Journal of Ubiquitous Computing and Communication Technologies (ISSN: 2582-337X)
www.irojournals.com/jucct/

Indian Machinery and Transport Equipment Exports - Forecasting with External Factors Using Chain of Hybrid Sarimax-Garch Model

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

Associate Director[1], Project Engineer[2], Project Intern[3], Student[4]

[1,2,3]C-DAC, Noida, Uttar Pradesh, India

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

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

Abstract

To choose the best forecasting model, it is essential to comprehend time series data since external influences like social, economic, and political events may affect the way the data behave. This study considers outside variables that could have an impact on the target variable used in improving the predictions. India Machinery and Transport Equipment Dataset is gathered from various sources, are cleaned, pre-processed, the missing values are removed, data types are converted, and dependent variables are identified before being used. By incorporating the SARIMAX model with the GARCH model and experimenting with various parameters and conditions, the current study seeks to enhance it. The SARIMAX-GARCH Model is a time series forecasting method used to predict market swings and export values. A helper model is developed to forecast the exogenous value to forecast the export value, which is then used as input for the final model. The ideal parameters for boosting the hybrid model's performance were identified through hyperparameter tuning. The results of this study provide estimates for future export values and contribute to a better understanding of India's Machinery and Transport Equipment export market. This research work focuses on export

DOI: https://doi.org/10.36548/jucct.2023.2.005
Received: 03.05.2023, received in revised form: 03.06.2023, accepted: 18.06.2023, published: 28.06.2023

value forecasting with the use of future exogenous variables. Exogenous factors are essential for predicting market changes and, as a result, support the forecasting of precise export values.

Keywords: Time Series Forecasting, SARIMAX, External Factors, Export Prediction, GARCH, ARCH, Hybrid Forecasting model, Exogenous variable, Export Forecasting, India Machinery and Transport export.

1. Introduction

Forecasting export and import activity are difficult because of external variables and market volatility, particularly in specialized industries such as Indian machinery and transport equipment exports. Forecasting accuracy is critical for successful business planning and risk management, giving organizations important insights to make educated decisions. As a result, establishing trustworthy forecasting models that account for these aspects is critical.

The impetus for this research stems from the realization that precise forecasting of export and import activity may considerably assist international trade enterprises. Forecasts of this type enable businesses to anticipate market needs, modify production and inventory levels, optimize resource allocation, and eventually increase their overall competitiveness. Furthermore, good forecasting aids risk management by allowing organizations to respond proactively to possible disruptions such as changes in currency rates, economic circumstances, or trade rules.

The fundamental goal of this study is to improve the SARIMAX [9] model by creating a chain of SARIMAX-GARCH models to anticipate external factors impacting export and import operations, allowing for more accurate forecasts. SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) models are widely used in time series analysis to capture both trend and seasonality in data [11], whereas GARCH (Generalized Autoregressive Conditional Heteroskedasticity) [12] models capture volatility and residual clustering. This study constructs a robust forecasting framework that accounts for both the internal dynamics of the time series data and the effect of external events by combining these two models.

However, it is critical to recognize the limits of the suggested SARIMAX-GARCH model chain. To begin, forecast accuracy is strongly dependent on the availability and quality of data on external factors. Predictions may be less reliable if the data is inaccurate or incomplete. Second, future forecasts, like any forecasting model, are subject to inherent uncertainty, particularly when dealing with complicated and dynamic market situations.

Finally, the model may be restricted to comparable industries or areas, and its generalizability to other sectors or nations may necessitate extra validation and adaptation.

2. Related Work

In recent years, there has been a lot of interest in using hybrid models and machine-learning approaches to anticipate economic variables. Several studies have looked at the use of these models in other sectors, such as forecasting gold prices, exports, and international oil prices.

