Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Fuel Sales Forecasting with SARIMA-GARCH and Rolling Window

Ramneet Singh Chadha[1], Jugesh[2], Shahzadi Parveen[3], Jasmehar Singh[4]

[1]Associate Director, C-DAC, Noida, Uttar Pradesh, India

[2,3]Project Engineer, C-DAC, Noida, Uttar Pradesh, India

[4]Student, Shiv Nadar University, Noida, Uttar Pradesh, India

E-mail: [1]rschadha@cdac.in, [2]prajapatijugesh@gmail.com, [3]parveenshahzadi13@gmail.com, [4]jasmeharsc@gmail.com

Abstract

This research article proposes an innovative strategy to improve long-term forecasting accuracy for gasoline sales in Canada. The SARIMA-GARCH model was used with the rolling window forecasting technique to successfully address varying seasons, changing patterns, and conditional variance on the historical data of gasoline sales in Canada (1993-01-01 to 2015-12-01) with the sample size of 276. The rolling window forecasting technique was used to forecast one-step-ahead value and update the model to fresh observations while minimizing look-back bias and attaining good long-term forecasting accuracy. The findings revealed considerable improvements in forecasting accuracy. The proposed SARIMA-GARCH model with rolling window forecasting produced a RMSE of 151026.28 and a Mean Absolute Percentage Error (MAPE) of 0.0340. This outperformed other baseline models, including simple SARIMA model which had a RMSE of 329,689.88 and a MAPE of 0.0786, and the GARCH model which had a RMSE of 316,168.33 and a MAPE of 0.0685. The data shows that the proposed approach is effective for accurate long-term forecasting of gasoline sales in Canada. The study provides significant data for politicians, industry professionals, and energy investors, assisting them in making informed decisions about resource allocation, strategic planning, and risk management.

Keywords: Time Series Forecasting, SARIMA, Gasoline Prediction, GARCH, ARCH, Hybrid Forecasting Model, SARIMA-GARCH, Expanding Rolling Window Forecasting

DOI: https://doi.org/10.36548/jscp.2023.3.007
Received: 08.09.2023, received in revised form: 03.10.2023, accepted: 16.10.2023, published: 30.10.2023

1. Introduction

In Canada, gasoline sales heavily depend on seasons; in winter, because of snowfall, gasoline sales on the road reduce sharply. However, achieving precise long-term forecasts is a complex task, considering the presence of seasonality, the interplay of various factors, and the inherent volatility in gasoline sales data. Accurate and reliable gasoline sales forecasts are essential for several reasons. Gasoline sales are closely linked to economic activity, consumer behavior, and transportation patterns. Accurate forecasts help policymakers, businesses, and investors anticipate economic fluctuations and make informed decisions. As the energy industry relies heavily on effective resource allocation, precise forecasts enable companies to manage their supply chains efficiently, ensuring that the demands are met without excessive overstocking or understocking. Energy investors and businesses are exposed to various risks, including market volatility and supply disruptions. Reliable forecasts help in identifying potential risks and developing risk mitigation strategies. This research proposes a unique approach that blends the SARIMA-GARCH hybrid model with the rolling forecasting technique to overcome these problems and improve predicting accuracy. The foremost goal of the study is to develop a robust and dependable forecasting model that captures the seasonality and trend found in gasoline sales data. The SARIMA-GARCH model is a powerful combination of two well-known time series forecasting models that can account for both seasonal trends and conditional variance in data. Moreover, the incorporation of the rolling window forecasting technique ensures that the hybrid model adapts to evolving data, consistently enhancing its predictive accuracy over time.

