

Forecasts of Video Game Sales based on the ARIMA Model

Siyang Yang

College of Art, Tianjin University of Finance and Economics, Tianjin, 300000, China

yangsiyang@tjufe.edu.cn

Abstract. With the rapid development of the digital age, video games have become an important component of the cultural industry, and their sales forecasting has become a focus of industry attention. This article aims to predict the sales of video games based on the ARIMA model and explore effective methods to improve prediction accuracy. Through in-depth research, this article found that the ARIMA model performs well in capturing the time series characteristics of sales data and can accurately predict future sales trends. The ARIMA model provides a systematic and scientific method for predicting the future market performance of video games by comprehensively considering the trends, seasonality, and random changes in historical sales data. This research achievement not only provides strong decision support for game developers, publishers, and platform operators but also promotes the optimization of resource allocation and market risk avoidance in the video games industry, which is of great significance for promoting the healthy and sustainable development of the industry.

Keywords: Video game, ARIMA model, Time series.

1. Introduction

In the tide of the digital age, video games, as a shining pearl of the cultural and entertainment industry, are influencing and shaping our way of life at an unprecedented speed. With the rapid development of technology, the exquisite graphics, diverse gameplay, and realistic interactive experience of games have reached unprecedented heights. This has made video games not only the first choice for leisure and entertainment among players but also an important force in promoting the development of the cultural industry[1]. Although technological advancements have driven the development of the industry, they have also made market competition more intense, further increasing the difficulty of predicting game sales. Therefore, predicting game sales has become a common focus of attention for game developers, publishers, and platform operator.

The prediction of video game sales is not only related to the economic benefits and market position of enterprises, but also a comprehensive test of game creativity, quality, and market acceptance. Through in-depth mining of historical data, keen insight into market trends, and detailed analysis of player behavior, we can reveal the key factors that affect game sales and construct scientifically reasonable prediction models based on them. These models can not only provide strong support for enterprise decision-making, help optimize resource allocation and avoid market risks but also promote the development of the entire video gaming industry towards a healthier and more sustainable direction[2].

Therefore, this article aims to explore the methods and strategies for predicting the sales of video games, based on the ARIMA model to predict the sales of video games, and provide useful references

and guidance for the future development of the video game industry. We believe that with the continuous deepening of research and the continuous advancement of technology, the prediction of video game sales will become more accurate and efficient, contributing greater strength to the prosperity and development of the entire industry.

2. Data and Methodology

2.1. Data Sources

The data used in this article is sales, which is sourced from Kaggle. The sales unit used in the article is millions of copies. This article presents a statistical analysis of video game sales from 1970 to 2024, obtaining complete time series data in years. This dataset includes 63989 observations and 8 fields.

2.2. Methods

Autoregressive Integral Moving Average Model (ARIMA) is a statistical model widely used for time series data analysis and prediction. It was proposed by Box and Jenkins in the 1970s and is widely regarded as one of the most effective methods in time series prediction. The ARIMA model combines the characteristics of the autoregressive (AR) model, difference (I) model, and moving average (MA) model, and can capture various dynamic characteristics in time series data[3].

Due to the requirement of the ARIMA model for data to be a stationary sequence, it is necessary to perform stationarity tests on the original time series data[4]. If the sequence does not pass the stationarity test, differential processing is required to make the data stationary[5,6]. Stable data has the following characteristics: stable mean, and the mean of the sequence does not change over time; Variance stability; Autocorrelated stability; No trend; Having periodicity; Predictability.

The original time series does not meet the requirement of stationarity (Figure 1). After second-order differencing, the P value is 0.0019707012085912463, almost approaching 0. At the same time, according to the ACF and PACF graphs (Figure 2 and Figure 3), it shows significant tailing and truncation, which meets the requirements of time series modelling. Therefore, establishing an ARIMA model based on this data for predicting subsequent sales data has a certain reference significance.