The authors of the study "Hybrid SARIMA-GARCH Model for Forecasting Indian Gold Price" by Dileep Kumar Shetty and Sumithra (2018) propose a hybrid model for modeling and forecasting the Indian gold price that combines the linear seasonal autoregressive moving average (SARIMA) and the non-linear generalized autoregressive conditional heteroscedasticity (GARCH) models. The research assesses predicting ability using metrics such as root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) and examines the goodness of fit using the Akaike information criterion (AIC). The data indicate that the SARIMA-GARCH model is a better technique for forecasting the Indian gold price.[6]

Y. Xiang (2022) explores the fluctuation patterns of international oil prices in his article "Using ARIMA-GARCH Model to Analyze Fluctuation Law of International Oil Price." To analyze and forecast worldwide oil prices, the study used the autoregressive integrated moving average (ARIMA) model and a combination model of ARIMA and generalized autoregressive conditional heteroskedasticity (GARCH). According to the findings, the ARIMA (1,1,0)-GARCH (1,1) combination model is more suited for short-term forecasting, with higher accuracy than the ARIMA model.[7]

To anticipate the long-term performance of Saudi Arabia's electrical industry, Alharbi and Csala (2022) suggested a seasonal autoregressive integrated moving average with exogenous components (SARIMAX) model. SARIMAX used historical data on power consumption, generation, peak load, and installed capacity. The study found that the SARIMAX model outperformed simpler autoregressive integrated moving average-based forecasting methodologies. The model's capacity to handle various-sized sequential datasets as well as the seasonal and external affecting elements all contributed to its promising performance.[8]

3. Proposed Work

The research methodology includes the quantitative research design, the machine learning life cycle, and the data collection technique. It also involves the identification of appropriate data sources and forecasting using the SARIMAX-GARCH combination.

3.1 Data Collection

This study will make use of the following data sources:

Time Series Data on Indian Machinery and Transport Equipment Exports: Obtained from the World Integrated Trade Solution (WITS) database, spanning 1988 to 2020.[1]

Indian and global inflation data derived from the World Bank's "One-Stop Source: A Global Database of Inflation" by Ha, Jongrim, Kose, M. Ayhan, and Ohnsorge, Franziska (2021).[2]

Oil prices can affect transportation costs which can affect export. Data on oil prices were obtained from the Brent Oil Prices dataset on Kaggle [3].

The death rate is a good indicator of natural disasters, pandemics, and wars. Feature engineering is used for the global death rate by averaging all country death rates.

Global Death Rate and Death Rate in India: Data from the World Bank's data repository, especially the indicator "SP.DYN.CDRT.IN."[4]

India Population: Data were taken from the World Bank's data repository using the indicator "SP.POP.TOTL".[5]

Table 1. India Machinery and Transport Equipment Exports (30 rows and 7 columns)

Field Name	Units
Index	Date Time (1990-2020)
Transport Equipment& Machinery	US Dollar
Indian Inflation	Consumer Price Index
Global Inflation	Consumer Price Index
Oil Price	US Dollar/barrel
Global death Rate	death/1000
India death Rate	death/1000
India Population	Natural Number

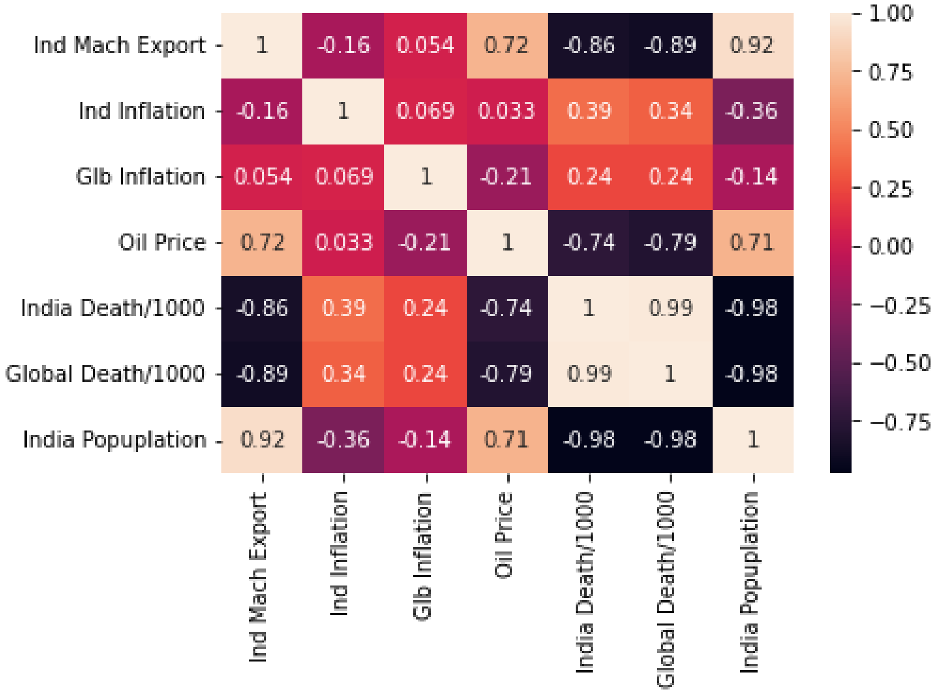

Figure 1. Correlation Heatmap, this Graph Shows Ind Mach Export (Transport Equipment & Machinery) is Highly Negatively Correlated with the Death Rate and Positively Correlated with India's Population.

