Academic Journal of Engineering and Technology Science
ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

Predicting Sales Data with Holt-Winters Method

Junning Li

School of Management, Shanghai University, Shanghai 200444, China
18621661843@163.com

ABSTRACT. *This article selects the sales data of a store that somewhere a down jacket from January 2001 to December 2007, through the R language for data processing and analysis, using the Holt - Winters exponential smoothing method to fitting and forecasting of time series, through the prediction error of time curve and histogram, Holt - Winters exponential smoothing method for the sales data of down jacket store provides a suitable prediction model. It provides decision-making information for the later operation and sales of the store.*

KEYWORDS: *time series analysis, Holt - winters prediction, auto correlation*

1. History of time series analysis

The use of time series data dates back as far as 7,000 years ago in ancient Egypt, where the Nile's fluctuations were recorded day by day. According to the recorded data, the ancient Egyptians found that the Nile rose and fell regularly. They found that the Nile began to flood about 200 days after Sirius first rose at the same time as the sun. The Babylonian astronomers predicted the astronomical events according to the data sequence of the relative positions of stars and satellites. After decades of observation and recording, German amateur astronomer and pharmacist Schwabe finally found that sunspot activity has a cyclical rule of about 11 years [1]. These are the earliest use of visual sequence data. These time series analysis is usually through the intuitive data comparison or mapping observation, looking for the development law contained in the sequence, this analysis method is called descriptive time series analysis. The method is simple, intuitive and easy to understand and still one of the commonly used method, but it is strong subjectivity, conclusion raw sugar, as research in the field of gradually widened and the research of complicated problems, the simple descriptive analysis can no longer meet the needs of the development of the random variables in probability theory and mathematical statistics put forward some conclusions and methods, the research emphasis has shifted from the summary on the phenomenon of surface gradually on the inherent nature of random sequence correlation analysis [2], of course, the development of statistical timing analysis is not plain sailing, but it is a long process, full of contradictions, the focus of the

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

problem mainly focus on the different understanding of the role of random motion, the classical approach doesn't make any sense to argue that these random motion is residual, modern methods according to their impact on the time series of all components are random, the time series of movement should be regarded as a random process, the analysis of time series is just a result of the data generating process [3].Therefore, contradictions gradually converted to explore random items and related structure of complex between them, in this context, in other words, it is in the process of this development, gradually gave birth to a series of basic concepts and analysis methods, such as, difference index and moving average, fluctuation, the percentage deviation and time series decomposition and thought, of course, in turn, statisticians development and use of these technological tools, has a certain extent to stimulate the higher and more advanced analysis methods, eventually led the formation of modern time series analysis.

Modern time series analysis methods mainly include frequency domain analysis and time domain analysis. In 1906, the periodogram method proposed by the German scholar schuster was the beginning of the frequency domain analysis. In 1927, the model initiated by the British statistician yule was the origin of the time domain analysis method. In fact, many indicators in the natural world and social economy can be regarded as time series data over time, such as the number of sunspots, a country's gross national product, consumer price index, price index, and even the sales volume of certain products and carbon dioxide emissions in the atmosphere. So, as a basic method of processing data, time series analysis discipline has developed rapidly, becoming mature, especially with the use of computer technology and the rising popularity of relevant statistical software, the subject has been widely applied to psychology, meteorology, hydrology, biological medicine, management, earthquakes and military, and many other fields. In particular, the combination with the financial field opens up the direction of financial time series analysis [4].

2. The case study of Time series analysis

This article selects the monthly sales data of a down jacket specialty store from January 2001 to December 2007. R language is used to analyze the data, and the results are shown in table 1:

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

Table 1 sales data of the store unit: yuan

Draw the time series diagram according to the obtained data:

	Jan	Feb	Mar	Apr	May	Jun
2001	120648.1	123975.3	28407.1	35472.9	37529.6	37147.4
2002	100218.2	151982.4	47170.2	48060.3	54964.3	49513.4
2003	152251.4	157026.3	59211	53047.8	64924.3	66308
2004	201766.2	158145.8	48266.4	63697.7	76091.2	72247.5
2005	200930.6	164702.3	76150.3	88211.7	87223.7	102094.8
2006	256373.9	198496.9	102432.4	115873.3	93325.6	130820.9
2007	605584	112668.8	186015.3	159977.9	173573.3	261551.5

