UDC 33

ARIMA Forecasting Model for Monthly Sales Volume in Manufacturing Enterprises

Zhang Jianhua

PhD in Economics,
Institute of EMIT,
Russian Presidential Academy of National Economy and Public Administration,
119571, 82, Vernadskogo ave., Moscow, Russian Federation;
e-mail: zhang_jianhua@vk.com

Abstract

The article discusses the ARIMA (Autoregressive Integrated Moving Average) model as a key tool for forecasting economic and financial time series. The main focus is on the principles of the ARIMA model, which optimizes production planning and inventory management by considering seasonal fluctuations and long-term sales trends based on five-year data. Using SPSS for time series analysis, an ARIMA model was developed, which passed tests for data smoothness and white noise, allowing the final model parameters (p, d, q) to be established. The ARIMA model demonstrates high forecasting accuracy and ease of use, making it an indispensable tool for time series analysis.

For citation

Zhang Jianhua (2024) ARIMA forecasting model for monthly sales volume in manufacturing enterprises. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 14 (10A), pp. 291-299.

Keywords

ARIMA, economic forecasting, SPSS, first-order difference, serial charts, time series model.

Introduction

Economic forecasting is the speculation and estimation of future scenarios of economic phenomena, based on the history and current situation of economic development, using scientific forecasting methods to reveal the development laws of economic phenomena and the interrelationships between various economic phenomena [Meshram, Deodas, 2012]. The purpose of time series analysis is to study the path observation of time series and establish a model to describe the data structure and predict the future values of time series [Lindenberg, Sonja, 2012]. It has a wide range of applications in business, economics, finance, and computer science, with the aim of serving economic decision-making for future problems.

The most common prediction method we use is the univariate "autoregressive moving average (ARMA)", which combines autoregressive (AR) and moving average (MA) models [Ayodele Ariyo Adebiyi, 2014]. Univariate "autoregressive integrated moving average (ARIMA)" is a special type of ARMA. Multivariate ARIMA models and vector autoregressive (VAR) models generalize univariate ARIMA models and univariate autoregressive (AR) models by allowing multiple constantly changing variables.

Material and method

A time series, also known as a dynamic sequence or time series, is a sequence arranged in chronological order that reflects the process and characteristics of socio-economic phenomena over time [A.Fernández-Manso, 2011]. Compared with regression analysis, time series analysis focuses on the correlation between data before and after time, while regression analysis focuses on the correlation between independent variables and dependent variables (variables with randomness). Based on the long-term trend of the time series, with time as the independent variable and sequence indicators as the dependent variable, fit the function equation y=f (t) to make extrapolation predictions. Based on mathematical statistics methods and stochastic process theory, this study investigates the statistical laws followed by random time series data, solves and predicts practical problems, and predicts future numerical values [ZHANG G H, 2019].

Random and non random time series - The average and variance of the observed values (y_t) in the time series gradually increase or decrease over time.

Steady and non-stationary time series - Over time, the average and variance of the observed values (y_t) in the series are very close.

Stationarity describes the concept of how time series will remain invariant in the future, and its statistical properties include mean invariance, variance invariance, and covariance independent of time [ZHANG Y J, 2019]. Time series prediction models require stationary time series because they are easier to model and have constant statistical properties. Therefore, if the time series is not stationary, it should be made as stationary as possible. In time series analysis, data with stationarity means that the above statistics do not undergo significant changes. Stable time series is the most important special type in time series analysis, and time series analysis is basically based on stable time series. In time series analysis, ensuring the stationarity of data is a necessary prerequisite before proceeding with subsequent time series analysis steps. For non-stationary time series, statistical analysis is limited by methods and theories. Due to the instability of time series, statistical characteristics obtained from historical data are meaningless for the future. This is a basic assumption that must be made before analyzing any problem [ZHENG M G, LI Q, 2020]. It can be determined through visual methods and unit root test statistical methods.