3.2 SARIMAX-GARCH Model

The accuracy of time series forecasting can be increased by taking both the mean and volatility of the data into consideration by combining a SARIMAX model with a GARCH model. A common name for the combined model is SARIMAX-GARCH.

The GARCH model calculates the conditional variance, while the SARIMAX model calculates the forecast of the time series. The combined model can be expressed as follows:

Equation 1. ARMA-GARCH Equation

$$y_t = \mu_t + \varepsilon_t$$

$$\mu_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \gamma_1 x_{1t} + \gamma_2 x_{2t} + ... + \gamma_k x_{kt}$$

$$\varepsilon_t = \sigma_t \varepsilon_{t-1}$$

$$\sigma^2_t = \omega + \alpha_1 \varepsilon^2_{t-1} + \beta_1 \sigma^2_{t-1}$$

where y_t is the time series being modeled, x_t is a vector of exogenous variables, ε_t is the error term, and σ^2_t is the conditional variance. The parameters α, β, and γ represent the coefficients of the SARIMAX model, while ω, α_1, and β_1 represent the coefficients of the GARCH model. p is the order of the autoregressive component of the SARIMAX model, and k is the number of exogenous variables. [13]

3.3 Understanding Approach for Forecasting with Unknown Exogenous Variable

In this research, a chain of SARIMAX-GARCH models was created, one final model for "Transport Equipment & Machinery Export" using exogenous variables "India Inflation", "Global Inflation," "Oil Price", "India Death Rate", "Global Death Rate", and "India Population". However, since the values of these variables in the future are unknown, separate SARIMAX-GARCH models were created for each external factor to improve forecasting accuracy.[14] The target variable can be accurately forecasted by linking the exogenous model with the final model. Hyperparameter tuning will improve the performance of these models.[15]

