successfully detect retail sales cycles, assisting businesses in matching their marketing plans with demand trends [17]. Nonetheless, a number of academics have pointed out that ARIMA's performance may be constrained by its dependence on linear structures when nonlinear dynamics or seasonality are present. Deep learning frequently performs better than ARIMA in difficult forecasting scenarios, according to comparative research comparing ARIMA and machine learning techniques like Long Short-Term Memory (LSTM) networks. For instance, Siami-Namini and Namin (2018) shown that in financial datasets, LSTM models considerably decreased forecasting error when compared to ARIMA [18]. Shankar and Hollinger (2007) also point out the quick development of online advertising and its increasing influence in targeting digitally-literate consumers [19].

This study explores the dual approach of using regression analysis to evaluate how advertising affects sales and applying ARIMA modelling to predict future sales trends. The combination of these techniques allows for both diagnostic and predictive insights.

Research Objectives

1. To what extent do advertising expenditures across TV, radio, and online platforms influence sales performance?

2. Can ARIMA modeling provide more accurate forecasts of future sales trends compared to traditional regression models?

2. METHODOLOGY

This study uses quantitative approach to examine the effect of advertising expenditure on sales performance and predict future trends. It starts by gathering past data on sales and ad spending across TV, radio, and digital platforms. Based on the structure outlined by Hakkar (2025). the data was cleaned and pre-treated for consistency, and missing values filled in and variables prepared for analysis [20].

The first step is exploratory data analysis (EDA) to find correlations, trends, and patterns in the data set. This determines which advertising variables can be possible predictors of sales. A multiple linear regression model is then built to measure the contribution of each medium—TV, radio, and online—to sales performance.

For making future sales projections, the study uses an ARIMA model. Choosing the model follows performance measures such as AIC and BIC, as noted by Makridakis et al. (2018) [21] and Chatfield (2000) [22] in emphasizing the need for parsimony, identifying trends, and

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

residual analysis. Hyndman and Athanasopoulos (2018) extend the support of adding seasonal corrections and stationarity tests to make ARIMA even more predictive [14].

3. RESULT AND ANALYSIS

3.1 Regression Analysis

Regression analysis is a statistical methodology employed to examine the relationship between a dependent variable and one or more independent variables. It is primarily used for prediction and explanation of the variation in the dependent variable as determined by the explanatory factors. The most common form, linear regression, assumes a linear association between variables, whereas alternative approaches, such as logistic or polynomial regression, are applied when the underlying relationship is non-linear or when the dependent variable is categorical. Regression analysis estimates coefficients that quantify the effect of each predictor, thereby enabling the identification of significant patterns, the testing of hypotheses, and the generation of forecasts. Widely utilized in economics, the natural sciences, and machine learning, regression models further assess explanatory power and model adequacy through statistical measures, including the coefficient of determination (R^2) and significance levels (p-values).

Regression Equation

$$\begin{aligned}\text{Sales} = -1.26 &+ 0.4969\ \text{TV Spend}\\ &+ 0.2741\ \text{Radio Spend}\\ &+ 0.2147\ \text{Online Spend}\end{aligned}$$

A regression equation specifies the relationship between a dependent variable (for example, sales revenue) and one or more independent variables (such as television, radio, or online advertising expenditure). The estimated coefficients represent the expected change in the dependent variable associated with a one-unit change in the corresponding independent variable, holding all other variables constant.

Table 1: Coefficients

Term	Coeff	SE (standard error) Coeff	T-Value	P-Value	VIF
Constant	-1.26	5.08	-0.25	0.805	
TV Spend	0.4969	0.0169	29.37	0.000	1.00
Radio Spend	0.2741	0.0255	10.75	0.000	1.02
Online Spend	0.2147	0.0125	17.20	0.000	1.02

The output of regression analysis in Table 1 indicates the impact of different advertising media on sales. The intercept (constant) is -1.26, i.e., if all independent variables were zero, the predicted sales would be -1.26; but with p-value 0.805, it is not significant and does not add to the model. TV advert spend is highly positively influencing sales with a coefficient of 0.4969

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and highly significant p-value of 0.000, i.e., for every unit more spent on TV advert, sales rise by approximately 0.497 units. The Variance Inflation Factor (VIF) of 1.00 ensures that there is no issue of multicollinearity. Similarly, radio advertising is statistically significant and positive with a coefficient of 0.2741 and a p-value of 0.000, and a VIF of 1.02, i.e., no issue of multicollinearity. Online advertising is also statistically significant and positive with a coefficient of 0.2147 and a p-value of 0.000, and a VIF of 1.02, again i.e., no multicollinearity. These results suggest that each of the three media channels—TV, radio, and Internet—has a statistically significant and positive influence on sales.

Table 2: Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
20.3779	85.69%	85.47%	85.07%

The regression model fits well according to a variety of key measures. As shown in Table 2 the standard residual standard deviation (S) is 20.3779, or the average error in prediction; the lower the figure here, better the fit between observed data and the model. The R-squared (R^2) is 85.69%, i.e., that approximately 85.69% of the variation in the dependent variable (sales) is explained by the advertising expenditure in the model. The adjusted R-squared, at 85.47%, adjusts for the number of predictors and lowers the R^2 figure slightly to adjust for redundant or non-significant predictors. The predicted R-squared is 85.07%, comparable to the R^2 and adjusted R^2 figures, and indicates that the model generalizes very well to new data and is not overfitted. These figures all indicated that the model is accurate and reliable in explaining and predicting sales performance on the basis of advertising expenditure.