In the ARIMA model, differencing is the only way to smooth time series data, and differencing is also a differencing operation. From a computational perspective, it is the difference between the values of two adjacent variables, the difference between the time series at time t and t-1, that is, the difference between the latter value and the former value, which reflects the changes between discrete variables [YOU Y L, LI J, 2020]. In time series data, subtracting the previous variable value from each variable value in reverse chronological order is the first-order difference, corresponding to a difference order of 1; Performing first-order differencing on the data after first-order differencing results in second-order differencing, rather than differencing the source data twice with an order of 2, and so on. Non stationary sequences can be transformed into stationary sequences through differencing [Banjul, Bhattacharyya, 2014], and differencing can help convert unstable sequences into stable ones

The ARIMA model can predict economic indicators, and quantitative forecasting methods are the most commonly used scientific methods and models in economic forecasting. The AR model believes that 'history determines the future'. On the other hand, the MA model believes that "time series are relatively stable, and their fluctuations are determined by accidental factors", but in reality, it is difficult to maintain this assumption of "stability" for time series. The ARIMA model combines the above two models. The basic idea of the ARIMA model is that the label value at a time point is influenced by both the label value in the past period and the accidental events in the past period [YAN X X, 2020]. Guided by this central idea, the formula of the ARIMA model is expressed as:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

In the formula, the meanings of each variable are exactly the same as those in the MA and AR models, and the solving methods for each variable are basically the same as those in the original MA and AR models. It is obvious that the first half of the formula is the AR model, and the second half is the part about "volatility" in the MA model. It is worth noting that the mean representing long-term trends in the MA model does not exist in the formula of the ARIMA model, because the "predicting long-term trends" function in the ARIMA model is performed by the AR model, so the AR model replaces the original ,In the ARIMA model, β_0 is a constant term that can be 0. ϵ_t is the error.

The findings of the research

The meanings of p and q in the ARIMA (p, d, q) model are exactly the same as those in the original MA and AR models, and p and q can be set to different values, while d is the order of difference required by the ARIMA model. For the ARIMA model, determining the values of p and q has two meanings: determining whether to use AR, MA, or ARIMA models, as shown in the table below:

Table 1- Models corresponding to ARIMA parameter settings

Parameter settings	Corresponding ARIMA model
ARIMA(p,0,0)	ARIMA is equivalent to autoregressive model AR
ARIMA(0,0,q)	ARIMA is equivalent to the moving average model MA
ARIMA(p,0,q)	ARIMA model

In statistics, the first step is to draw ACF and PACF images to determine parameter settings. For any time series, when the ACF image exhibits tailing and the PACF image exhibits truncation, the AR model is applied to the current time series, and the lag order of PACF truncation is the ideal value of the hyperparameter p

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon_t$$

For the AR model, the value of ACF corresponds to the effect of y_{t-1} on y_t, and the significance of PACF corresponds to the significance of [Demir V, 2020]

For any time series, when the PACF image exhibits tailing and the ACF image exhibits truncation, the current time series applies the MA model, and the lag order of ACF truncation is the ideal value of the hyperparameter q

$$y_t = \beta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

For the MA model, the significance of ACF corresponds to the significance of, and the value of PACF corresponds to the influence of ϵ_{t-1} on ϵ_t

Table 2-Parameters p,q and determination conditions of model

Model	ACF	PACF
AR(p)	Tail	Post truncated p-order
MA(q)	Q-order truncation	Tail
ARMA(p,q)	Q-order trailing	P-order trailing

For any time series, when both ACF and PACF images show no trailing state, regardless of whether the images are truncated or not, the time series is applicable to the ARIMA model. At this time, ACF and PACF images cannot help us determine the specific values of p and q, but it can be confirmed that both p and q are definitely non-zero.