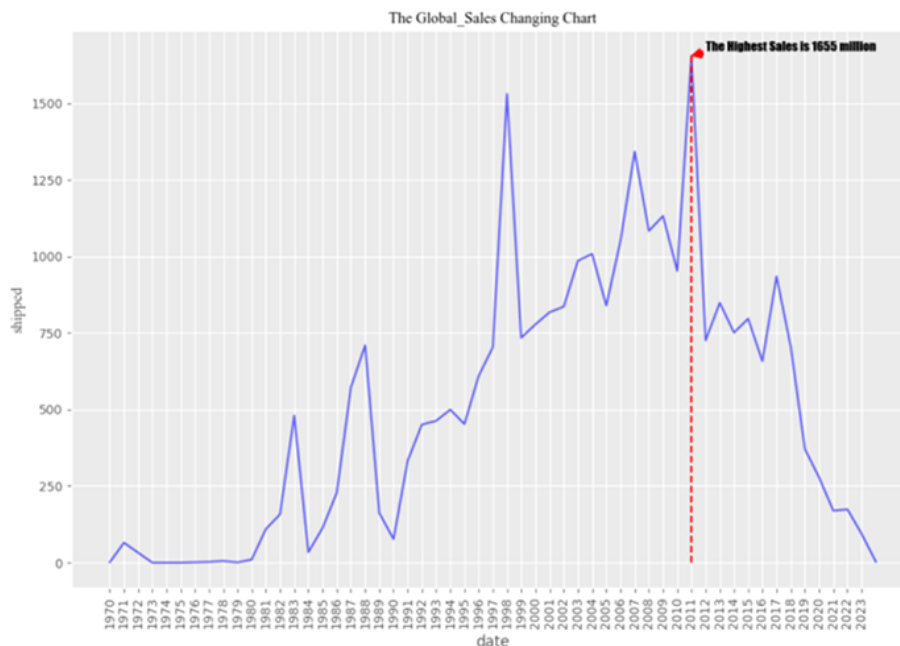


Figure 1. The Global Sales Changing Chart.

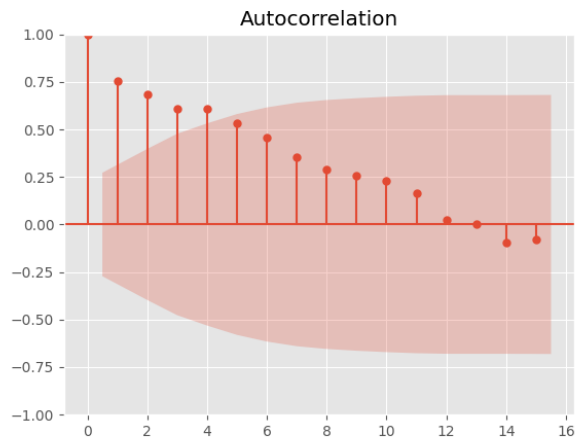


Figure 2. Autocorrelation diagram of second-order difference.

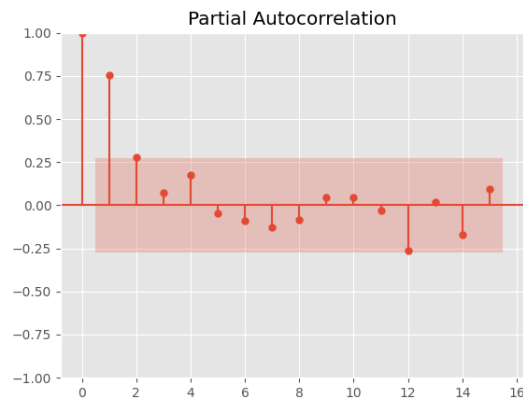


Figure 3. Partial autocorrelation diagram of second-order difference.

3. Results

3.1. Forecast of video game sales based on the ARIMA model

Figure 4 shows the trend changes in raw and predicted data of video game sales from 1970 to 2024. It can be seen that comparing the differences between the two-time series can provide intuitive references for analyzing data trends, predicting future trends, or studying the interrelationships between data [7,8].

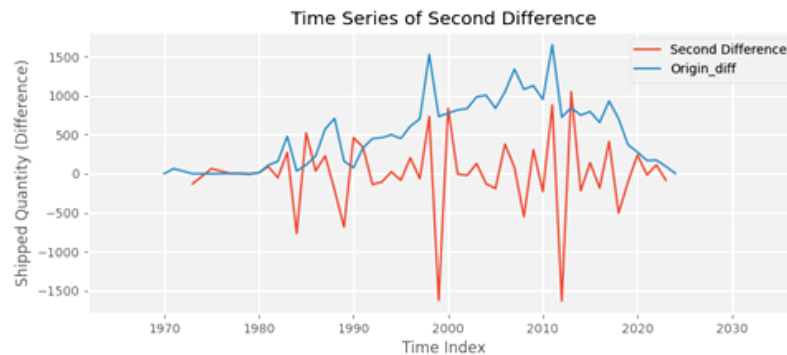


Figure 4. Time Series of Second Difference

Both ACF and PACF plots show that all coefficients are close to 0 and there is no significant tailing or truncation phenomenon, indicating that the residual sequence is already white noise (Figure 5). This means that the ARIMA model has successfully extracted all predictable information from the original time series, and the residual sequence no longer contains any predictable patterns or trends, further demonstrating the applicability of the model[6].

