MSC 91B84 **DOI: 10.14529/jcem250404**

MODELING THE SALES VOLUME OF COMPULSORY LIABILITY INSURANCE POLICIES OF HAZARDOUS FACILITY OWNERS

M. O. Sokolov[1]**,** mosokolov20@gmail.com**,**
N. N. Ovchinnikova[1]**,** ovchinnikovann@susu.ru**,**
E. I. Nazarova[1]**,** nazarovaei@susu.ru
[1]South Ural State University, Chelyabinsk, Russian Federation

This paper performs modeling the sales forecast of compulsory insurance policies of hazardous production facilities for an insurance company in the Chelyabinsk region. The time series analysis of insurance policy sales was carried out and the data were tested for normal distribution and stationarity. ARIMA and SARIMA models were used to generate a forecast. The developed models were validated and demonstrated good forecast quality, confirming their practical applicability for assessing insurance sales dynamics.

Keywords: modeling; forecast; insurance policies; statistics.

Introduction

Insurance is an essential element of the modern economy, ensuring the redistribution of risks among business participants. Compulsory civil liability insurance for owners of hazardous production facilities (CLIHF) plays a special role, as it aims to compensate for damage in case of emergencies. In the context of uncertainty in the insurance market, a model for forecasting insurance policy sales volumes becomes crucial. For insurance companies, accurate sales forecast allows for the business process optimization, insurance reserve creation, and risk management [1, 2]. Modern forecasting methods include statistical and machine-based approaches, among which ARIMA and SARIMA models hold a special place [3]. The use of ARIMA and SARIMA models allows taking into account trends, seasonality, and random fluctuations, which makes them universal tools for forecasting economic indicators and financial time series [4, 5].

A number of theoretical results have been obtained in the field of insurance sales forecasting. However, their application to a specific insurance market participant is not always effective. These studies typically model the activities of specific companies and construct models of their insurance product consumption. Furthermore, the development of the insurance market itself depends largely on the current socioeconomic situation in a given region. Therefore, building a model to forecast sales volumes of compulsory insurance policies of hazardous industrial facilities for a specific company will help it optimize its business processes.

1. Time Series Research

To build a mathematical model for forecasting the sales volume of CLIHF policies, an insurance company in the Chelyabinsk region provided data for the period 2022–2024. Fig. 1 shows the dynamics of the cost of CLIHF policies (in conventional monetary units) over this period.

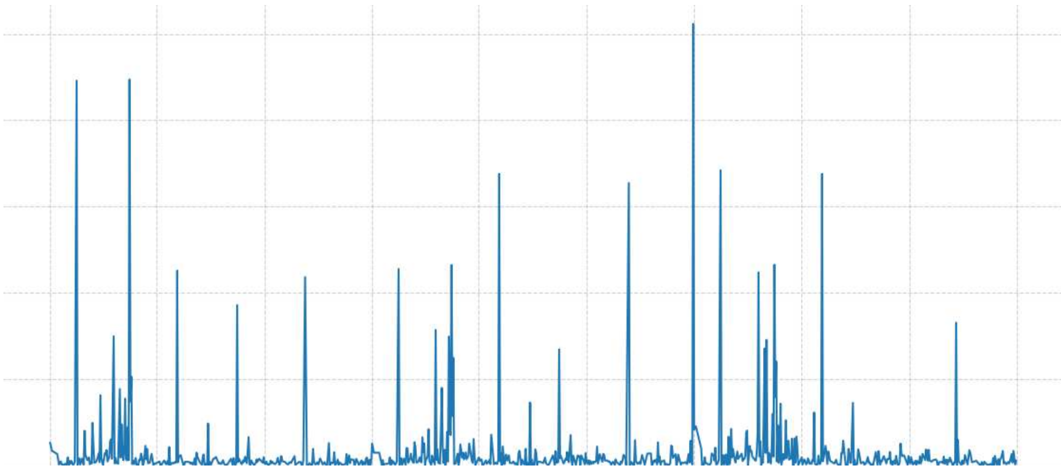

Fig. 1. Initial time series of CLIHF policy sales

The time series contains peaks that exceed the basic level values; these are outliers or anomalies. Further analysis revealed that the data were widely dispersed and asymmetrical. To separately examine the most frequently sales of policies, the so-called mass segment, and policies that are much less common, and to eliminate distortions in the analysis of frequent sales, the dataset was split into two groups (Figs. 2–3). This approach takes into account the specific characteristics of each model.