Taking the monthly sales data of China's tobacco industry as an example, the ARIMA model of SPSS tool is used for monthly sales forecasting. Firstly, the original time series is analyzed

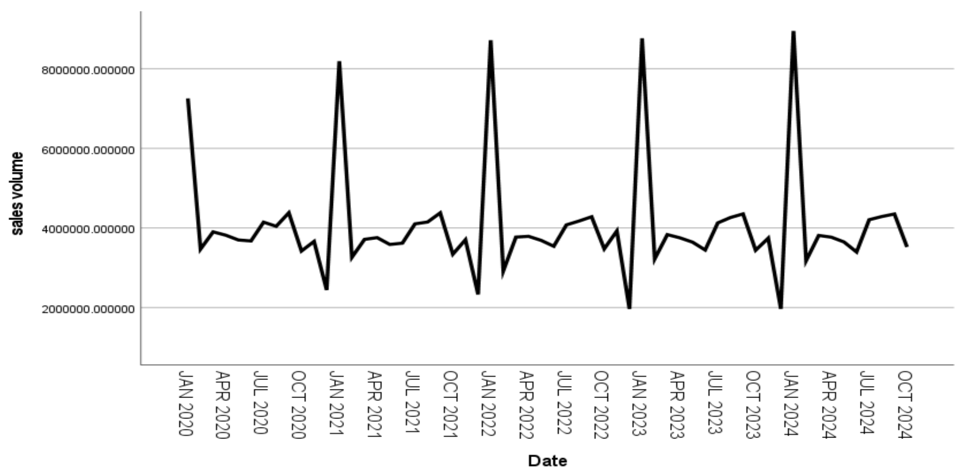

Figure 1- Original time series of monthly sales

Analyze sales indicators based on sequence diagrams and determine that the original time series is non-stationary. If the same sequence does not meet the stationarity condition, differential stationarity processing is performed. SPSS is used to analyze the autocorrelation ACF and partial autocorrelation PACF, and the autocorrelation and partial autocorrelation graphs are obtained, both of which are tailing

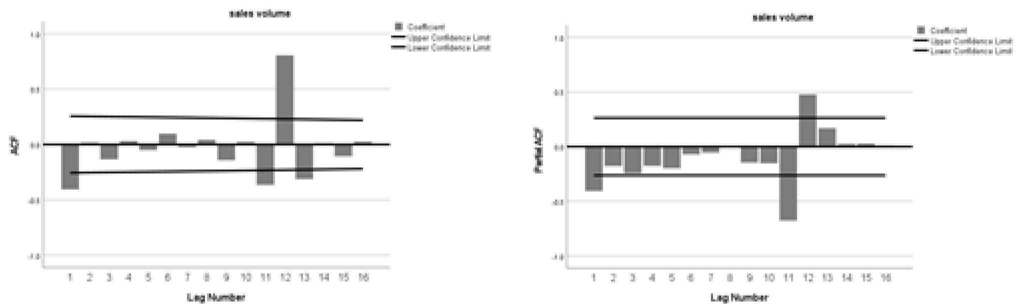

Figure 2- Original time series ACF and PACF plots

The monthly sales difference series is basically evenly distributed on both sides of the 0 scale, indicating that the original time series is non-stationary. The coefficients fluctuate in a small range around the zero axis, so it can be considered that the seasonal 1st order difference sequence is stationary, resulting in a comprehensive difference order d=1.