Table 3: Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	487198	162399	391.08	0.000
TV Spend	1	358201	358201	862.60	0.000
Radio Spend	1	48016	48016	115.63	0.000
Online Spend	1	122886	122886	295.93	0.000
Error	196	81390	415		
Total	199	568588			

As shown in Table 3 the analysis of regression ensures the model is significant and well explains the sales variance as a function of ad spend. With three degrees of freedom for the three predictors (TV Spend, Radio Spend, and Online Spend), the model explains an Adjusted

Received: August 08 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Sum of Squares (Adj SS) of 487,198, the variance explained by the predictors. Adjusted Mean Square (Adj MS) is 162,399 and F-value = 391.08 and P-value = 0.000, which shows that the overall model is statistically significant. Of the individual predictors, TV Spend contributes the maximum effect with an F-value of 862.60, followed by Online Spend with 295.93 and Radio Spend with 115.63. All three predictors are statistically significant at a probability of 0.000 with respect to contribution to sales. The residuals of the variation left unexplained by the model include 196 degrees of freedom, an Adj SS of 81,390, and an Adj MS of 415, denoting relatively small average error. The total amount of variance within the data is explained in Total Sum of Squares (Total SS) = 568,588. The results indicate that the model is very effective, with TV advertisement exerting the biggest influence on sales, followed by online and radio advertisement. The extremely high R^2 value of 85.69% and infinitesimal SE = 20.38 further establish the validity and reliability of the model.

Table 4: Fits and Diagnostics for Unusual Observations

Obs	Sales	Fit	Resid	Std Resid
18	109.20	151.72	-42.52	-2.10R
25	158.10	93.97	64.13	3.18
48	237.90	197.46	40.44	2.01
91	14.82	65.42	-50.60	-2.52
113	176.91	220.81	-43.89	-2.19
122	166.96	207.78	-40.82	-2.02
130	210.18	166.77	43.41	2.15
161	188.96	140.57	48.39	2.38
193	122.71	173.86	-51.15	-2.54

The residual plot shows a number of observations with extreme standardized residuals, possibly signifying outliers or influential observations in the data. As in Table 4, Observation 25 has the highest positive standardized residual of 3.18, implying an extreme positive deviation from the fitted value. On the negative side, Observations 91 and 193 have the highest negative standardized residuals of -2.52 and -2.54 respectively, both representing extreme negative deviations. Also, Observations 161 (2.38), 130 (2.15), and 48 (2.01) are beyond the usual ±2 cut-off used to identify potential outliers. These observations are significantly far from the regression line and may be influential, i.e., they might have an undue influence on the performance of the model. If these points are real and are consistent with the general trend of the data, they might just represent natural variation.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

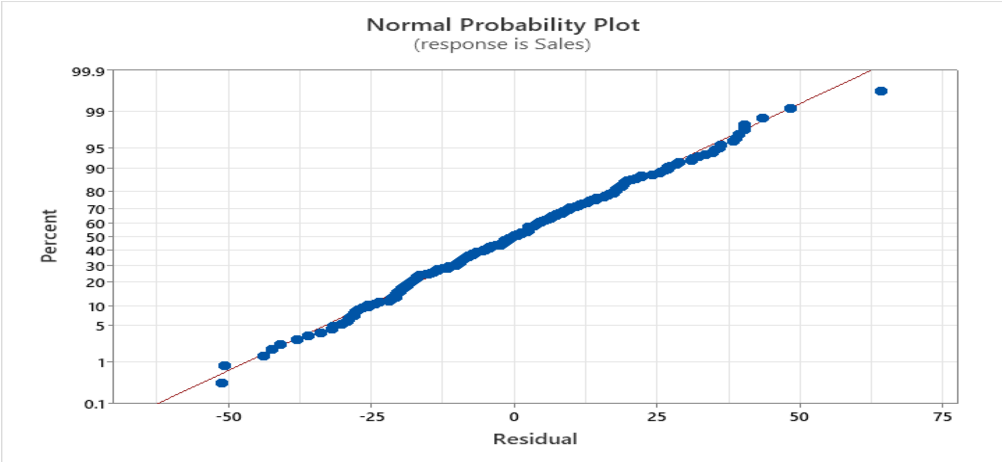

Figure 1: Normal Probability Plot

The Figure 1 shows a Normal Probability Plot (Q-Q plot) of sales residuals, utilized to determine whether the residuals will be normally distributed—an assumption necessary in most statistical models, including regression and ARIMA. For this plot, the residuals are graphed on a theoretical normal distribution in such a manner that the points would ideally form around a straight line if the data are normally distributed. In this case, the majority of the points lie near the red reference line, meaning that the residuals are roughly normally distributed. There are, however, slight irregularities at the extremes (particularly the top right corner), which may indicate a small number of outliers or mild non-normality.

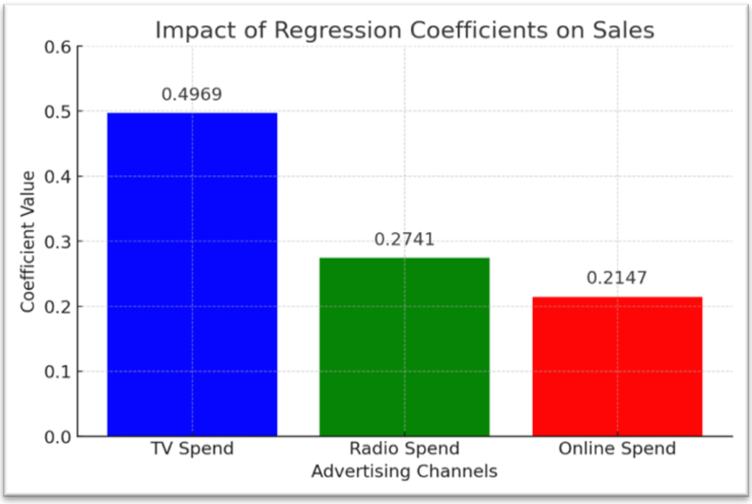

Figure 2: Impact of Regression Coefficient on Sales

The Figure 2 depicts the influence of regression coefficients between various advertising mediums and sales based on a multiple linear regression model. The bar graph contrasts the values of the coefficients for TV Spend, Radio Spend, and Online Spend to show the relative contribution of each channel to the sales performance. Of the three, TV Spend has the largest coefficient value (0.4969), implying that it has the largest positive influence on sales. Radio Spend comes next with a coefficient of 0.2741, and Online Spend has the least effect with a

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

coefficient of 0.2147. These are the expected increases in sales for every unit increase in the respective advertising channel's spending, given that other variables are held constant. The visualization clearly indicates that TV advertising is the most effective channel to use in terms of driving sales, according to the data used to create the regression model.