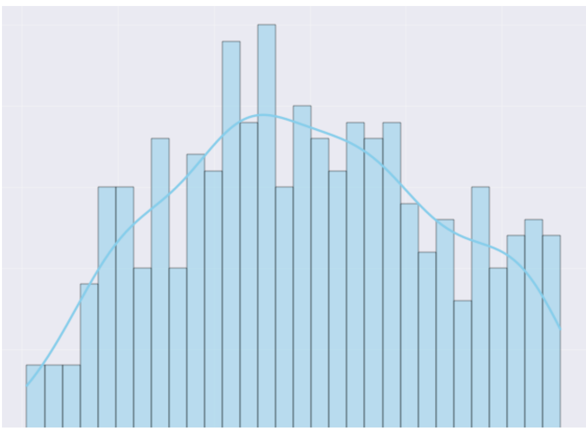

Fig. 2. Frequency histogram for group 1

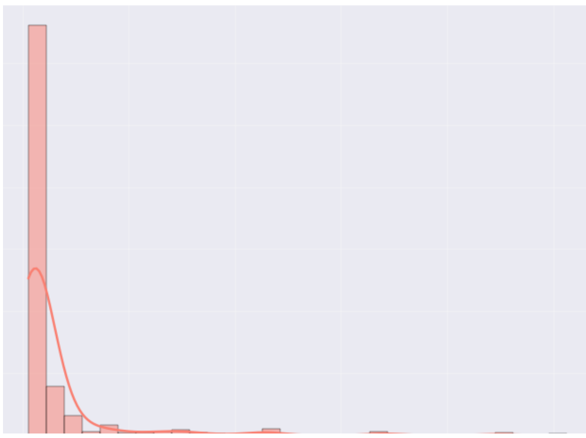

Fig. 3. Frequency histogram for group 2

Using the Shapiro-Wilk test, the data were tested for normal distribution. The results are presented in Table 1.

Table 1

Shapiro-Wilk test results

Group	W statistic	p-value
Group 1	0.98	0.000003
Group 2	0.96	0.000017

Two groups demonstrate low p-values. The particularly noticeable deviation in group 2 indicates strong asymmetry. To ensure stability of variance and reduce the asymmetry of the time series distribution, the Box-Cox transformation was applied, which normalizes the data and improves the accuracy of model parameter estimates. Therefore, for the subsequent stages of statistical analysis, the data were reduced to a normal distribution (Figs. 4–5).