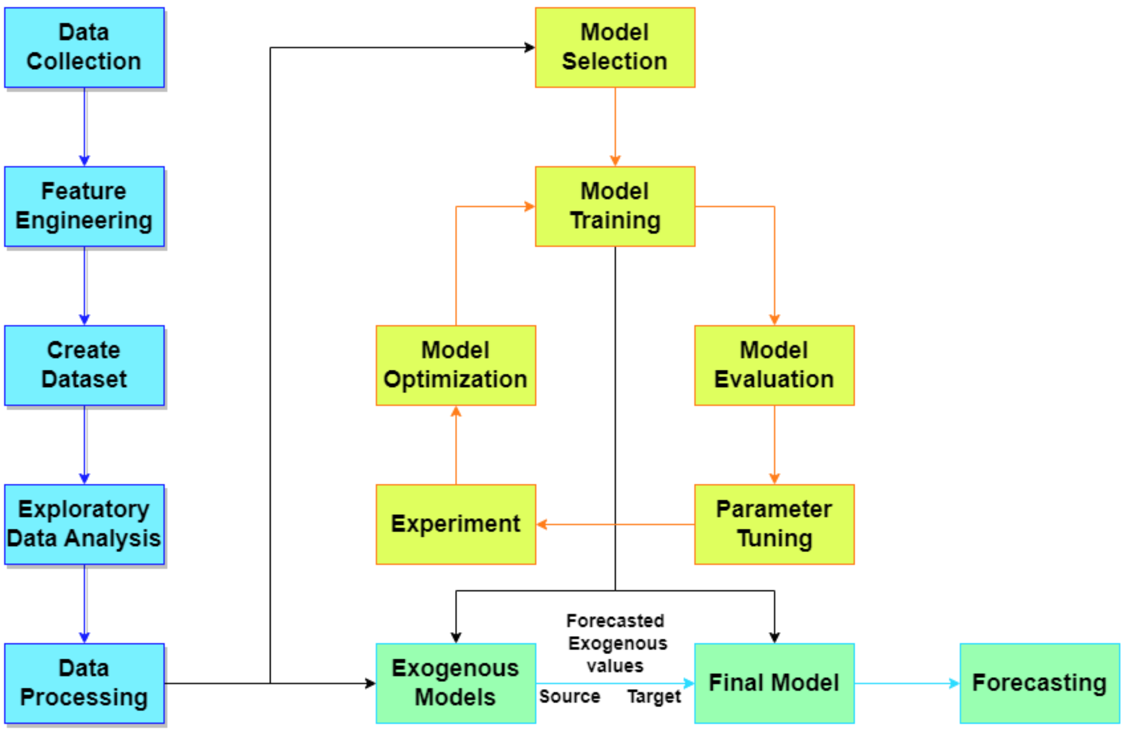

Figure 2. Flow Diagram

Figure 2 (flow diagram) shows all the processes carried out in this research, data collection from various sites, feature engineering for required data such as global death rate, and global inflation rate by averaging all country data, creating data set by combining all different data set into one data frame, analyzing data and their correlation, processing data checking if any none value or outlier exists in the dataset.

In model selection, the combination of exogenous variables, scaled exogenous variables, residuals of exogenous variables, scaled residuals of exogenous variables, and scaled target variables were used to examine the SARIMAX- GARCH model.

The sklearn library's Min-Max-Scaler function was used to conduct the scaling[17], which scales data to a range between 0 and 1. Green Block represents the chain of the exogenous model and the final model, exogenous model forecasts the external factors so they can be used in the final model. And final model forecast the export value of the targeted variable.

4. Experiments and Discussion

In this research, total 30 sample were used (yearly data 1990-2020), 90% for training and 10% for testing purposes in different scenarios shown in graph below. Table 2 represents the RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) scores of the Final models on the test part.

The Root Mean Square Error (RMSE) is the square root of the MSE (mean square error). RMSE is used to convert MSE back into the same units as the actual data.

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}}$$

The average of absolute errors divided by actual observation values is the mean absolute percentage error (MAPE). MAPE should not be utilized if the data contains zeros or near- zeroes.

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|}{n} \times 100$$

Where A is actual value at time t and F is the forecast value at time t and n is the number of periods. [16]

Graph interpretation-

- The Red line shows the SARIMAX model prediction on the training part.
- The Magenta line shows SARIMAX-GARCH prediction on the training part
- The Cyan line shows SARIMAX-GARCH forecasting on the testing part.
- The Green line shows only SARIMAX forecasting on the testing part.
- The Blue line shows actual value of Indian Machinery and Transport Equipment Exports (1990-2020).

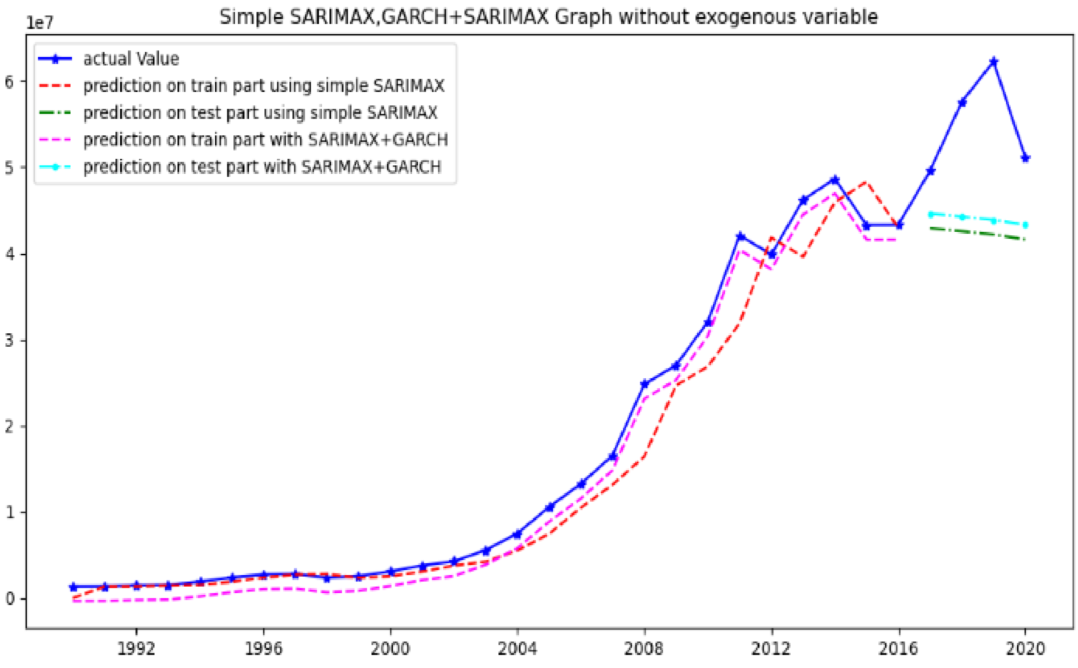

Figure 3. Experiment 1- Sarimax (order (1,0,0)) and GARCH (mean='Constant', vol='GARCH', p=1, o=1, q=1, dist='generalized error') model without adding exogenous variable