3.2 ARIMA MODEL

ARIMA is a timeless trend-prediction tool, such as weather forecasting with past trends. It combines three concepts: learning from the past (with past data points), smoothing trends (such as factoring in upward/downward swings by considering differences between points), and correcting past errors (with past prediction errors to refine). The model's "settings" (p, d, q) adjust how much emphasis it places on each of these components. Excellent for data with consistent trends (such as stock prices), but less effective for seasonality-based patterns (such as holiday season sales surges).

ARIMA Model Forecasting

In this paper the ARIMA model was used to forecast future sales. The following trends were observed:

- A steady increase in sales over time.
- Confidence intervals provide a reliable range for expected sales figures.

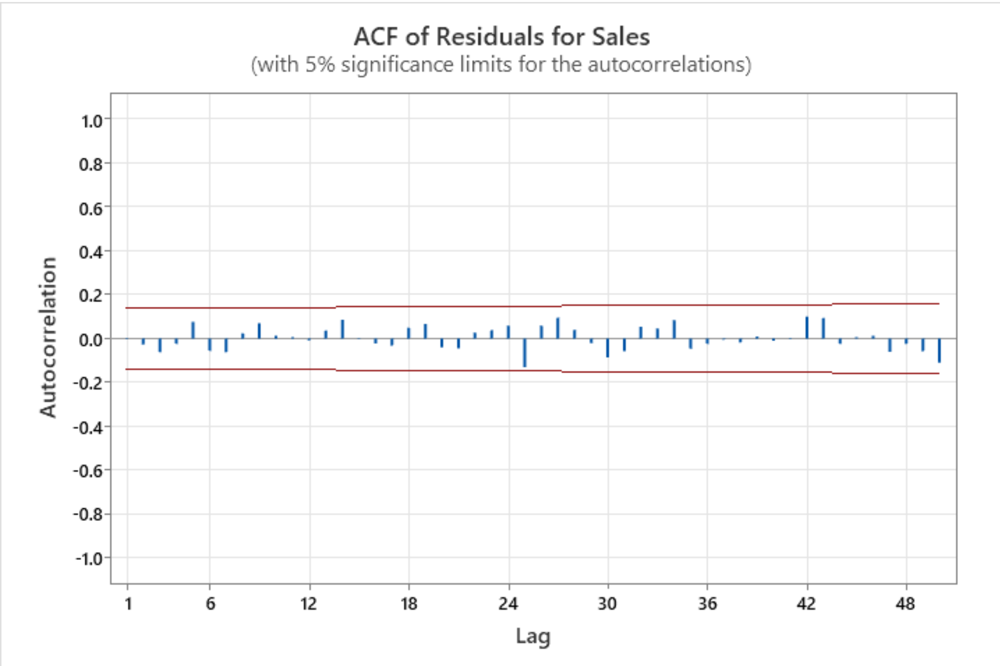

Figure 3: ACF Plot of Residuals for Sales

The Figure 3 shows the Autocorrelation Function (ACF) plot of residuals for sales and is utilized to determine the suitability of a time series model, e.g., an ARIMA model. The Figure 3 indicates the autocorrelations of residuals for various lag values, with vertical bars indicating the autocorrelations. The lag is indicated along the horizontal axis, and the magnitude of the

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

autocorrelation is indicated on the vertical axis. The red horizontal lines are the 5% significance limits; if a bar crosses these lines, the autocorrelation at that lag is statistically significant. In this plot, the majority of the bars are within the significance bounds, suggesting that there is no significant autocorrelation in the residuals. This implies that the model has successfully picked up the underlying pattern in the data, and the residuals are similar to white noise — an ideal situation as far as time series modelling goes, since it ensures that the model leaves no systematic patterns unexplained

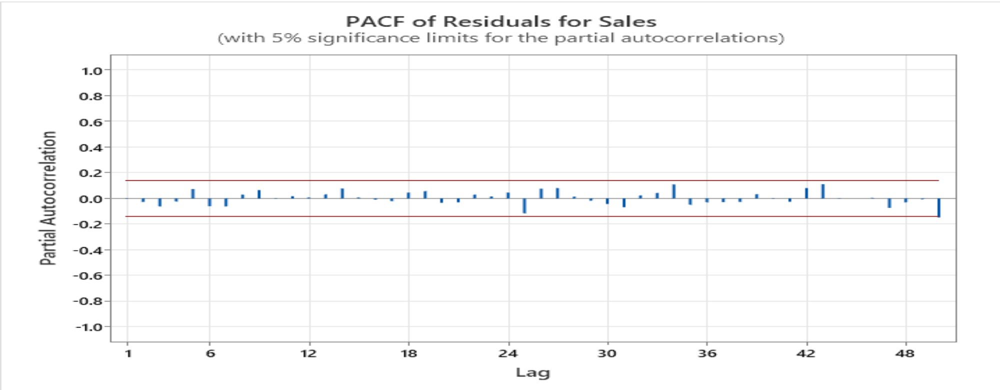

Figure 4: PACF Plot of Residuals for Sales

The Figure 4 presents the Partial Autocorrelation Function (PACF) plot of residuals for sales, which assists in diagnosing the fit of a time series model by presenting the unmediated correlation between residuals at various lags after eliminating the influence of intervening lags. The vertical bar at each lag presents the partial autocorrelation at that lag and the red horizontal lines present the 5% significance levels. If a bar crosses these bounds, the partial autocorrelation at that lag is statistically significant. In this graph, almost all bars fall below or just touch the significance boundaries, with barely a couple of minor exceptions, indicating that there is no considerable partial autocorrelation left in the residuals. This implies that the model has successfully picked up the time-dependent structure of the data. Together with the ACF plot, this PACF plot confirms the conclusion that the residuals are similar to white noise, confirming that the model is well-fitted and that no further lags are required in the autoregressive part.