2. Literature Review

Many researchers experimented and implemented the time series regression models, for energy demand forecasting more specifically for gasoline demand and its price prediction. To estimate gasoline prices in China, forecasting based on five prevalent time-series models: ARIMA-GARCH, grey system, support vector machines, exponential smoothing, and neural network was studied. The study compares the performance of these models and finds that for this specific time series, an ARIMA model predicts gasoline prices best for a short time horizon, while support vector regression (SVR) and feedforward neural network (FNN) models outperform others for medium and long forecasting horizons, respectively (Xu et al., 2018). In Forecasting of annual demand for gasoline and the decline in fuel production in Indonesia using time series data from 2017 to 2019, the study compares the Holt-Winters additive model to the autoregressive integrated moving average model. Because the Holt-Winters model produces more reliable findings, it is used to forecast total gasoline consumption from 2020 to 2022. Furthermore, the study combines the Holt-Winters model with a neural network to estimate gasoline 92 demand, resulting in fewer forecasting errors when compared to utilizing the Holt-Winters technique alone. The study suggests that total gasoline consumption is expected to rise, but demand for specific gasoline components shows disparities. Demand for gasoline 90 is likely to rise, while demand for gasoline 92 and gasoline 88 is expected to fall (Mardiana et al., 2020). The research article "Data-driven short-term natural gas demand forecasting with machine learning approaches" discusses the necessity of precise forecasting for natural gas demand in natural gas supply system planning and operation. (Sharma et al., 2021). The study investigates alternative machine learning algorithms for forecasting natural gas demand and tests the models with real data from the 2018 nPower forecasting competition. The study offers artificial neural network (ANN)-based models showing higher accuracy than gradient boosting (GB)-based models, and also three hybrid models which combine estimates from individual models and outperform solo models in terms of Mean Absolute Percentage Error (MAPE) by roughly 15%. The study underlines the need for precise demand forecasting for local distribution companies (LDCs) and the possible economic benefits of the suggested hybrid forecasting model in real-world energy operations. It also emphasizes the significance of accurate forecasting in balancing natural gas supply and demand and decreasing penalties associated with prediction errors[6]. Some studies show that the hybrid model and technique are best for time series, to project one-day-ahead daily air quality index (AQI) levels in 16 Taiwanese cities using a rolling window approach with a hold-out set of 375 days, or just over a year. This study estimates forecast performance for air quality levels using four models: ARX-GARCH (autoregressive models with exogenous variables) model; AR logistic regression; SVM method; and neural network autoregression with exogenous variables NNARX model. The results indicate that SVM and autoregressive multinomial logistic regression are the most effective models for AQI-level predictions in terms of accuracy rates. The study concludes that the proposed methodology may accurately anticipate one-day-ahead AQI levels, allowing authorities to adopt appropriate AQI-level prediction activities (Chen & Chiu, 2021)[7]. The SARIMAX-GARCH model predicts Indian Machinery and Transport Equipment export values by taking into account both the internal dynamics of time series data and the influence of external events. The results show that the Garch-Sarimax model outperforms the standard SARIMAX model in terms of RMSE and MAPE score, giving it a better choice for time series prediction (Ramneet Singh Chadha et al., 2023). Dileep Kumar Shetty and Sumithra (2018) has proposed a hybrid model “SARIMA-GARCH” for forecasting the Indian gold price. The study analyses the goodness of fit using the Akaike information criteria (AIC) and evaluates forecasting capacity using measures such as MAE, RMSE, and MAPE. According to the statistics, the SARIMA-GARCH model with a 4.14353 MAPE score is superior for forecasting the Indian gold price[13]. Overall, the literature review underlines the importance of selecting effective forecasting methodologies that are suited to specific datasets and research aims. ARIMA, SARIMA, GARCH, one-step-ahead, rolling forecasting, and hybrid models all contribute to the advancement of time series forecasting and have important implications for industries, governments, and enterprises looking for credible predictions for planning and operational purposes. This study aims to fill gaps in the long-term forecasting of gasoline sales in Canada. This study aims to construct an enhanced forecasting framework that ideally handles seasonality, captures trend dynamics, and adjusts to new observations by merging the SARIMA-GARCH model with the rolling window forecasting technique.

3. Methodology

This study follows the Machine learning approach depicted in Figure (1):

Figure 1. General Machine Learning Approach

The machine learning methodology was used in this research, which includes data preprocessing, where historical gasoline sales data was transformed. After that, the SARIMA-GARCH model was used and tweaked using the AdFuller test, ACF, PACF, and AIC score. To test the model's predictive capabilities, performance metrics MAPE and RMSE were used.

3.1 Data Collection and Data Analysis

The data set being used in this research is illustrated in Table (1) below

Table 1. Monthly Sales of Gasoline used for Road Motor Vehicles

Publisher - Current Organization Name	**Statistics Canada**
Series Issue ID	"Table 23100080; Formerly CANSIM Table 405-0003"
Record ID	e75562a3-3dea-471b-b143-a8b918a49a2c
URL	"https://open.canada.ca/data/en/dataset/e75562a3-3dea-471b-b143-a8b918a49a2c"
Data Size	276
Range	1993-01-01 to 2015-12-01
Frequency	month
Unit	liters
Type of Fuel Sales	'Net sales of gasoline
Geography	Canada

Table 2. Dataset Description

Description	**Value**
count	276
mean	3151497
std	304501
min	2373299
25%	2955517
50%	3160843
75%	3342615
max	4154359

Data Description Table (2) shows the data sample size is 276 of which 95% were used for the training part and 5% for the testing part, values ranging between 2373299 to 4154359 it may vary in the future and have no outliers and null values in this dataset. As shown in Figure (2), Gasoline sale data includes trend, seasonality, and residual, making forecasting extremely difficult.