	Jul	Aug	Sep	Oct	Nov	Dec
2001	35663.4	43496.1	64234.8	24998.1	76006	197562.1
2002	59008.8	61791.2	68351	47521.5	126000.8	285417.2
2003	96905	99575.8	85731.7	73496.2	151518.4	340610.1
2004	81212.2	79792.5	124212.5	84767	179146.6	301144.1
2005	112765.5	125522.2	136068.9	96387.7	218221.1	450606.9
2006	167327.8	239333.8	253913.5	198886.1	360248	457217.1
2007	285865.2	308213.3	305054.1	218268.4	466343.8	706606.7

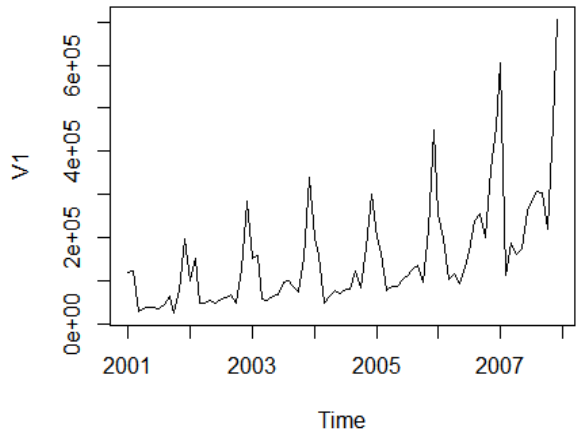

Figure. 1 The sales data of the store

After analyzing the data, it can be found that the seasonal fluctuations and random changes presented in the figure 1 above gradually increase with the time series. Therefore, the additive model is not suitable to describe this time series. The time series need to be transformed to obtain a time series that can be described by the additive model.

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

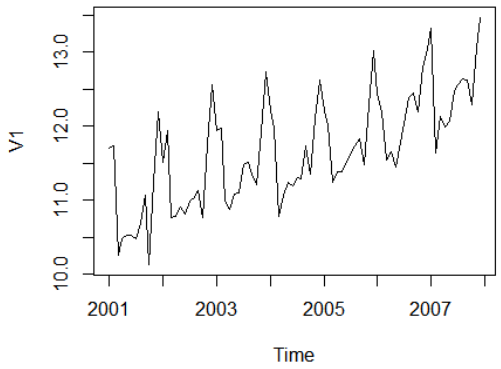

Figure. 2 the sales data after taking the natural logarithm

It can be seen from figure. 2 that after the time series in figure. 1 achieved logarithmic transformation, it is roughly constant with the passage of time in terms of the magnitude of seasonal fluctuation and random fluctuation, and do not depend on the level of time series. Therefore, the logarithmic transformed time series may be described by an additive model.

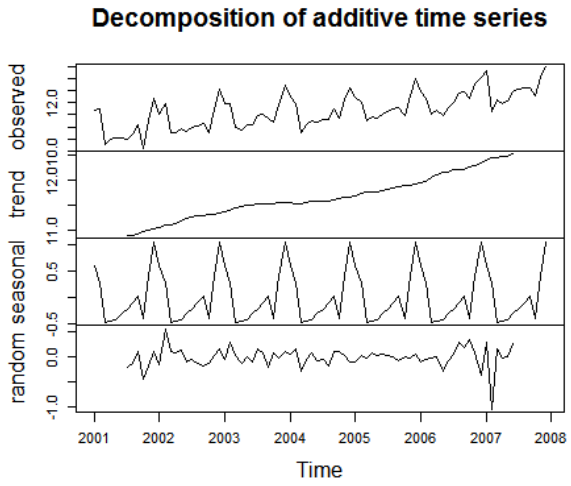

Figure. 3 the trend chart of sales data

Figure 3 shows the original time series (the top part), estimated trend (the second part), estimated seasonality (the third part), and estimated irregularity (the bottom part), respectively. According to the seasonal trend, due to the return season sales of down jackets in specialty stores, the sales volume of down jackets in specialty stores generally increases in July and August every year, which is reflected in the trend

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

chart in the form of small peak sales volume. The annual peak of down jacket sales is roughly between November and January of the following year, which is related to the sufficient demand of consumers for down products in winter. Between February and May of each year, the sales volume of down products is always at a low level, during which consumers generally reduce the demand for them.

The Holt-winters exponential smoothing method can be used to make short-term predictions based on the fact that sales data are a time series with an increasing trend and seasonality that can be described as an additive model.