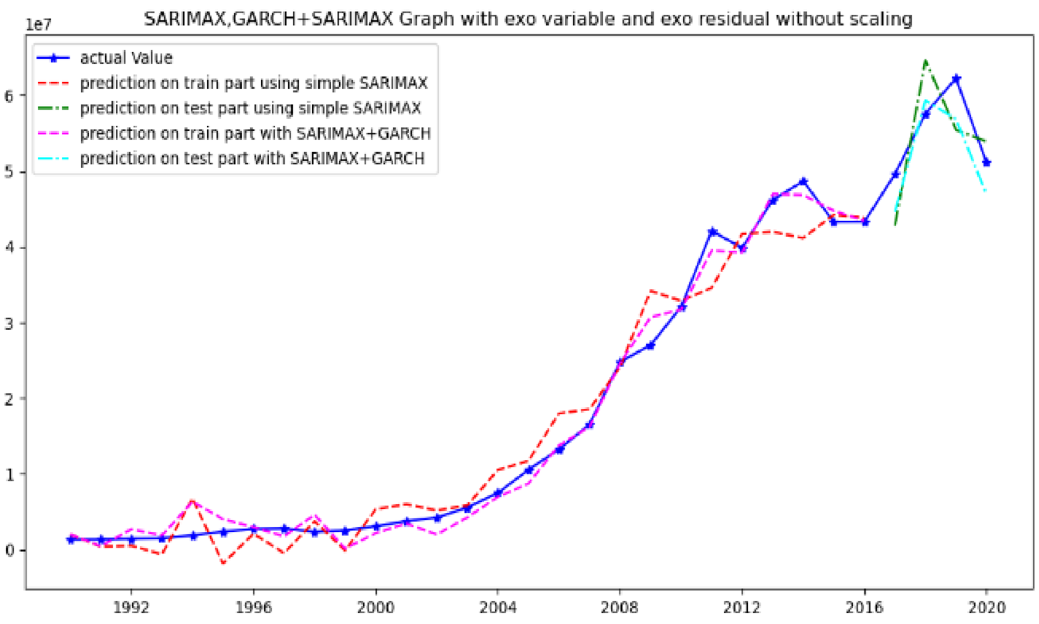

Figure 4. Experiment 2- SARIMAX (order= (1, 0, 0)) Model with Exogenous Variable and GARCH (mean='ARX', vol='GARCH', p=1, o=1, q=1, dist='generalized error') Model with Exogenous Residual without Scaling