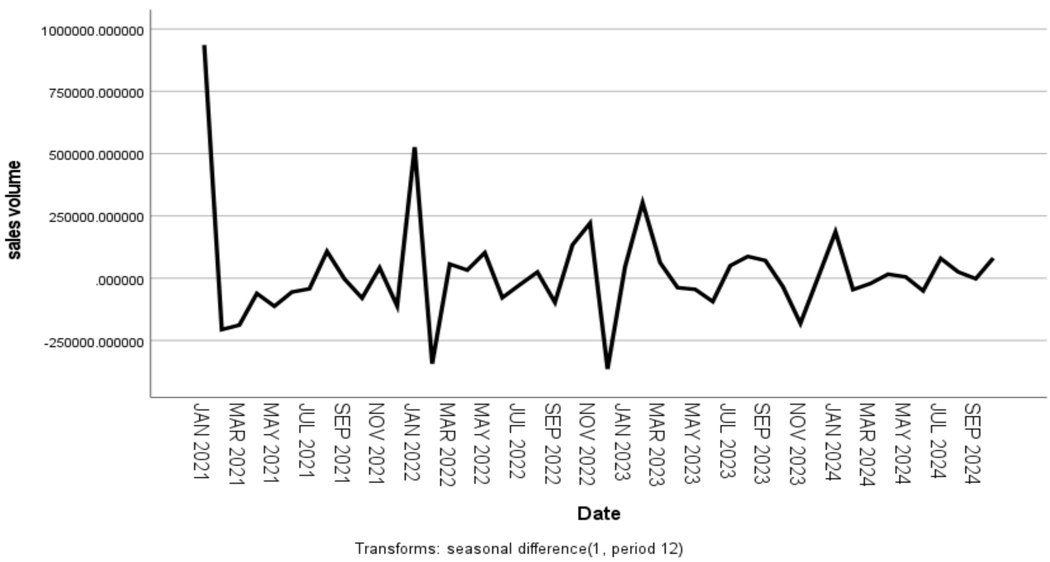

Figure 3 -Shows the sequence diagram of the first-order differential sequence.

Looking at the ACF and PACF plots of the residuals again, it can be seen that the seasonal first-order differences are stationary.

From the above figure, it can be seen that the ACF and PACF plots of the original sequence after first-order differencing are both first-order tails. Therefore, the first-order differenced sequence is a stationary sequence, and finding appropriate parameters is a process of repeated attempts, where p and q are both equal to 1. Therefore, further analysis can be conducted through differencing, such as analyzing the autocorrelation ACF and partial correlation PACF plots of non-stationary series sales. The ACF and PACF plots of the differenced series are both trailing. Therefore, a seasonal ARIMA (p, 1, q) model can be established for the original series, and after repeated experiments, the model is determined to be ARIMA (0,0,0) (1,1,1).

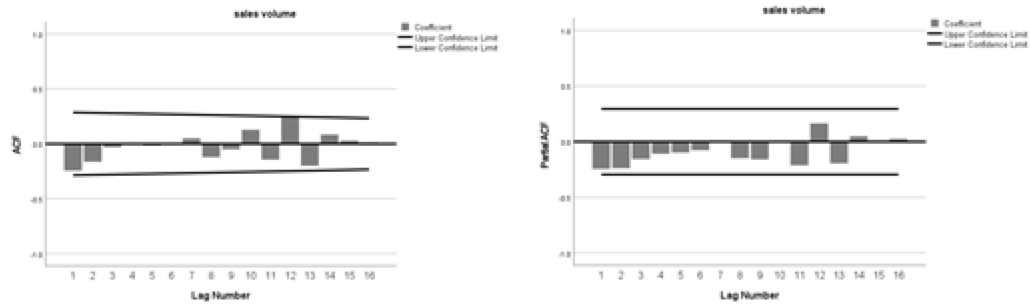

Figure 4- ACF and PACF plots of first-order differential time series

Table 3-Model Description

			Model Type
Model ID	sales volume	MODEL_1	ARIMA(0,0,0)(1,1,1)

Based on the previously determined parameters p, d, and q, the ARIMA model (0,0,0) (1,1,1) was determined and modeled in SPSS. The model statistics are shown in Table 3.

Table 4-Model Statistics

Model	**Number of Predictors**	**Model Fit statistics**		**Ljung-Box Q(18)**			**Number of Outliers**
		Stationary R-squared	**R-squared**	**Statistics**	**DF**	**Sig.**	
sales volume-MODEL_1	0	.052	.983	12.147	16	.734	0

The fixed stationary R-squared statistic is an estimate of the proportion of total variation explained by the model in the sequence, and the higher the value (maximum value of 1.0), the better the model fit. Based on the table, the model fitting effect is very perfect, with a Stationary R-squared value of 0.052 and R-squared value of 0.983.