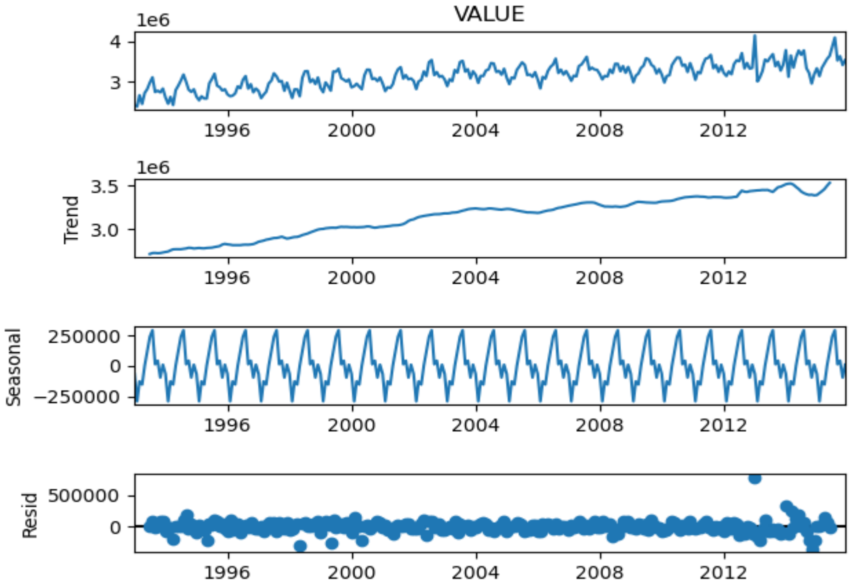

Figure 2. Sale of Gasoline in Canada (1993-2015) Time Series Decomposition

3.2 Data Transformation

Gasoline columns before data transformation.

Index “['REF_DATE', 'GEO', 'DGUID', 'Type of fuel sales', 'UOM', 'UOM_ID', 'SCALAR_FACTOR', 'SCALAR_ID', 'VECTOR', 'COORDINATE', 'VALUE', 'STATUS', 'SYMBOL', 'TERMINATED', 'DECIMALS']”

After filtering rows based on ['Type of fuel sales'=='Net sales of gasoline' & 'GEO'=='Canada'], this study selected the column ['REF_DATE',' VALUE'] using Pandas functions. While importing this dataset, column 'REF_DATE' was converted as date-time index. After the data Transformation column remains ['REF_DATE', 'VALUE'] .

3.3 SARIMA

An ARIMA (Auto Regressive Integrated Moving Average) model is a general form of SARIMA model that combines a moving average, differencing, and autoregression model. This model is defined as:

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\,\varepsilon_t$$

Where, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ is the autoregressive operator of order p;

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ is moving average operator of order q; ε_t is the error term at time t. $(1-B)^d$ is the d^{th} difference; and B is backward shift operator ;

where (ϕ_1, ϕ_2, ..., ϕ_p) are the autoregressive coefficients of the model, each ϕ(phi) term corresponds to the impact of the lagged values (up to lag *p*) of the differenced series on the current data point at time *t*. (θ_1, θ_2, ..., θ_q) are the moving average coefficients of the model each θ (theta) term corresponds to the impact of the past forecast errors (up to lag *q*) on the current data point at time *t*.

By integrating additional seasonal variables in the ARIMA models, a seasonal ARIMA model is constructed. It's written like this: ARIMA *(p, d, q) (P, D, Q) m*

Where, (*p*,*d*,*q*) are non-seasonal parts of the model, *p* is the autoregressive order, *d* is differencing and *q* is the moving average order, *m* is observation number/year, *(P, D, Q)m* = models seasonal parts, whereas *P* is the seasonal autoregressive order, *D* is seasonal differencing and *Q* is the seasonal moving average order.[9]

3.4 GARCH

The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model forecasts volatility in time series data. The GARCH model is a modification of the ARCH (Autoregressive Conditional Heteroskedasticity) model that allows for greater flexibility in modeling conditional variance. (Bollerslev, 1986). The GARCH model is typically written in the following form:

$$\sigma^2_t = \omega + \alpha_1\varepsilon^2_{t-1} + ... + \alpha_p\varepsilon^2_{t-p} + \beta_1\sigma^2_{t-1} + ... + \beta_q\sigma^2_{t-q}$$