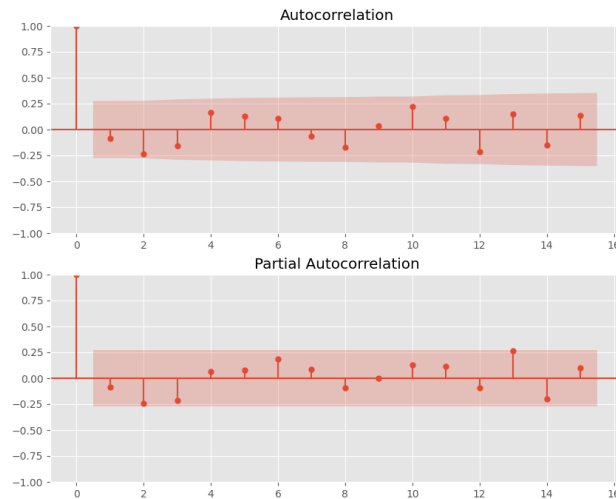


Figure 5. Autocorrelation plots and partial autocorrelation plots of residual sequences

As shown above, in the prediction of video game sales, the ARIMA model can analyze the cyclical, trending, and stochastic components in historical sales data, and thus construct an accurate prediction model. By adjusting model parameters, ARIMA can flexibly adapt to the characteristics of different game sales data, providing powerful market forecasting support for the gaming industry and helping enterprises make scientific decisions and plans.

This article applies the ARIMA model to conduct an in-depth analysis of time series data on video game sales. By carefully selecting model parameters (including autoregressive term p , difference order d , and moving average term q), we have successfully constructed a highly fitted prediction model[7,8]. The results show that the model can not only accurately capture seasonal fluctuations and long-term trends in sales data, but also effectively handle short-term random changes [9,10].

According to the forecast data, the sales of video games will continue to grow. In the subsequent development of the video gaming industry, enterprises can use predictive data to conduct in-depth research on market dynamics and adjust marketing strategies accordingly to attract consumers.

4. Conclusion

In the field of exploring video game sales forecasting, the ARIMA model, as a classic and powerful time series analysis tool, has demonstrated its unique value and potential. The ARIMA model provides us with a systematic and scientific method to predict the future market performance of video games by comprehensively considering the trends, seasonality, and random changes in historical sales data.

This study not only validates the effectiveness of the ARIMA model in predicting video game sales but also reveals the complex dynamic relationships behind sales data. By finely adjusting the parameters of the model (such as autoregressive term p , difference order d , moving average term q), we can more accurately capture subtle differences in sales changes, thereby improving the accuracy of predictions.

Looking ahead, with the booming development of the esports industry and the increasing diversification of player preferences, predicting the sales of video games will become more complex and crucial. The ARIMA model, as a foundation, can be further combined with other advanced data analysis techniques such as machine learning and deep learning to form a more comprehensive and

powerful prediction system. This will help game developers, publishers, and market analysts better understand market dynamics and develop more scientific and reasonable market strategies and product plans.

In summary, the research on video game sales forecasting based on the ARIMA model has opened a window for us to gain insights into the market and predict the future. With the continuous advancement of technology and the accumulation of data, we have reason to believe that research in this field will continue to make breakthroughs, contributing more wisdom and strength to the prosperous development of the video games industry.

References

- [1] Zhu Y Y, Sun Y L and Lee G 2022 Visualization analysis of video game research in China in the past ten years based on CiteSpace Southeast Communication 02 pp 137-140
- [2] Yao Q F, Tan B and Xiong Y H 2005 Research on the development trend of China's online game market Information Networks 09 pp 51-54
- [3] Wang Y M, Yin W Q and Yang L 2024 Multiplicative seasonal model based on ARIMA for prediction of prevalence trend of carbapenem-resistant *Acinetobacter baumannii* Chinese Journal of Nosocomiology 17 pp 2575-2579
- [4] Zhang L 2008 Study of the Time Series Forecasting Algorithm and System Realization Based on the ARIMA Model Doctoral dissertation, Jiangsu University
- [5] Xiang, Y, Lee S C and Li X 2005 The variables of effecting customer loyalty in chinese online game market IEEE
- [6] Liu F, Wang R J and Li C X 2009 Application of ARIMA model in agricultural product price forecasting Jisuanjii gongcheng yu yingyong 45(25) pp 3
- [7] Tang Y 2007 Research and application of time series analysis Master's thesis Northeast Agricultural University
- [8] Zhuang G X, Zhang H and Liu X 2013 The Economic Trend of Video Game Industry CoRR
- [9] Wang Z, Tang J K, Hou S S, Wang Y J, Zhang A A, Wang J R and Han B 2023 Landslide displacement prediction from on-site deformation data based on time series ARIMA model Frontiers in Environmental Science
- [10] Tang D 2021 Innovation in the British Video Game Industry since 1978 Business History Review 3 pp 543-567