Table 5-ARIMA Model Parameters

					Estimate	**SE**	**t**	**Sig.**
sales volume-MODEL_1	sales volume	No Transformation	Constant		18082.285	41588.468	.435	.666
			AR, Seasonal	Lag 1	.664	.304	2.183	.035
			Seasonal Difference		1			
			MA, Seasonal	Lag 1	.345	.406	.849	.400

After all parameter settings are completed, the model displays the results. The R-squared reaches 0.983, indicating a good fit. The coefficients for seasonal AR and MA are 0.664 and 0.345, respectively. The significance levels of AR autoregression and MA moving average are 0.035 and 0.400, respectively. If the significance value is greater than 0.05, the difference is not significant

Table 6-Forecast

Model		**Nov 2024**	**Dec 2024**	**Jan 2025**	**Feb 2025**	**Mar 2025**	**Apr 2025**
sales volume-MODEL_1	Forecast	3722620.151597	1933118.546583	9058780.707001	3190698.519491	3821889.611644	3779852.887151
	UCL	4108068.255443	2318566.650429	9444185.555118	3576103.367609	4207294.459762	4165257.735269
	LCL	3337172.047750	1547670.442737	8673375.858883	2805293.671373	3436484.763526	3394448.039033

Based on the output results of the model, the upper control line (UCL) and lower control line (LCL) values are used to obtain the predicted sales volume from November 2024 to April 2025.

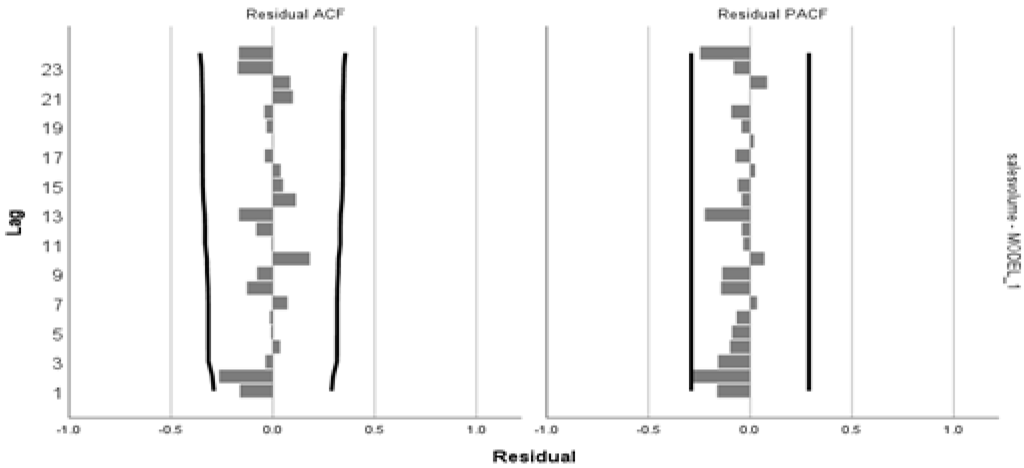

Figure 5-ACF and PACF plots of sales Amount residuals

After constructing the AIRMA model, it is generally required that the model residuals be white noise, meaning that there is no autocorrelation in the residuals. The ACF and PACF plots of the sales residuals can be seen to be stationary, so ARIMA (0,0,0) (1,1,1) is reasonable.