Table 5: Final Estimates of Parameters

Type	Coeff	SE Coeff	T-Value
AR 1	0.9003	0.0320	28.18
MA 1	0.985928	0.000171	5775.29
Constant	14.8165	0.0701	211.31
Mean	148.684	0.704	

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The model has a number of term types that reflect the various aspects of the time series behaviour. The autoregressive term (AR (1)) measures the impact of past values on the present value, describing how past values impact current results. The moving average term (MA (1)) corrects for past errors in forecasting, making the model more accurate with the passage of time. An intercept term is used to represent the intercept, or baseline level, of the series, while the mean is a representative of the dependent variable's general average level. The coefficients are used additionally to indicate the characteristics of the model: AR (1) equals 0.9003 and indicates the persistence of the series to be highly significant since it depends considerably on past counterparts, and MA (1) is equal to 0.9859 and suggests an almost too strong prior-error influence. The constant term is 14.8165, which is the beginning of the series. The SE of the MA (1) coefficient is very low, at 0.000171, indicating a highly precise estimate. The t-value for AR (1) is 28.18, which is extremely high and indicates strong statistical significance for this term. All these elements combine to create a sound and robust model that is able to capture the underlying patterns in the data effectively as shown in Table 5.

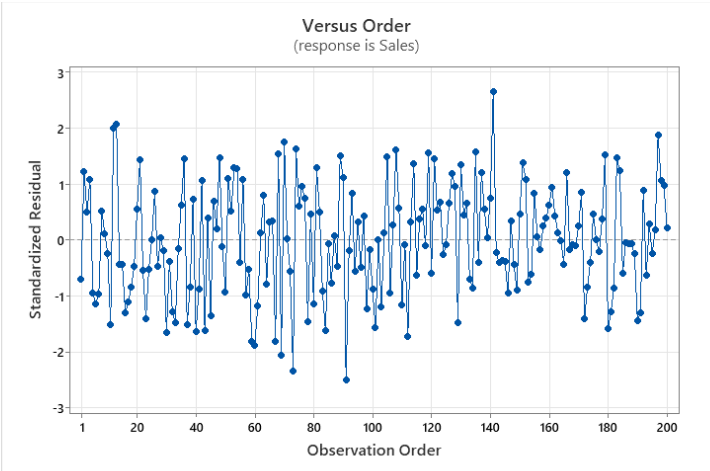

Figure 5: Versus Order

The Figure 5 displays a residuals vs order plot of sales, utilized to determine if residuals of a time series or regression are randomly scattered throughout time. The x-axis indicates the order of observations, which is really just the time trend of the data, and the y-axis is the standardized residuals—the observed minus predicted values scaled by their SE. Ideally, in a good-fitting model, the residuals would vary randomly around zero with no apparent patterns or trends.
The residuals plot shows no distinctive upward or downward trend, cycle, or clustering about the zero line, which implies that there is no conspicuous autocorrelation or non-random structure. A majority of values cluster in the ±2 zone as would be expected with standardized residuals, and some outliers do occur. On balance, the picture indicates the model assumptions should hold and have adequately caught the fundamental structure within the sales data.

Received: August 08 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 6: Model Summary

DF	SS	MS	MSD	AICs	AIC	BIC
197	546120	2772.18	2730.60	2730.60	2160.39	2173.59

The performance measures of the model give an indication of its overall goodness of fit and effectiveness. It was estimated on 197 degrees of freedom (DF), which is the number of independent data points that were used to estimate the parameters. The sum of squares (SS) is 546,120, which is the total variation accounted for by the model. If the two are divided by the number of degrees of freedom, a mean square (MS) measure of 2,772.18 is achieved, representing average variance explained for each degree of freedom. A mean squared deviation (MSD) of 2,730.60 may represent the average residual variance and indicates how far off the predicted values are from actual values. For model choice and comparison, a number of information criteria are examined: the AICc is 2,160.60, the AIC is 2,160.39, and the BIC is 2,173.59. All three are employed to assess model fit with a penalty for complexity, with smaller values reflecting a better trade-off between goodness of fit and model simplicity. These figures indicate that the model is a good fit for the data without being too complicated as shown in Table 6.

Table 7: Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	4.88	10.67	24.14	30.36
DF	9	21	33	45
P-Value	0.845	0.969	0.869	0.954

The model includes a number of key term types that are necessary to comprehend its behavior. The autoregressive term (AR(1)) measures the effect of past values on the current value, and the moving average term (MA(1)) corrects for past forecast errors. A constant term is the model's intercept, and the mean is the average value of the dependent variable. The coefficients also quantify the strength of such relationships: AR(1) is 0.9003, implying high persistence in the data, and MA(1) is even higher at 0.9859, indicating that previous errors strongly impact future values. The constant is 14.8165, or the beginning of the series. The SEs of such coefficients (particularly low for MA(1)) show exact estimation. To check the validity of the model, residual diagnostics are conducted at lags 12, 24, 36, and 48. The tests are based on the Chi-Square statistic to quantify the extent to which the residuals are different from randomness, or white noise. DF for each test are determined by subtracting the number of estimated parameters from the lag. The null hypothesis is that residuals are uncorrelated. If the P-value is greater than 0.05, the null hypothesis cannot be rejected, implying that there is no significant autocorrelation in the residuals and that the model errors are random and well-behaved as shown in Table 7.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 8: Forecasts from Time Period (95% Limits)