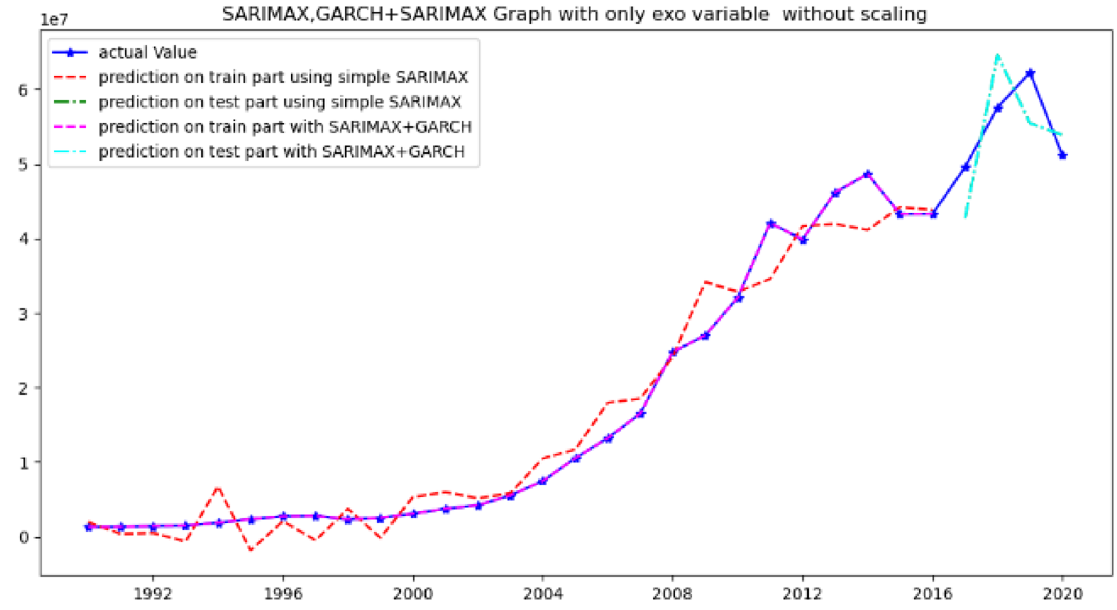

Figure 5. Experiment 3 -SARIMAX (order= (1, 0, 0)) Model with Exogenous Variable and GARCH (mean='Constant', vol='GARCH', p=1, o=1, q=1, dist='generalized error') Model without Adding Exogenous Residual

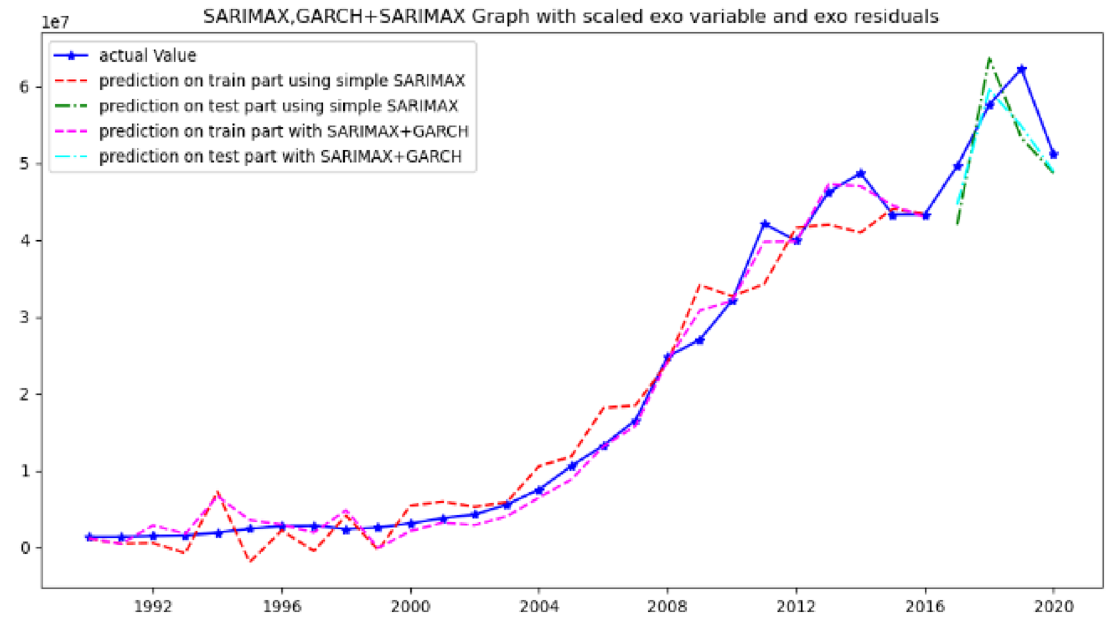

Figure 6. Experiment 4- SARIMAX ((1, 0, 0)) Model with Scaled Exogenous Variable and GARCH (mean='ARX', vol='GARCH', p=1, o=1, q=1, dist='generalized error') Model with Scaled Exogenous Residual

Table 2. Evaluation Table of All Experiments

Experiments	**Test RMSE Score (GARCH-SARIMAX)**	**Test RMSE Score (SARIMAX)**	**Test MAPE value ranging between 0 to 1 (GARCH-SARIMAX)**	**Test MAPE value ranging between 0 to 1 (SARIMAX)**
Experiment 1	12,253,678.10	13,811,537.65	0.194	0.225
Experiment 2	4,317,400.15	6,117,565.89	0.074	0.105
Experiment 3	6,118,153.68	6,117,565.89	0.105	0.105
Experiment 4	4,736,160.60	6,747,412.82	0.074	0.112

4.1 External Factors Forecasting for Final Model

SARIMAX and GARCH models were created in this study to forecast the values of each exogenous variable. A data frame containing the forecast values of exogenous variables and their residuals was then generated using the predicted values. The final model used this data frame as an exogenous input.