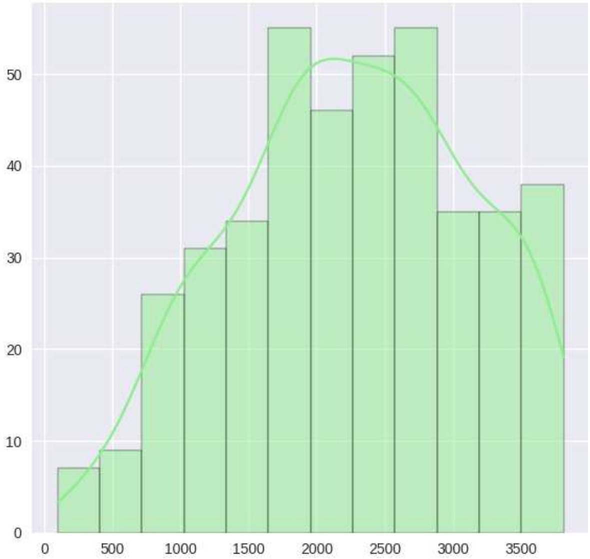

Fig. 4. Updated frequency histogram for group 1

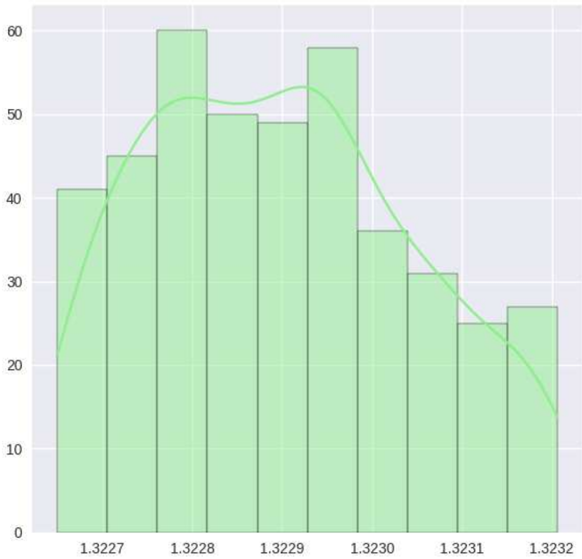

Fig. 5. Updated frequency histogram for group 2

The Box-Cox transformation brought both groups closer to a normal distribution. However, the p-values remained very small, preventing the acceptance of the hypothesis of a normal sample distribution.

The augmented Dickey-Fuller (ADF) test was used to test the stationarity of the time series. The test rejected the null hypothesis of a unit root at the 5% significance level, demonstrating the stationarity of the time series and their suitability for ARIMA modeling (Table 2).

Table 2

Results of the Dickey–Fuller test for stationarity of series 1 and 2

Group	ADF statistic	p-value	Critical value (5%)
Group 1	-4.50	0.0002	-2.87
Group 2	-12.48	$3.02 \cdot e^{-23}$	-2.87

We compared the means of the two groups to ensure that the groups actually differed in their means. A one-way analysis of variance (ANOVA) was used.

Levene's test for the homogeneity of sample variances allows for the conclusion on the heterogeneity of variances. Welch's t-test, which does not require the assumption of homogeneity, was used to accurately compare means. The test results showed a statistically significant difference in means (Table 3), confirming the difference in the indicator level between the groups under study.

Table 3

Results of the Welch's t-test

Test	Statistic	p-value
Welch's	53.14	< 0.001

2. Modeling Sales Volumes of Insurance Policies

After splitting the data into two groups with different statistical characteristics, we construct forecast models for each group. To select a forecast model, it is important to examine the autocorrelation structure of the data (Figs. 6–7).