Where, σ^2_t reflects the conditional variance at time t, which is an estimate of the time series' variance at that moment based on past data, ω is a constant term that represents the model's intercept, $(\alpha_1, \ldots, \alpha_p)$ are the coefficients of the lagged squared residuals ($\varepsilon^2_{t-1}, \ldots, \varepsilon^2_{t-p}$). These coefficients determine the persistence of shocks in the volatility by capturing the impact of past squared residuals on the current conditional variance. $(\beta_1, \ldots, \beta_q)$ are the coefficients of the lagged conditional variances $\sigma^2_{t-1}, \ldots, \sigma^2_{t-q}$ .These coefficients reflect the influence of previous conditional variances on the current conditional variance. The larger the β, the longer the duration of the impact. ε_t represents the error term (residual) at time t, It is intended to have a zero mean and constant variance, and to be independently and identically distributed (i.i.d.). [10][11][12]

3.5 SARIMA-GARCH

By combining a seasonal ARIMA model with a GARCH model, the forecasting accuracy can be boosted by taking both the mean and volatility of the data into account. The GARCH model computes the conditional variance, whereas the SARIMA model computes the time series forecast. [8] The developed model is stated in the equation shown below:

$$\varphi_p(B)\phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\,\Theta_Q(B^s)\in_t ,$$

Where, $\in_t = \sigma_t z_t$, $\sigma^2_t = \omega + \alpha_1\varepsilon^2_{t-1} + \ldots + \alpha_p\varepsilon^2_{t-p} + \beta_1\sigma^2_{t-1} + \ldots + \beta_q\sigma^2_{t-q}$

Where y_t represents the time series; $\phi_P(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \ldots - \phi_P B^{PS}$ is seasonal autoregressive part; $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots - \Theta_Q B^{QS}$ is seasonal moving average part, S is the seasonal period; D is the seasonal difference; (p, q) is the order of GARCH process; ω is a constant term, α and β are coefficients of GARCH model; ε_t is the error term; σ_t^2 is the conditional variance of $\in_t$; $\{ z_t \}$ is the identical and independently distributed random variables that has zero mean and constant variance. The final prediction is obtained by merging the results of the seasonal ARIMA and GARCH models.[12][13]

3.6 Parameter Tuning

Hyperparameter optimization is a very essential part of every research for the high accuracy of the model, this research uses the Adfuller test, ACF/PACF plots, and AIC score for selecting parameters of the SARIMA and GARCH model. The ADF test, also known as the Augmented Dickey-Fuller or AdFuller test, is a statistical test that determines the appropriate value of the ARIMA model's differencing parameter (d), which indicates the number of times the time series must be differenced in order to stay stable. The null hypothesis in the ADF test assumes that there is a unit root (non-stationary data), while the alternative hypothesis assumes that there is no unit root (stationary data). If the p-value is lower than 0.05 it suggests the time series is stationary otherwise it is non-stationary. This study discovered that the smallest value of *d* that makes the Gasoline time series stationary was 1 by conducting the ADF test on it [15]. A time plot normally cannot tell what values of *p* and *q* of SARIMA model are acceptable for the time-series data. However, the "ACF(Auto-Correlation function)" plot and its closely related "PACF (Partial Auto-Correlation function)" plot were used to find optimal values for *p* and *q*. The "ACF plot" measures the direct and indirect autocorrelations between two variables "y_t and y_{t-k}"for diverse *k values*. While the PACF plot estimates the direct connection across *two* variables of k values when the effects in between the lags are removed $1,2,3,\ldots,k-1$.[9]

This study focuses on these SARIMA(p,d,q) orders based on the AdFuller test, ACF/PACF plots, and pmdarima (pmdarima is a Python library that enables an automated implementation of the ARIMA modeling process, including model selection and hyperparameter tuning) the selected orders are show in Table (3) below

Table 3. Selected Orders

Order(p, d, q)	AIC score
(0,1,0)	7034.38
(1,1,0)	7014.06
(0,1,2)	7016.35
(2,1,0)	7015.55

AIC (Akaike's Information Criterion) is beneficial in picking predictors for regression, and can also be used to determine the order of the GARCH and ARIMA model. It defines as:

$$AIC=-2\log(L)+2k$$

where k is the number of parameters (including σ^2) = $p+q+c+1$ where $c = 1$ if there is a constant term otherwise $c = 0$, and L is the likelihood of the data.