Table 3. Forecasting External Factors Value for (2021-2022)

Years	**India Inflation Rate**	**Global Inflation Rate**	**Oil Price**	**India Death Rate**	**GlobalDeath Rate**	**India Population**
2021	7.02	42.81	43.59	7.97	8.07	1.386083e+09
2022	7.21	30.68	43.59	8.67	8.01	1.398480e+09

Table 4. Forecasting External Factors Residuals for (2021-2022)

Years	India Inflation Rate	Global Inflation Rate	Oil Price	India Death Rate	GlobalDeath Rate	India Population
2021	3.65	-23.14	14.25	6.54	7.58	8.474908e+08
2022	3.65	-23.14	14.25	6.54	7.58	8.474908e+08

5. Result

The SARIMAX model for Indian machinery and transport equipment export was improved by incorporating the GARCH model, which is capable of predicting volatility and capturing fluctuations in export data. Furthermore, the proposed method addresses the issue of unknown future values of external factors needed for final forecasting. This is accomplished by developing individual SARIMAX-GARCH models for each external factor, which are then linked to the final SARIMAX-GARCH model. This comprehensive framework improves forecasting accuracy by considering both internal and external factors that influence export dynamics.

The value of India Machinery and Transport Equipment export for the next two years has been predicted after getting the exogenous values to be used in the final model.

Table 5. Forecasting India Machinery and Transport Equipment Export for (2021-2022)

Year	Value
2021	79230762.78
2022	98237774.65

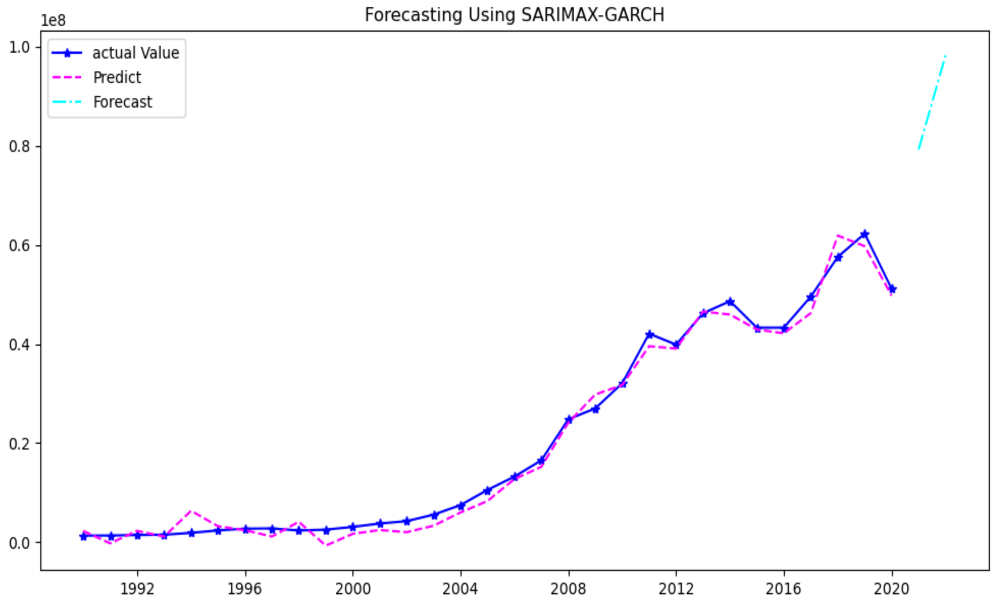

Figure 7. Graph of in-Sample Forecasting and Out-Sample (2021-2022) Forecasting of India Machinery and Transport Equipment export

6. Conclusion

Table 2 shows that for all four experiments, the Garch-Sarimax model consistently beats the simple Sarimax model in terms of the RMSE score. Experiment 2 has the lowest test RMSE score for both Garch-Sarimax and plain Sarimax models. Overall, the Garch-Sarimax model seems to be a better choice for predicting the data as compared to the simple Sarimax model. This study uses a thorough methodology to forecast India's Machinery and Transport Equipment export for the following two years. The predicted values are 79230762.78 and 98237774.65 over the subsequent two years (2021-2022). This method offers a solid projection for India's Transport Equipment export, which decision-makers, business professionals, and other stakeholders can use.