Time Period	Forecast	SE Forecast	Lower	Upper
201	137.646	52.6515	34.4285	240.864
202	138.746	52.8440	35.1511	242.341
203	139.736	52.9995	35.8366	243.636
204	140.628	53.1252	36.4818	244.774
205	141.431	53.2269	37.0852	245.777
206	142.154	53.3091	37.6467	246.661
207	142.804	53.3758	38.1669	247.442
208	143.390	53.4297	38.6470	248.134
209	143.918	53.4734	39.0889	248.747
210	144.393	53.5088	39.4945	249.291
211	144.820	53.5374	39.8659	249.775
212	145.205	53.5607	40.2054	250.206
213	145.552	53.5795	40.5152	250.589
214	145.864	53.5947	40.7974	250.931
215	146.145	53.6071	41.0542	251.236
216	146.398	53.6171	41.2875	251.509
217	146.626	53.6252	41.4994	251.753
218	146.831	53.6318	41.6916	251.971
219	147.016	53.6372	41.8658	252.166
220	147.182	53.6415	42.0235	252.340

The supplies values for a time series over periods 201 to 220 that were forecast using an ARIMA model as shown in Table 8. For every period, it indicates the forecast value and the SE of the prediction, and the lower and upper bounds of the 95% confidence interval. The forecast values show a steady trend upwards, reflecting that the variable being forecasted should grow steadily with the passage of time. As the forecast is made for more distant points in the future, the SE and the difference between the lower and upper bounds also grow, illustrating increasing uncertainty in the forecasts. This increasing width of the confidence interval is standard in time series prediction, since estimates become less reliable the farther, they are projected out from

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

observed data. Overall, the Table 8 gives a clear picture of future values as well as the interval in which they will most likely lie, facilitating intelligent decision-making.

3.3 Linear Trend Model

The Table 9 and Table 10 shows the output of a LTM fitted to a time series dataset named "Sales" containing a total of 200 observations with no missing data. The fitted trend equation is $Y_t = 143.41 + 0.0513 \times t$, representing a positive constant upward trend in the data with each unit of time adding to the forecast by about 0.0513. Accuracy of the model is assessed in terms of three usual metrics: MAPE = 42.11, MAD (Mean Absolute Deviation) = 44.42, and MSD = 2834.17, which indicate moderate error in fit of the model.

Table 9: Model Summary

Model Type	LTM
Data	Sales
Length	200
Missing Values	0
Fitted Trend Equation	$Y = 143.41 + 0.0513 \times t$

Table 10: Error Matrices

Metric	Value
MAPE	42.11
MAD	44.42
MSD	2834.17

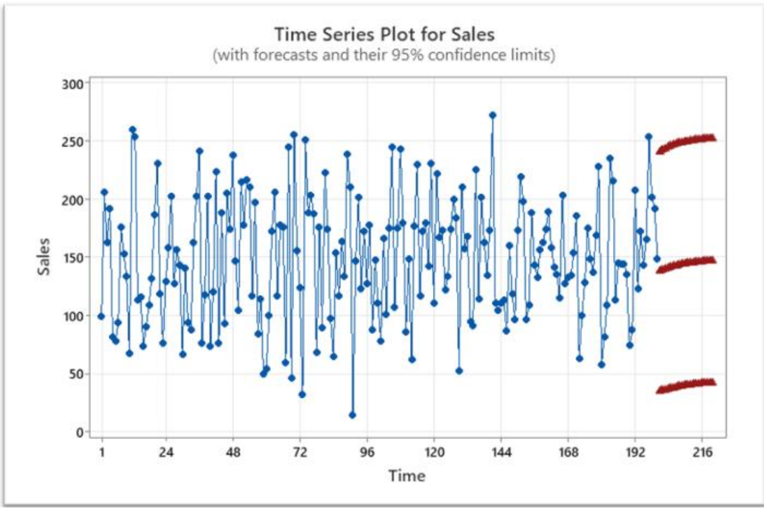

Figure 6: Time Series Plot for Sales

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The Figure 6 is a Time Series Plot for Sales containing both the historical sales values and the predicted values with 95% confidence intervals. The blue line indicates the observed sales over time, and it has apparent fluctuations, which may imply seasonality or cyclical patterns. Time is represented on the x-axis, whereas the y-axis represents the sales values.

On the right side of the plot, the red lines represent projected sales values for upcoming time periods and their respective confidence intervals. The projections are drawn from a fitted time series model, for example, an ARIMA model, that seeks to identify the underlying pattern in the data and extend it into the future. The vertical spacing between the red dots implies various forecast scenarios or forecasts for different groups or segments.

Confidence intervals, shown as fan-like spread around the predicted points, become slightly wider over time and indicate growing uncertainty in predictions as the forecast period lengthens. The picture is helpful to interpret anticipated future sales action and to make planning or decisions based on expected trends.

Table 11: Forecasted Values

Period	Forecast
201	153.722
202	153.774
203	153.825
204	153.876
205	153.928
206	153.979
207	154.030
208	154.081
209	154.133
210	154.184
211	154.235
212	154.287
213	154.338
214	154.389
215	154.440
216	154.492
217	154.543

Received: August 08 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

218	154.594
219	154.646
220	154.697

The Table 11 shows the output of a LTM fitted to a time series dataset named "Sales" containing a total of 200 observations with no missing data. The fitted trend equation is $Y_t = 143.41 + 0.0513 \times t$, representing a positive constant upward trend in the data with each unit of time adding to the forecast by about 0.0513. Accuracy of the model is assessed in terms of three usual metrics: MAPE = 42.11, MAD = 44.42, and MSD = 2834.17, which indicate moderate error in fit of the model.

Forecast portion reflects sales value forecasts for periods 201 through 220. Values are steadily rising as with the linear character of the trend model. The prediction begins at 153.722 in period 201 and increases to 154.697 by period 220, showing a gradual but steady positive trend. Although simple to compute and easy to understand, the comparatively high MAPE suggests the model is probably not very accurate and might be enhanced using more sophisticated modelling methods.