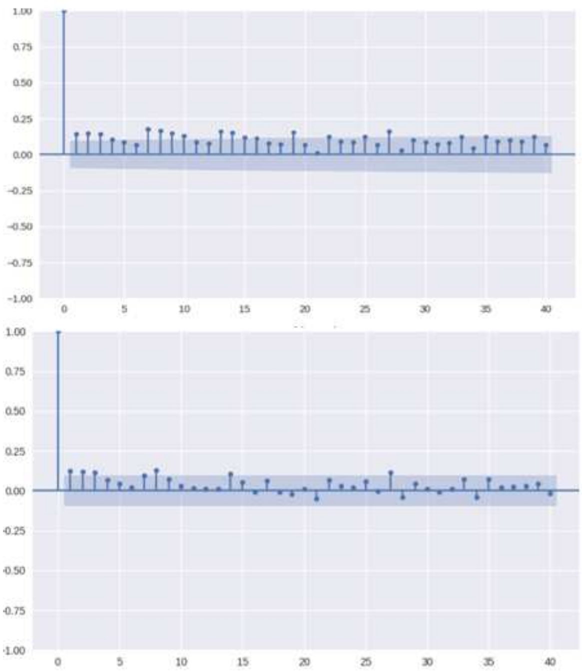

Fig. 6. ACF and PACF graphs for group 1

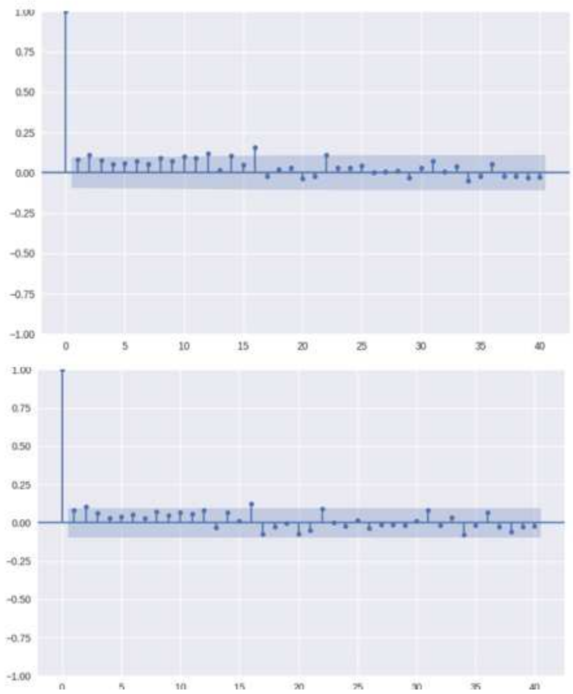

Fig. 7. ACF and PACF graphs for group 2

For group 1, correlogram analysis reveals weak seasonality of order $s = 4$ (local peaks at lags 4, 8, and 12). The PACF decreases stepwise up to lag 3, corresponding to an AR component of order $p = 3$, and the ACF has a significant first lag – $q = 1$. For group 2, seasonality tests are negative, and the correlograms correspond to short-term dependencies of order $(1, 1)$.

The SARIMA model was constructed for the first data group:

$$y_t = 1.0138y_{t-1} + 0.0158y_{t-2} - 0.0298y_{t-3} + 0.9974y_{t-4} + \\ +0.0732y_{t-8} - 0.0719y_{t-12} + \epsilon_t - 0.9750\epsilon_{t-1} - 0.9948\epsilon_{t-4}.$$

The forecast error (MAPE) was 17%. Figure 8 presents the forecast result.

For the second group, the ARIMA model was built:

$$y_t = -0.0298 + 1.3229y_{t-1} + 0.0158\epsilon_{t-1} + \epsilon_t.$$

The forecast error (MAPE) was 24%. Figure 9 presents the forecast result.