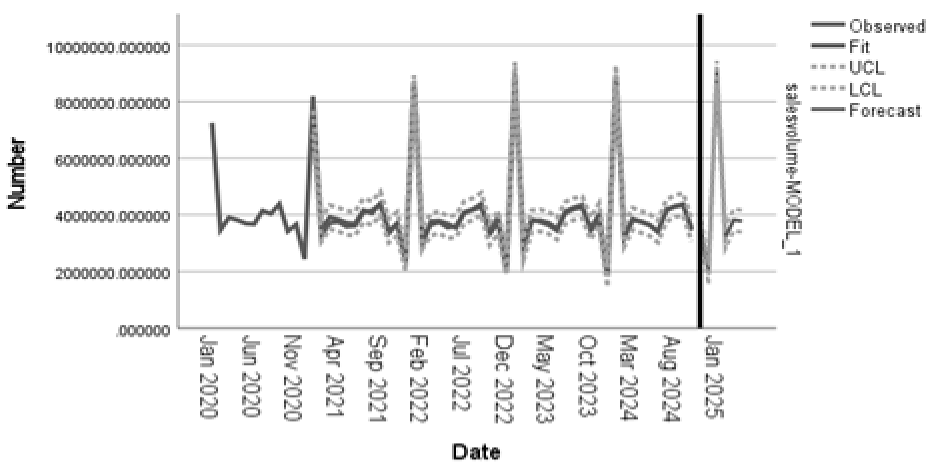

Figure 6- Monthly Sales Forecast Fitting Chart

Figure 6 shows the fitting graph of the ARIMA model, with a good overlap between the fitted values and observed values. It is also believed that the residuals of this sequence follow a random sequence distribution, and there are no outliers present. Therefore, the ARIMA model fits the sales volume well. The final part is the prediction curve for the sales data from November 2024 to April 2025.

Conclusion

This article uses ARIMA for modeling and prediction. In the process of analyzing and predicting, the modeling data first needs to meet the condition of stationarity. If it does not meet the condition, it needs to be processed to pass the test, and differential processing is carried out to determine the differential order. Then, based on the ACF and PACF graphs analyzed by SPSS tools, the parameters q, d, and p of the ARIMA prediction model are determined. Then, the model needs to pass the significance test of the parameters and the white noise test of the residuals to increase its credibility. The basic idea of ARIMA model is to establish a model that can describe the characteristics of time series data through autoregression, moving average, and differencing transformations, and use this model to predict future data changes.

References

1. Zhang J. (2024) Proekt empiricheskoi shkal'y tsifrovoi transformatsii proizvodstvennykh predpriyatii Kitaya [Project of empirical scale of digital transformation of manufacturing enterprises in China]. Ekonomika i Upravlenie: Problemy, Resheniya [Economics and Management: Problems, Solutions], 2024. DOI: 10.36871/ek.up.p.r.2024.02.03.005
2. Meshram D. (2012) Seasonal Arima Model for Pomegranate Evapotranspiration [Seasonal Arima Model for Pomegranate Evapotranspiration]. LAP Lambert Academic Publishing.
3. Lindenberg S.K. (2012) ARIMA vs. Spectral Analysis [ARIMA vs. Spectral Analysis]. AV Akademikerverlag.
4. Ayodele A.A., Adewumi A.O., & Ayo C.K. (2014) Stock price prediction using the ARIMA model [Stock price prediction using the ARIMA model]. In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.
5. Fernández-Manso A., Quintano C., & Fernández-Manso O. (2011) Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale [Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale]. International Journal of Remote Sensing, 32(6), 1595-1617.
6. Zhang G.H. (2019) Analysis and Prediction of Precipitation Trend in Weinan City Based on ARIMA Model [Analysis and Prediction of Precipitation Trend in Weinan City Based on ARIMA Model]. Value Engineering, 34, 197-199.
7. Zhang Y.J. (2019) Based on ARIMA model Simple comparison analysis with index prediction results [Based on ARIMA model Simple comparison analysis with index prediction results]. Marketing Research, 11, 23-26.
8. Zheng M.G., & Li Q. (2020) Scenario prediction of China's oil resource demand in 2020-2030 [Scenario prediction of China's oil resource demand in 2020-2030]. Advances in Earth Science, 35(3), 286-296.
9. You Y.L., Li J., Gao S.Y.J., et al. (2020) Application of the combination of ARIMA model and SVM model in the prediction of infectious diarrhea [Application of the combination of ARIMA model and SVM model in the prediction of infectious diarrhea]. Journal of Medical Pest Control, 5, 432-435.
10. Banjul B., Bhattacharyya R.D., et al. (2014) Forecasting of Maize Production in Andhra Pradesh by ARIMA Modeling [Forecasting of Maize Production in Andhra Pradesh by ARIMA Modeling]. Environment & Ecology.
11. Yan X.X. (2020) Using ARIMA Model to Predict Green Area of Park [Using ARIMA Model to Predict Green Area of Park]. Computer Science, DOI:10.11896/jsjkx.200300099, 47(S2), 531-534+556.
12. Demir V., Zontul M., & Yelmen I. (2020) Drug Sales Prediction with ACF and PACF Supported ARIMA Method [Drug Sales Prediction with ACF and PACF Supported ARIMA Method]. In 2020 5th International Conference on Computer Science and Engineering (UBMK). DOI:10.1109/UBMK50275.2020.9219448