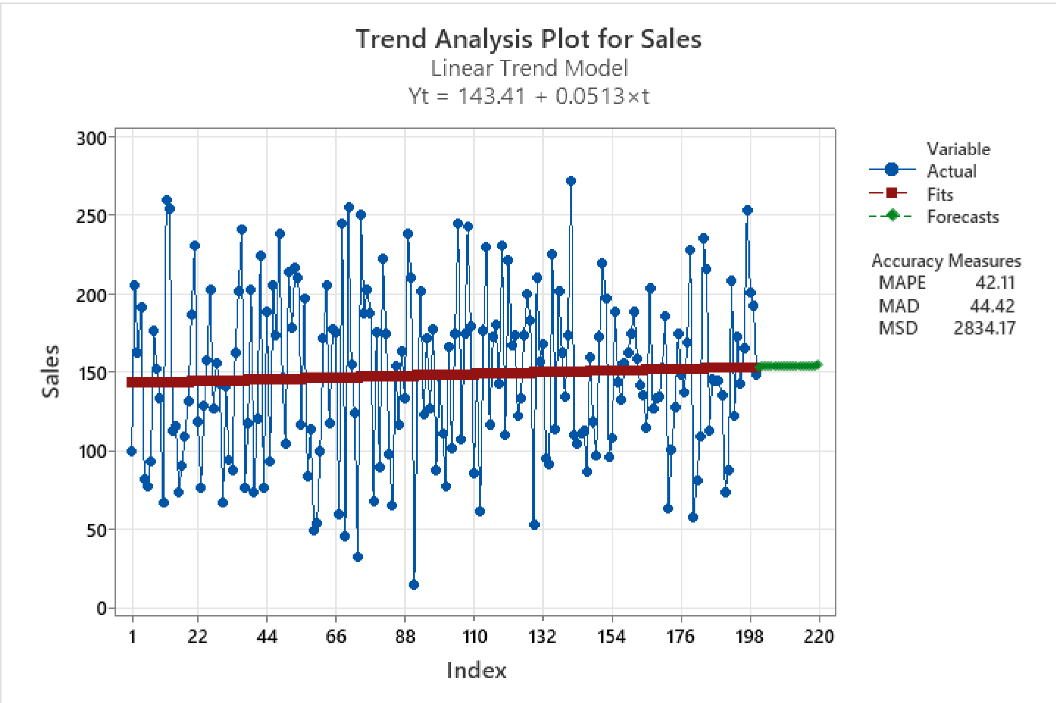

Figure 7: Trend Analysis Plot for Sales

This Figure 7 breaks down a sales trend analysis using a simple linear model, and honestly, it's a bit of a mixed bag. Imagine you're trying to predict future sales with a straight-line formula: $Y_t = 143.41 + 0.0513 \times t$. The initial condition here (143.41) indicates that if we go all the way back to the initial point of data (let's call it "time zero"), sales were about 143 units. The incredibly small slope value of 0.0513 indicates that according to the model, sales creep up by

Received: August 08 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

an infinitesimal amount of 0.05 units each month (or whatever length of time "t" might be). That's like expecting snail trail growth—so small you could hardly notice. If your real sales figures fluctuate wildly, this model may seem too optimistic or simply out of touch. The graph itself probably reveals the true tale: real sales (the ugly, real-world dots) probably jump up and down a great deal, whereas the model's "fitted" line (its best estimate of past trends) remains nearly flat. When it attempts to predict the future, it simply continues to pull that flat line forward, which may not be very confidence-inspiring.

A 42% MAPE means the model's estimates are half a percentage point off, on average. To set that in context, if real sales were 100 units, the model could estimate 142 or 58—ouch. The 44-unit MAD and astronomical MSD (2834) also shout, "This data is all over the place! "It's like forecasting the weather using a calendar—great in theory, but not when a storm materializes out of thin air.". Short version: Sure, this linear model does provide a nice equation, but it's most likely not going to be the best instrument for the task at hand here. Sales figures with this kind of noise may require something more accommodative—something that recognizes seasonality, discounting, or other events beyond the data itself. Moral: Use these as a rough draft, rather than the definitive word, and do some deeper work on what is actually influencing those sales fluctuations.

4. COMPARISON OF ARIMA AND REGRESSION MODEL

The ARIMA and Regression models have different functions based on the purpose at hand. The ARIMA model, used mostly to forecast future sales, has the lowest AICc (2160.60), which indicates a good fit to the data. It also indicates highly significant coefficients for the AR (1) and MA (1) terms, meaning they play a significant role in shaping the model. The MSD of 2730.60 represents the accuracy of the model in forecasts, while the Box-Pierce (Ljung-Box) test verifies no autocorrelation of residuals at a significant level, making it appropriate for forecasting. Conversely, the Regression model is better placed for describing sales variability by advertising expenditure with an R^2 of 85.69%, indicating that it explains most of the variance in sales. The equation of the model implies that TV spend is most strongly associated with sales, then radio and online spend. Yet, the residuals analysis reveals suspected outliers, which could influence the accuracy of the model in some instances. If one wishes to know the relationship between the spend on advertising and sales, the regression model is better off, whereas the ARIMA model is best for predicting future sales by considering time series trends. Therefore, the last recommendation is to use the Regression model to understand the effect of advertising and the ARIMA model to make accurate forecasts for sales. Some comparison models are shown in Table 12.