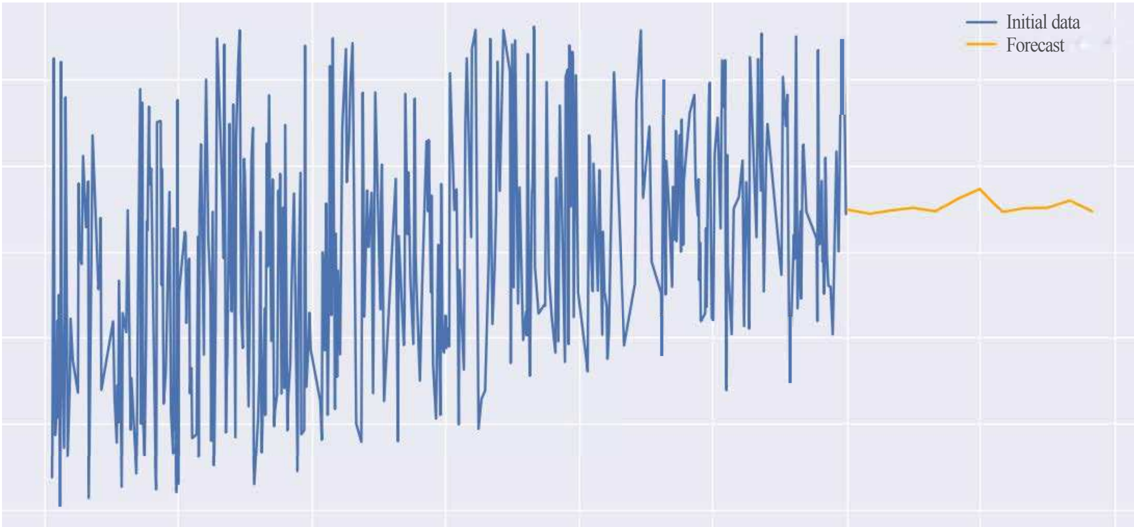

Fig. 8. SARIMA forecast model

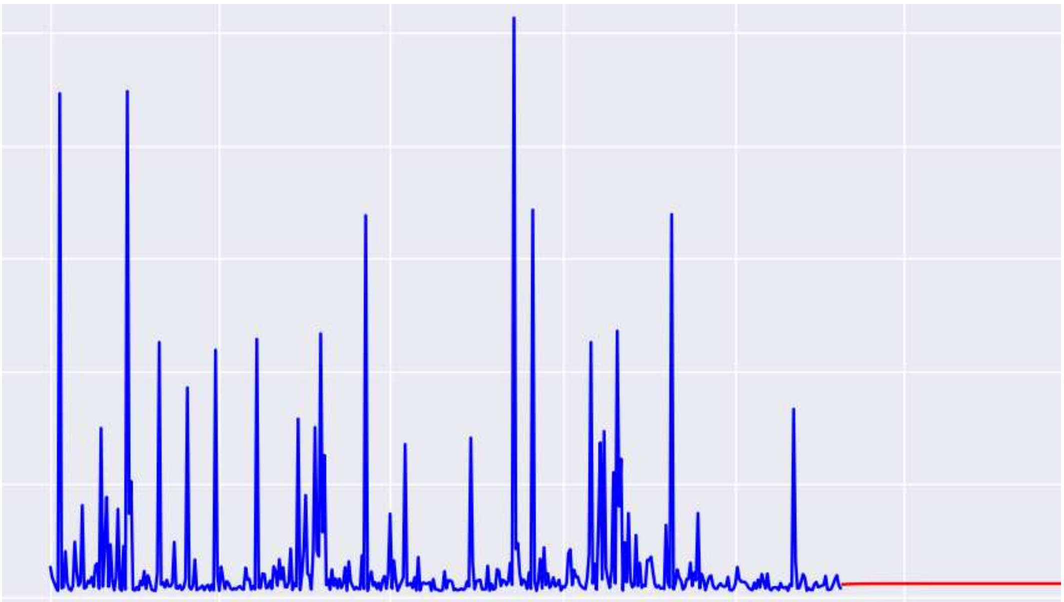

Fig. 9. ARIMA forecast model

To test the adequacy of the models, a residual analysis was performed. The residuals of both models passed tests for randomness (runs test), normality (Shapiro-Wilk test), and absence of autocorrelation (Durbin–Watson test). The resulting DW statistic values were 1.98 and 2.00, indicating independence of the residuals. Therefore, the constructed models are adequate and suitable for forecasting.

Conclusion

The comparative analysis revealed that the SARIMA model provides a more accurate description of seasonal fluctuations, while the ARIMA model demonstrates high robustness in the absence of seasonality. The forecast values showed a decline in sales volumes at the beginning of the forecast period and a gradual recovery by the end of the year. The obtained results can be used to formulate sales plans and marketing strategies for insurance companies, optimize insurance policy inventory, and manage risks.

The forecast error (MAPE) for the first group was 17%, which is a good result for forecasting sales volume in monetary terms. For the second group, the model quality score was 24%, which is higher than for the first group but still satisfactory. Accuracy in the second group dropped due to highly noisy data.