It is important to note that these information criteria are only useful for establishing the values of *p* and *q*, not for determining the differencing *(d)* of a model. Because differencing modifies the data on which the likelihood is estimated, the AIC values of models with different orders of differencing are incomparable. So AdFuller test used to select *d*, and then we may use the AIC to select *p* and *q* [9]. For the GARCH model, this study try 324 combinations of parameters and calculated the AIC score, in which (mean ='Constant', p=1,o=1,q=2, dist= 'normal') was the best parameter with an AIC score of 7459.20.

3.7 Approach

This study combines an expanding rolling window forecasting technique with a hybrid SARIMA-GARCH model to anticipate one-step-ahead gasoline sales that will be added to historical data.

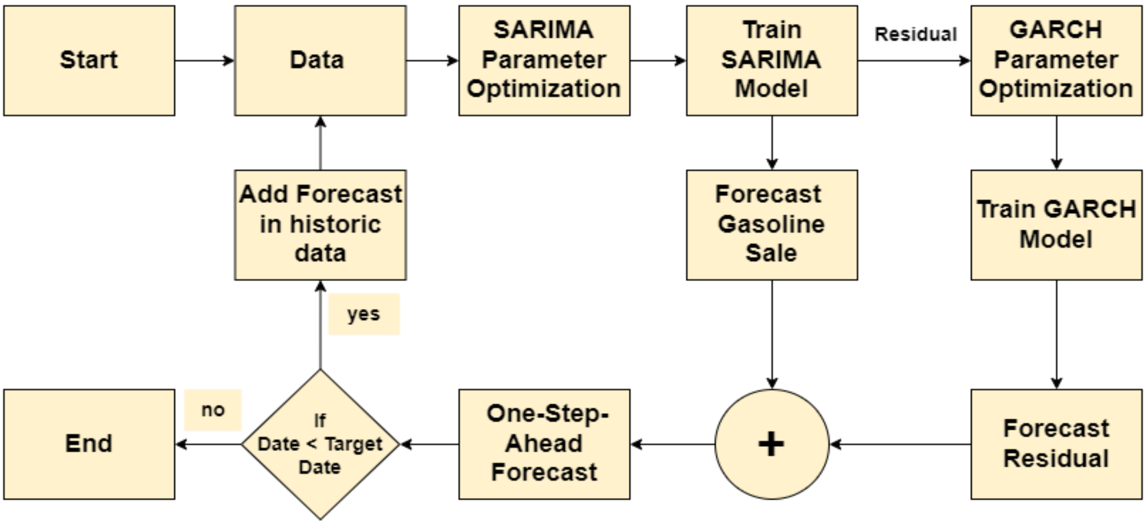

Figure 3. Flowchart of the Procedure for Hybridization of SARIMA and GARCH Model.

As depicted in Figure (3), this study initially collected and preprocessed gasoline sales data to meet the prerequisites of models. Subsequently, it employed the SARIMA model, which was trained with selected optimal parameters, to forecast gasoline sales in Canada. Then model's residuals were forwarded to the GARCH model, which was trained on these residuals to predict the fluctuations in gasoline sales. By combining the forecasts of SARIMA model for gasoline sales as well as the GARCH model for sales volatility, it obtained a one-step-ahead actual forecast of gasoline sales. This forecast was iteratively incorporated into the historical data, and the entire process was repeated until the target date was reached. This technique is called expanding window forecasting. This comprehensive approach ensures that the forecasting model remains adaptive and accurate over time, effectively addressing the complex dynamics of gasoline sales data**.**

Expanding Window Forecast: It Adds fresh data points to the sample regularly. It eliminates lookback bias and adapts forecasts to fresh observations, making the model less susceptible to overfitting and very accurate for long-term forecasting.[12] The time series expanding window forecasting is depicted in Figure (4) below.

Figure 4. Expanding Window Forecast.[12]

3.8 Model Evaluation

The MSE (mean square error) is the squared error of mean/average difference across the actual and the estimated values. The RMSE (Root Mean Square Error) is the square root of the MSE, it converts back the value into the same units as the original data. MSE is defined as:

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2$$

The MAPE (mean absolute percentage error), is designed to be sensitive to relative errors, it is independent of the global scaling of the target data. MAPE is defined as:

$$\mathrm{MAPE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \frac{|y_i - \hat{y}_i|}{max(\epsilon, |y_i|)}$$

Where $\hat{y}_i$ is the predicted value at ith value and y_i is the corresponding true value, n_{samples} is the number of samples, and ϵ is an arbitrarily small yet strictly positive number to avoid undefined results when y is zero.[14] Forecast accuracy scale is illustrated in Table (4)