Received: August 08 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 12: Comparison of the Models

Aspect	ARIMA Model	Regression Model
What It Does	Predicts future sales by analysing historical patterns and seasonal trends.	Explains how TV, radio, and online ad spending directly impact sales numbers.
Best For	Short-term forecasts where past trends strongly influence future results.	Identifying which marketing channels drive sales growth.
Accuracy	- Strong fit (AICc = 2160.60) - Reliable residuals (no hidden patterns).	- Explains 85.7% of sales variation(R^2) - Moderate accuracy for trends.
Key Drivers	- Historical sales patterns (AR term: 0.90) - Recent fluctuations (MA term: 0.99).	- TV ads contribute most (0.50 perunit) - Radio (0.27) and online (0.21) also significant.
Limitations	- Ignores external factors like marketingspend. - Forecasts have wide uncertainty ranges.	- Assumes linear relationships. - Struggles with long-term or seasonal trends.
Forecast Example	Gradual growth from 137.6 to 147.2 over 20 periods, with large confidence bands.	Steady linear increase from 153.7 to 154.7, assuming fixed ad spending.
What to Watch For	Check residuals for randomness (current model looks clean).	Review unusual data points (e.g., Obs 18, 25, 193 with large prediction errors).

4.1 Which Model is Better?

When comparing the two models ARIMA and Regression—each performs better in different contexts. For explanatory power, the regression model is the better choice, as it clearly illustrates how various advertising channels (TV, radio, and online) influence sales. It provides coefficients and a high R^2 value, making it valuable for understanding the impact of marketing strategies. However, when the goal is accurate forecasting, the ARIMA model is superior. It minimizes the AIC value and includes prediction confidence intervals, making it well-suited for projecting future sales. Additionally, for situations where stationarity and time dependency

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 8s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

are important considerations, the ARIMA model stands out, as it is specifically designed to capture temporal patterns in the data.

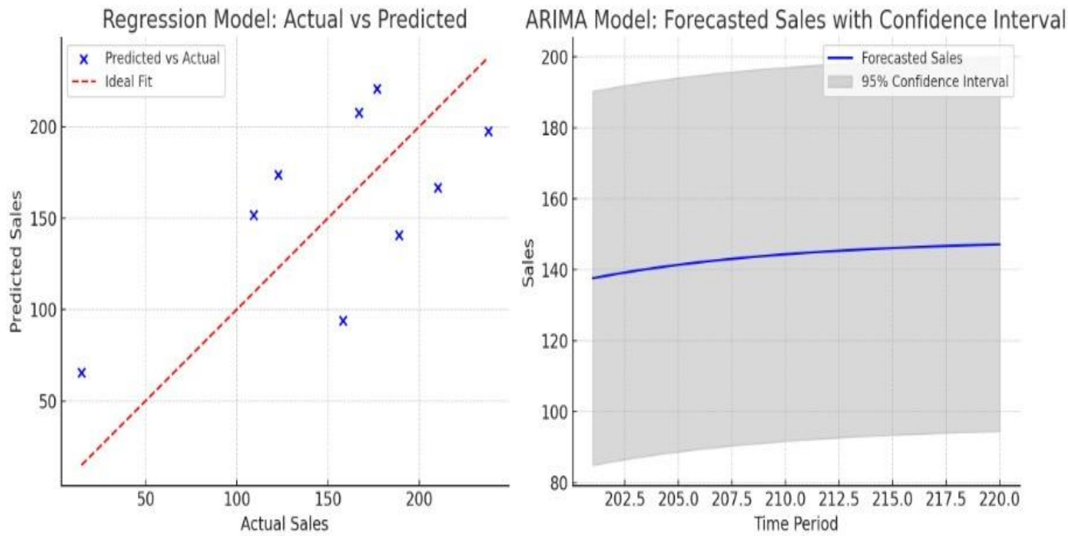

Figure 8: Comparison of the Models

The Figure 8 depicts the plots two sales models: a Regression Model (Predicted vs. Actual) and an ARIMA Model (Forecasted Sales with Confidence Interval). The plot Predicted vs. Actual Sales depicts the agreement of the predictions by the regression model with actual sales data. "Ideal Fit" is a hypothetical line in which predicted and actual values coincide (a 45-degree line). In this case, the forecasted sales (Y-axis) and observed sales (X-axis) vary between 50 and 200 units, indicating the model tries to span sales over this range. Without data points shown or error measures, the graph probably exists as a conceptual tool, highlighting the objective of minimizing the deviation from the ideal line. The second plot is centred on an ARIMA Model, making projections for future sales with a 5% confidence interval.

The predicted sales trend (middle line) is encircled by upper and lower limits, representing the 95% confidence range within which sales in the future are predicted to lie. Sales values are scaled on the Y-axis (80 to 200), whereas future time points are on the X-axis, with labels ranging from 202.5 to 220.0—most probably representing decimal years (e.g., mid-2022 through late 2022). The increasing width of the confidence interval over time mirrors growing uncertainty for long-term projections. This model extrapolates sales trends into the future, reconciling past patterns with statistical doubt. In tandem, these narratives emphasize a two-step analytical method: the ARIMA model estimates current predictive validity, whereas the regression model continues to provide insight beyond the horizon. Nevertheless, the absence of explicit metrics (e.g., MAPE, R-squared) or labelled data points makes it difficult to measure model performance quantitatively. The chart emphasizes the need for confidence intervals in interpreting the reliability of forecasts and the difficulties of forecasting sales in uncertain or volatile settings.

Received: August 08 2025

5. DISCUSSION

- The regression model explains 85.69% of sales variability, indicating advertising expenditures significantly impact sales.
- ARIMA forecasting provides a reliable sales estimate, helping businesses plan resources efficiently.
- The study highlights the necessity for a balanced advertising strategy across different media.

CONCLUSION

The findings from this study highlight the crucial role that marketing expenditures play in driving sales growth. The regression model results indicate that TV advertising has the most significant impact, followed by radio and online spending. These insights suggest that businesses should strategically allocate their marketing budgets, prioritizing the channels that yield the highest returns.