Модель прогнозирования ARIMA для ежемесячного объема продаж продукции на производственных предприятиях

Чжан Цзяньхуа

Кандидат экономических наук
Институт ЭМИТ,
Российская академия народного хозяйства
и государственной службы при Президенте РФ,
119571, Российская Федерация, Москва, пр-кт Вернадского, 82;
e-mail: zhang_jianhua@vk.com

Аннотация

В статье рассматривается модель ARIMA (Autoregressive Integrated Moving Average Model) как ключевой инструмент для прогнозирования экономических и финансовых временных рядов. Основное внимание уделяется принципам работы модели ARIMA, которая оптимизирует планирование производства и управление запасами, учитывая сезонные колебания и долгосрочные тенденции продаж на основе данных за пятилетний период. Используя SPSS для анализа временных рядов, была разработана модель ARIMA, прошедшая тесты на плавность данных и белый шум, что позволило установить окончательные параметры модели (p, d, q). Модель ARIMA демонстрирует высокую точность прогнозирования и простоту в использовании, что делает её незаменимым инструментом для анализа временных рядов.

Для цитирования в научных исследованиях

Чжан Цзяньхуа. ARIMA forecasting model for monthly sales volume in manufacturing enterprises. // Экономика: вчера, сегодня, завтра. 2024. Том 14. № 10A. С. 291-299.

Ключевые слова

ARIMA; экономический прогноз; СПСС; разница первого порядка; серийные диаграммы; модель временных рядов.

Библиография

1. Чжан Цзяньхуа.Проект эмпирической шкалы цифровой трансформации производственных предприятий Китая // Экономина и управление: проблемы, решения. 2024.// DOI: 10.36871/ek.up.p.r.2024.02. 03. 005
2. Meshram, Deodas.Seasonal Arima Model for Pomegranate Evapotranspiration[M].LAP Lambert Academic Publishing,2012(05)
3. Lindenberg, Sonja Katajun.ARIMA vs. Spectral Analysis[M].AV Akademikerverlag,2012(07)
4. Ayodele Ariyo Adebiyi, Aderemi Oluyinka Adewumi, Charles Korede Ayo.Stock price prediction using the ARIMA model[C].//2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014.
5. A.Fernández-Manso, C.Quintano, O.Fernández-Manso.Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale[J].International Journal of Remote Sensing, 2011, 32（6）：1595-1617.
6. ZHANG G H.Analysis and Prediction of Precipitation Trend in Weinan City Based on ARIMA Model[J]. Value Engineering, 2019(34):197-199.
7. ZHANG Y J.Based on ARIMA model Simple comparison analysis with index prediction results[J.Marketing Research,2019(11);23-26.
8. ZHENG M G,LI Q. Scenario prediction of China's oil resource demand in 2020-2030[J].Advances in Earth Science,2020,35(3):286-296.
9. YOU Y L,LI J,GAO S Y J,et al. Application of the combination of ARIMA model and SVM model in the prediction of infectious diarrhea[J].Journal of Medical Pest Control ,2020(5):432-435.
10. Banjul,Bhattacharyya, Ramesh D ,et al.Forecasting of Maize Production in Andhra Pradesh by ARIMA Modeling[J].Environment & Ecology, 2014.
11. YAN X X.Using ARIMA Model to Predict Green Area of Park[J].Computer Science ,DOI:10.11896/jsjkx.200300099,2020,47(S2):531-534+556,
12. Demir V , Zontul M , Yelmen I .Drug Sales Prediction with ACF and PACF Supported ARIMA Method[C]//2020 5th International Conference on Computer Science and Engineering (UBMK).2020.DOI:10.1109/ UBMK 50275. 2020.9219448.