Table 4. Forecast Accuracy Scale [3]

MAPE%	Evaluation
<10%	High accuracy
10% <MAPE <20%	Good forecast
20% <MAPE <50%	Reasonable forecast
>50%	Inaccurate forecast

4. Experiment and Result

Table 5. Models and RMSE/MAPE Score

Model	RMSE score	MAPE score(0 to 1)
Sarimax ((2, 1, 0),(1, 0, 1, 12))	329689.88	0.0786
Sarimax ((0, 1, 0),(1, 0, 1, 12))	349433.61	0.0845
Sarimax((1, 1, 0),(1, 0, 1, 12))	335927.87	0.0801
sarimax((0,1,2),(1, 0, 1, 12))	328806.57	0.0784
Garch("Zero", "GARCH",1, 0, 1, "normal")	316168.33	0.0685
Sarimax-garch((0,1,2),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	225254.30	0.0525
Sarimax-garch((0,1,2),(1, 0, 1, 12), "GARCH",1, 0, 1, "normal")	232885.97	0.0548
sarimax_garch((2,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	225247.52	0.0525
sarimax_garch((0,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	210452.97	0.0483

sarimax_garch((1,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	226276.06	0.0529
Rolling-sarimax,(0,1,2),(1, 0, 1, 12)	152330.24	0.0347
Rollings-sarimax-garch((0,1,2),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	151867.76	0.0344
Rollings-sarimax(2,1,0),(1, 0, 1, 12)	152080.60	0.0346
Rollings-sarimax-garch((2,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	151614.65	0.0343
Rollings-sarimax (0,1,0),(1, 0, 1, 12)	148878.77	0.0350
Rollings-sarimax-garch ((0,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	149143.05	0.0351
Rollings-sarimax (1,1,0),(1, 0, 1, 12)	151486.25	0.0342
Rollings-sarimax-garch (1,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')	151026.28	0.0340

The studies involve applying several SARIMA and SARIMA-GARCH models to the Gasoline dataset with and without a rolling window technique and evaluating their performance using RMSE and MAPE scores as illustrated in Table (5) . The GARCH models are identified by their specific GARCH(p,o,q) parameters, which represent the autoregressive orders, the order of the lagged conditional variance terms, how many past values of the conditional variance are included as predictors for the current value, and moving average terms for the conditional variance respectively. The prediction and the forecasting results are depicted in Figure (5) and (6).

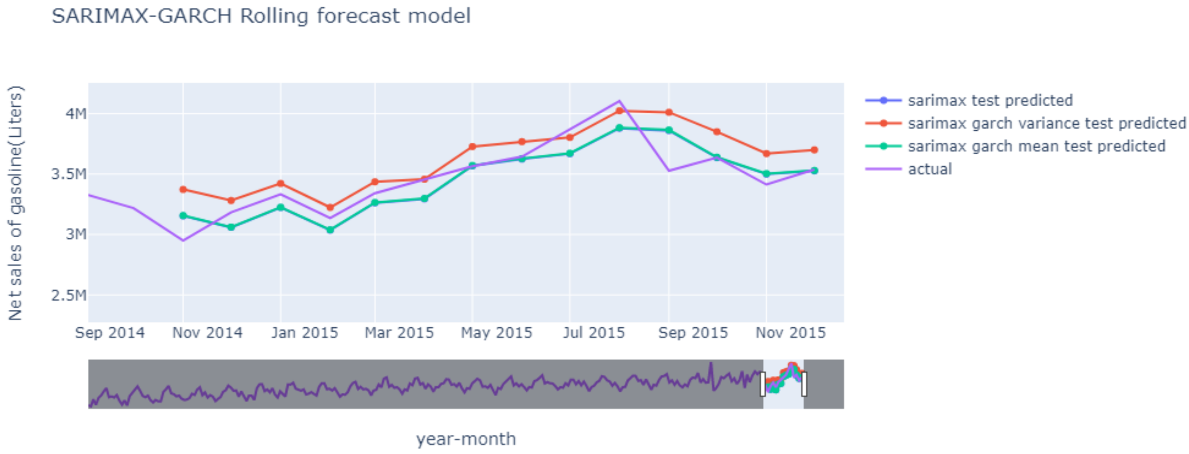

Figure 5. Prediction on Test Part using MODEL((1,1,0),(1, 0, 1, 12), "GARCH", 1, 1, 2, 'normal')