7. Future Scope

This study enhances SARIMAX forecasting models, especially for the export of machinery and transport equipment from India. By considering the unique features of the data, such as trends and potential outside impacts, the research provides a more precise and comprehensive predicting answer. By considering how external factors affect export trends, the model is better able to anticipate future events. This study has the drawback that long-term predicting accuracy deteriorates over time. This restriction opens the possibility for future research into alternative modeling approaches or the inclusion of new parameters that may increase the precision of long-term forecasts.

Acknowledgments

I would like to express our sincere gratitude to Mr. Ramneet Singh Chadha, our supervisor, for his guidance and support throughout this research work. I also thank our colleagues and research collaborators, for their valuable input and feedback. This research was supported by the CDAC-Noida. I would also like to thank the participants who took part in our study and the staff at CDAC institution for their assistance. Finally, I would like to thank our families and friends for their unwavering support and encouragement throughout this journey.

References

[1] India Machinery and Transport Equipment Exports to World US$000 1988 - 2020 | WITS Data.(2020,April13). https://wits.worldbank.org/CountryProfile/en/Country/IND/StartYear/1988/EndYear / 2020/TradeFlow/Export/Indicator/XPRT-TRD-VL/Partner/WLD/Product/Transp

[2] A Global Database of Inflation" by Ha, Jongrim, Kose, M. Ayhan, and Ohnsorge, Franziska (2021), published by the World Bank. A Global Database of Inflation. (n.d.). World Bank. https://www.worldbank.org/en/research/brief/inflation-database

[3] Brent Oil Prices. (n.d.). Brent Oil Prices | Kaggle. https://datasets/mabusalah/brent-oil-prices

[4] World Bank Open Data. (n.d.). World Bank Open Data. https://data.worldbank.org

[5] Population, total - India | Data. (n.d.). Population, Total - India | Data. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=IN

[6] Shetty, Dileep Kumar and Sumithra. "Hybrid SARIMA-GARCH Model for Forecasting Indian Gold Price." (2018).

[7] Xiang, Y. (2022, March 17). Using ARIMA-GARCH Model to Analyze Fluctuation Law of International Oil Price. Mathematical Problems in Engineering, 2022, 1–7. https://doi.org/10.1155/2022/3936414

[8] Alharbi, F. R., & Csala, D. (2022, October 16). A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. Inventions,7(4),94.https://doi.org/10.3390/inventions7040094

[9] statsmodels.tsa.statespace.sarimax.SARIMAX—Statsmodels. https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html

[10] Forecasting: Principles and Practice (2nd ed). Retrieved April 13, 2023, from https://otexts.com/fpp2/

[11] SARIMAX and ARIMA: Frequently Asked Questions (FAQ) — statsmodels. Retrieved Apri13,2023, from

https://www.statsmodels.org/dev/examples/notebooks/generated/statespace_sarimax_faq.html#Differences-between-trend-and-exog-in-SARIMAX

[12] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.

[13] ARIMA-GARCH forecasting with Python. Retrieved April 13, 2023, from https://medium.com/analytics-vidhya/arima-garch-forecasting-with-python-7a3f797de3ff

[14] Forecasting with Exogenous Regressors — arch 5.3.2.dev67+g00dbf506 documentation. Retrieved,April13,2023,from,https://arch.readthedocs.io/en/latest/univariate/univariate_forecasting_with_exogenous_variables.html

[15] David A. Dickey & Wayne A. Fuller (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74:366a, 427-431, DOI: 10.1080/01621459.1979.10482531

[16] Wright, D. (2016, February 27). Use Excel to Calculate MAD, MSE, RMSE & MAPE - Dawn Wright, Ph.D. Dawn Wright, Ph.D. https://www.drdawnwright.com/use-excel-to-calculate-mad-mse-rmse-mape/

[17] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, D. (n.d.). Scikit-learn: Machine Learning in Python. Scikit-learn: Machine Learning in Python. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Author's Biography

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 is a final Year Student of Master of Computer Application at C-DAC, Noida affiliated to GGSIPU. He Graduated in Software Development from GNDIT Rohini affiliatedto GGSIPU in Delhi, India. He is currently working as project inter at C-DAC Noida. He has great inspiration and deep interest in Research work in the field of Artificial Intelligence and Machine Learning.

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 Chadha
15-03-2022

Jasmehar Singh is first 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.