The proposed approach can be used to forecast other financial indicators and optimize business processes for insurance companies.

References

1. Mahdieva Yu.M. Modern Features and Prospects for the Development of Compulsory Liability Insurance in Russia and the Regions.*Regional'nye Problemy Preobrazovaniya Ekonomiki*, 2022, no. 12, pp. 40–46. DOI: 10.26726/1812-7096-2022-12-40-46 (in Russian)
2. Fookina S.P. Estimation of Model of Development of the Insurance Market of Russia. *Azimut Nauchnyh Issledovanij: Ekonomika i Upravlenie*, 2014, no. 2, pp. 77–80. (in Russian)
3. Box G.E.P., Jenkins G.M. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970.
4. Rukavishnikova N.G., Zavertkina E.G. [*Statistical Analysis of Data and Methods of Presenting the Results of the Study*]. Yaroslavl', YaGPU Publ., 2000. (in Russian)
5. Zerkina E.K., Nazarova E.I. Dynamics Modeling of Credit Interest Rate. *Journal of Computational and Engineering Mathematics*, 2023, vol. 10, no. 3, pp. 54–66. DOI: 10.14529/jcem230305

Maxim O. Sokolov, Undergraduate, Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), mosokolov20@gmail.com

Natal'ya N. Ovchinnikova, PhD(Educ), Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), ovchinnikovann@susu.ru

Elena I. Nazarova, PhD(Math), Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), nazarovaei@susu.ru

Received November 15, 2025

УДК 519.2 **DOI: 10.14529/jcem250404**

МОДЕЛИРОВАНИЕ ОБЪЕМА ПРОДАЖ ПОЛИСОВ ОБЯЗАТЕЛЬНОГО СТРАХОВАНИЯ ОСОПО

М.О. Соколов[1]***, Н.Н. Овчинникова***[1]***, Е.И. Назарова***[1]
[1]Южно-Уральский государственный университет, г. Челябинск, Российская Федерация

В данной статье выполнено моделирование прогноза объема продаж полисов обязательного страхования опасных производственных объектов в страховой компании Челябинской области. Проведен анализ временных рядов продаж страховых полисов, выполнена проверка данных на нормальность распределения и стационарность. Для построения прогноза применены модели ARIMA и SARIMA. Построенные модели прошли проверку на адекватность и продемонстрировали хорошее качество прогноза, что подтверждает возможность их практического применения для оценки динамики продаж страховых услуг.

Ключевые слова: моделирование; прогнозирование; статистика; страховые полисы.

Литература

1. Махдиева, Ю.М. Современные особенности и перспективы развития обязательного страхования ответственности в России и регионах / Ю.М. Махдиева // Региональные проблемы преобразования экономики. – 2022. – № 12. – С. 40–46.
2. Фукина, С.П. Оценка модели развития страхового рынка / С.П. Фукина // Азимут научных исследований: экономика и управление. – 2014. – № 2. – С. 77–80.
3. Бокс, Дж., Дженкинс, Г. Анализ временных рядов. Прогноз и управление: в 2 вып. / пер. с англ.; под ред. В.Ф. Писаренко. – М.: Мир, 1974.
4. Рукавишникова, Н.Г. Статистический анализ данных и способы представления результатов исследования / Н.Г. Рукавишникова, Е.Г. Заверткина. – Ярославль: ЯГПУ, 2000.
5. Zerkina, E.K. Dynamics Modeling of Credit Interest Rate / E.K. Zerkina, E.I. Nazarova // Journal of Computational and Engineering Mathematics. – 2023. – V. 10, № 3. – P. 54–66.

Соколов Максим Олегович, магистрант, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), mosokolov20@gmail.com

Овчинникова Наталья Николаевна, кандидат педагогических наук, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), ovchinnikovann@susu.ru

Назарова Елена Игоревна, кандидат физико-математических наук, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), nazarovaei@susu.ru

Поступила в редакцию 15 ноября 2025 г.