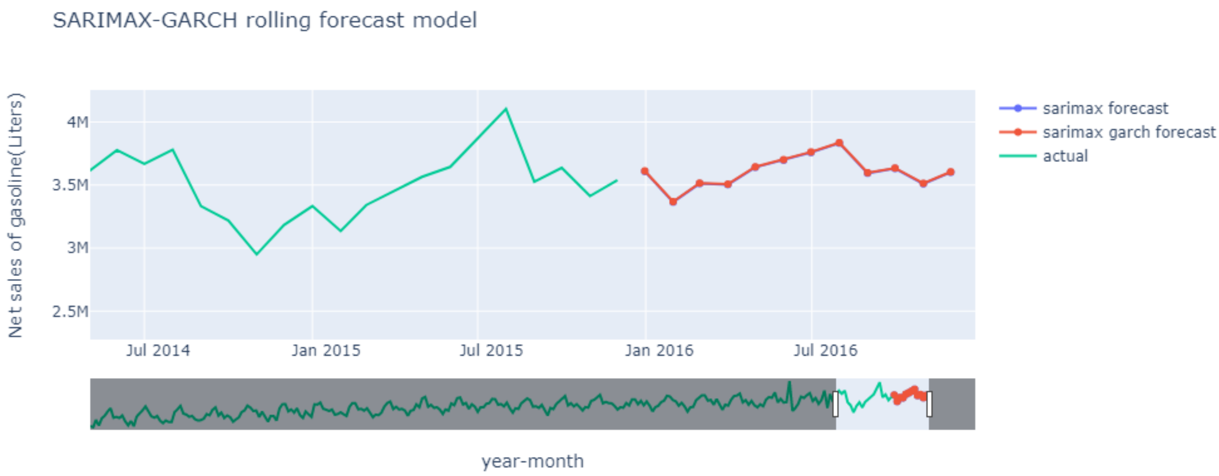

Figure 6: Forecast for Next 12 Month

The findings of this study revealed that SARIMA-GARCH model's efficiency in capturing both seasonal trends and volatility in gasoline sales data. The models with the "Rolling-sarimax" and "Rolling-sarimax-garch" labels incorporate an expanding rolling window technique. These models achieve the best forecasting performance, with the lowest RMSE and MAPE values (around 151,000 and 0.034, respectively). The rolling window technique, continuously adds fresh data points to the sample, allowing the model to adapt to shifting patterns and increase forecast accuracy over time. Furthermore, the rolling window forecast approach assisted in minimizing lookback bias, a prevalent issue in time series forecasting in which past data overly impact future projections; it also aids in long-term forecasting.

5. Conclusion

According to the findings, the SARIMA-GARCH hybrid models employing the rolling window forecasting technique outperform the standalone SARIMA models for forecasting Gasoline sales in Canada, as shown by RMSE and MAPE scores. Lower RMSE indicates that the forecasts are closer to the actual values, while lower MAPE indicates that the percentage mistakes are reduced, implying that the SARIMA-GARCH models are more accurate for forecasting gasoline sales. Finally, this study makes an important contribution to the field of time series forecasting, notably in the domain of gasoline sales in Canada. When paired with the rolling window technique, the suggested SARIMA-GARCH model provides a resilient and adaptive forecasting framework. This model can help policymakers, energy industry professionals, and investors make informed decisions based on precise and trustworthy estimates of gasoline sales in Canada, resulting in better resource planning in the dynamic world of energy use and demand.

References

[1] Vo, N., & Ślepaczuk, R. (2022, January 20). Applying Hybrid ARIMA-SGARCH in Algorithmic Investment Strategies on S&P500 Index. *Entropy*, *24*(2), 158. https://doi.org/10.3390/e24020158

[2] Xu, F., Sepehri, M., Hua, J., Ivanov, S., & Anyu, J. N. (2018, November 5). Time-Series Forecasting Models for Gasoline Prices in China. *International Journal of Economics and Finance*, *10*(12), 43. https://doi.org/10.5539/ijef.v10n12p43

[3] Mardiana, S., Saragih, F., & Huseini, M. (2020, October 10). FORECASTING GASOLINE DEMAND IN INDONESIA USING TIME SERIES. *International Journal of Energy Economics and Policy*, *10*(6), 132–145. https://doi.org/10.32479/ijeep.9982

[4] Suradhaniwar, S., Kar, S., Durbha, S. S., & Jagarlapudi, A. (2021, April 1). Time Series Forecasting of Univariate Agrometeorological Data: A Comparative Performance Evaluation via One-Step and Multi-Step Ahead Forecasting Strategies. *Sensors*, *21*(7), 2430. https://doi.org/10.3390/s21072430