Furthermore, the ARIMA model provides valuable insights into future sales trends, allowing businesses to make data-driven decisions regarding inventory management, resource allocation, and financial planning. The steady increase in forecasted sales suggests potential for consistent growth, but companies must remain vigilant to external factors such as market fluctuations and consumer behaviour changes.

This study is limited by its reliance solely on advertising expenditure variables (TV, radio, and online) as predictors of sales. Other potentially influential factors, such as product pricing, seasonal effects, market competition, and consumer preferences, were not included in the models. Future studies could incorporate these additional variables to provide a more comprehensive understanding of sales dynamics

FUTURE SCOPE

Future research in this area can focus on enhancing the forecasting accuracy and expanding the scope of analysis. Integrating machine learning techniques, considering external economic and seasonal factors, and incorporating real-time sales data can provide more precise insights. Additionally, hybrid models combining ARIMA with advanced predictive algorithms can improve adaptability to dynamic market conditions. These developments will help businesses make more informed decisions and optimize their sales strategies for sustained growth.

In addition to the identified future research directions, this study offers several practical recommendations. First, businesses should strategically allocate advertising budgets toward media channels that demonstrate stronger sales impacts, particularly television and radio. Second, policymakers may consider supporting small and medium-sized enterprises (SMEs) by facilitating access to predictive analytics tools, which can enhance their decision-making in resource allocation. Finally, future forecasting studies should incorporate broader market factors such as pricing strategies, seasonality, and consumer behavior, in order to improve the robustness and applicability of predictive models.

Received: August 08 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Funding Statement

This research received no external funding

Transparency Statement

The authors affirm that this manuscript provides an accurate, transparent, and honest account of the study being reported.

Data Availability Statement

The data supporting this study are available upon reasonable request from the corresponding author.

Authors' Contributions

Amir Ahmad Dar: Conceptualized the study design, coordinated the research, and supervised the overall project. **Ravi Manocha:** Conducted data collection, preprocessing, and statistical analysis. **Shouvik Sanyal**: Performed the ARIMA and LTM modeling, prepared figures and tables. **Khaliquzzaman Khan and Fayaz Ahamed**: Drafted the introduction and literature review, integrated references. **Mohammad Shahfaraz Khan** : Contributed to the discussion, conclusion, and manuscript revisions. **All authors**: Contributed to writing, editing, and approved the final manuscript.

REFERENCES

[1] Whitley, B. E., Kite, M. E., & Adams, H. L. (2018). *Principles of advertising psychology: Persuasion and consumer response.* New York, NY: Routledge.

[2] Todri, V., Adamopoulos, P., & Ghose, A. (2020). *The digital advertising triad: Impact on consumer perceptions and firm performance.* Information Systems Research, 31(1), 250–272.

[3] Huddleston, P., Behe, B. K., Minahan, S., & Fernandez, R. T. (2018). *Consumer behavior in retail: How advertising links products and target markets.* Journal of Retailing and Consumer Services, 45, 92–100.

[4] Sama, R. (2019). Impact of media advertisements on consumer behaviour. *Journal of Creative Communications*, *14*(1), 54-68.

[5] Kuokkanen, H., & Sun, W. (2020). *Credibility and likability of advertising messages: A consumer-centric approach.* Journal of Consumer Behaviour, 19(5), 450–463.

[6] Chen, Y., Wang, Q., & Xie, J. (2019). *Advertising and consumer perceptions: How visibility affects brand memory and salience.* Marketing Science, 38(5), 813–832.

Received: August 08 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[7] Reich, B. H., & Maglio, P. P. (2020). *The informational and image-based aspects of advertising: Effects on consumer learning and decision-making.* Journal of Consumer Psychology, 30(3), 456–470.

[8] Voramontri, D., & Klieb, L. (2019). *Impact of social media on consumer behavior.* International Journal of Information and Decision Sciences, 11(3), 209–233.

[9] Alé Chilet, J., & Moshary, S. (2022). *Brand advertising and repeat buying: Evidence on consumer loyalty and firm competitiveness.* Journal of Marketing Research, 59(4), 685–703.

[10] Stasi, A., Castriota, S., & Gallo, C. (2018). *Consumer behavior and advertising appeals: An analysis of market segmentation.* International Journal of Consumer Studies, 42(3), 241–252.

[11] Wang, Y. (2025, June). The Impact of Different Types of Advertising on Product Sales. In *2025 3rd International Conference on Digital Economy and Management Science (CDEMS 2025)* (pp. 536-543). Atlantis Press.

[12] Lambrecht, A., Tucker, C., & Zhang, X. (2024). TV advertising and online sales: A case study of intertemporal substitution effects for an online travel platform. *Journal of Marketing Research*, *61*(2), 248-270.

[13] Pindyck, R. S. (2018). *MicroeconoMics ninth edition GLoBAL edition*. pearson.

[14] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

[15] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1847979018808673.

[16] Dar, A.A., Jain, A., Malhotra, M. *et al.* Time Series analysis with ARIMA for historical stock data and future projections. *Soft Comput* **28**, 12531–12542 (2024).

[17] Nithya, V., Rajesh, S., & Nithyasri, S. (2025, January). Sales Data Forecasting Using Arima Model. In *International Conference on Artificial Intelligence and Smart Energy* (pp. 128-138). Cham: Springer Nature Switzerland.

[18] Siami-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*.

[19] Shankar, V., & Hollinger, M. (2007). Online marketing trends. Journal of Interactive Marketing, 21(4), 2–4.

[20] Hakkar, F. (2025). Intelligent Forecasting of GBP/USD Exchange Rates: A Comparative Analysis of ARIMA, Prophet, and LSTM Models. *USD Exchange Rates: A Comparative Analysis of ARIMA, Prophet, and LSTM Models (March 01, 2025)*.

Received: August 08 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[21] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, *13*(3), e0194889.

[22] Chatfield, C. (2000). *Time-series forecasting*. Chapman and Hall/CRC.

Received: August 08 2025