[5] Isiaka, A., Isiaka, A., & Isiaka, A. (2021, February 11). Forecasting with ARMA models. *International Journal of Research in Business and Social Science (2147-4478)*, *10*(1), 205–234. https://doi.org/10.20525/ijrbs.v10i1.1005

[6] Sharma, V., Cali, M., Sardana, B., Kuzlu, M., Banga, D., & Pipattanasomporn, M. (2021, November). Data-driven short-term natural gas demand forecasting with machine learning techniques. *Journal of Petroleum Science and Engineering*, *206*, 108979. https://doi.org/10.1016/j.petrol.2021.108979

[7] Chen, C. W. S., & Chiu, L. M. (2021, September 4). Ordinal Time Series Forecasting of the Air Quality Index. *Entropy*, *23*(9), 1167. https://doi.org/10.3390/e23091167

[8] Ramneet Singh Chadha, Shahzadi Parveen, Jugesh, & Jasmehar Singh. (2023, June). Indian Machinery and Transport Equipment Exports - Forecasting with External Factors Using Chain of Hybrid Sarimax-Garch Model. *Journal of Ubiquitous Computing and Communication Technologies*, *5*(2), 175–192. https://doi.org/10.36548/jucct.2023.2.005

[9] Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on (2023, July 28)

[10] Bollerslev, T. (1986, April). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

[11] arch.univariate.GARCH - Arch 5.5.0. Accessed on (2023, July 28) https://arch.readthedocs.io/en/latest/univariate/generated/arch.univariate.GARCH.html#arch.univariate.GARCH

[12] ARCH_GARCH Volatility Forecasting. Accessed on (2023, July 28) https://goldinlocks.github.io/ARCH_GARCH-Volatility-Forecasting/

[13] Dileep Kumar Shetty, Sumithra, & Ismail.B. (2018). Hybrid SARIMA-GARCH Model for Forecasting Indian Gold Price. RESEARCH REVIEW International Journal of Multidisciplinary, 03(08), 263–269. https://doi.org/10.5281/zenodo.1344062

[14] *3.3. Metrics and scoring: quantifying the quality of predictions*. (n.d.). Scikit-learn. Accessed on (2023, July 28) https://scikit-learn/stable/modules/model_evaluation.html

[15] statsmodels.tsa.stattools.adfuller - Statsmodels 0.14.0. Accessed on (2023, July 28) https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.adfuller.html

[16] *pmdarima: ARIMA estimators for Python — pmdarima 2.0.3 documentation*. (n.d.). Accessed on (2023, July 28) https://alkaline-ml.com/pmdarima/index.html

[17] *Sales of gasoline used for road motor vehicles, monthly, inactive - Open Government Portal*. Accessed on (2023, July 28) https://open.canada.ca/data/en/dataset/e75562a3-3dea-471b-b143-a8b918a49a2c

Author's biography

Ramneet Singh Chadha -Ramneet Singh Chadha is currently working as Associate Director in C-DAC, Noida. He has more than 25 years of experience in IT expertise. He has Master's in Management (IT) and M.S (Software Systems) from BITS Pilani. He was a College topper while pursuing his Bachelor in Computer Engineering from Nagpur University. He has vast experience in Health care Domain and Transit Domain. He has experience of developing and implementing National/ State wide Hospital projects and developed NCMC Standard for MOHUA in transit domain. He is experience in academia as well as industry.

Jugesh -Jugesh is a Project Engineer in Department of Embedded System at C-DAC, Noida. He completed his masters in Computer Application from C-DAC, Noida in 2023. He Graduated in Software Development from GNDIT Rohini affiliated to GGSIPU in Delhi, India. He has great inspiration and deep interest in Research work in the field of Artificial Intelligence and Machine Learning.

Shahzadi Parveen -Shahzadi Parveen is a Project Engineer in Department of Embedded System at C-DAC, Noida. She received her B.Tech in Computer Science from Jamia Hamdrad in Delhi, India in 2021. She has completed her schooling from Jamia Millia Islamia with good excellent academic records. She was Vice Chairperson in IEEE JHSB WIE during her graduation in engineering. She was awarded with Outstanding WIE Student Volunteer Award in 2021.

Jasmehar Singh -Jasmehar Singh is 2nd year student pursuing Computer Science from Shiv Nadar University with keen interest in Data science and Artificial Intelligence. He is passionate about exploring and solving the problems in the transit domain. He has great interest in Research work in the field of AI, ML and cyber security.