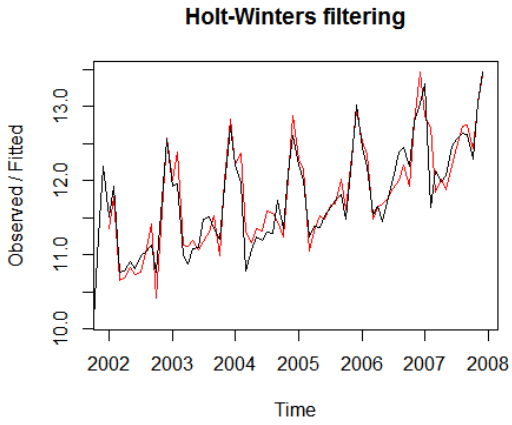

Figure. 4 results after data correction

The black line in figure 4 is the time curve of the original data, and the red line is the time curve of the predicted value. From the figure, holt-winters' exponential smoothing method successfully predicted the seasonal peak, which occurred approximately in December every year while there would be a small peak from July to August every year.

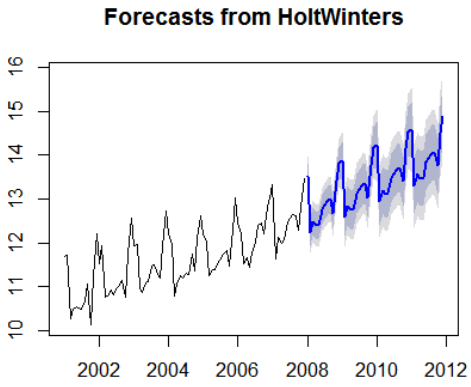

Figure. 5 prediction diagram

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

To predict the future of the time series, use the "forecast.Holt Winters ()" function in the "forecast" package. The original figures for sales in down jacket specialty stores were from January 2001 to December 2007. Based on the above existing data, the data from January 2009 to December 2012 are predicted, and the prediction chart is made, as we can see in figure 5.

The blue line shows the forecast, while the gray and light gray shades are 80 and 95 percent, respectively. Ljung-box test was conducted to check whether the prediction error in the sample was non-zero auto correlation when the delay was 1-20 orders, and to determine whether the prediction model could be optimized again.

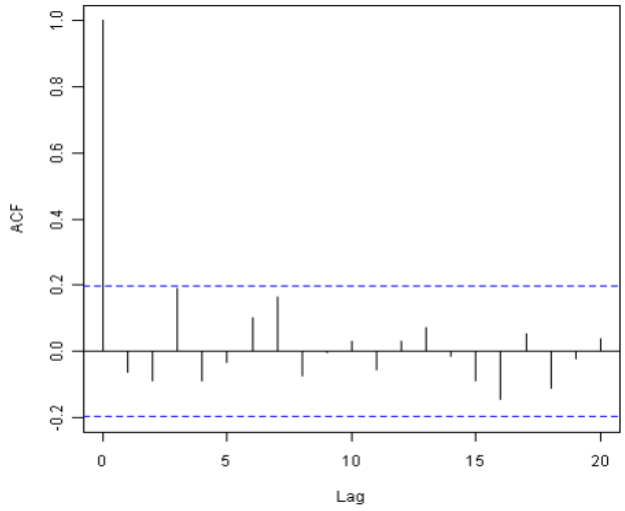

Figure. 6 Lag

The correlation graph of the prediction error in the sample did not exceed the confidence limit of the auto correlation coefficient in the order of 1-20 delay. Moreover, the P value of the ljung-box test was 0.6, which means that it is not sufficient to prove that the auto correlation of the order 1-20 delay is non-zero.

Test whether the prediction error is variance invariant and obeys the normal distribution of zero mean over the whole time period. Make the time curve and histogram of the prediction error.

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

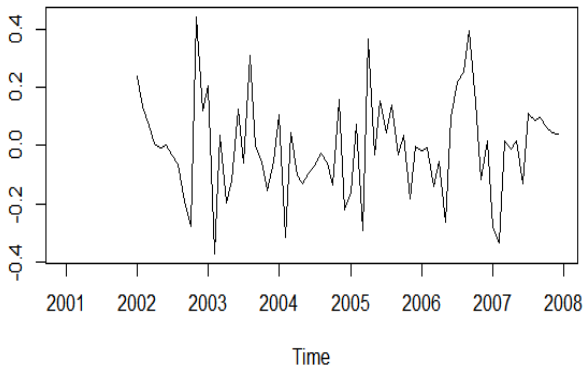

Figure. 7 time curve of prediction error

As can be seen from figure 7, the variance of the prediction error is invariant in the whole time period. As can be seen from the figure, it seems that the prediction error is normally distributed with zero mean.

Therefore, this is not enough to prove that the prediction error is auto correlated in the order of 1-20 of delay, and the prediction error presents a normal distribution with zero mean and invariant variance in the whole time period. This suggests that Holt-winters' exponential smoothing provides a suitable prediction model for the sales data of down jacket stores, and that the assumptions based on the prediction intervals are reasonable under the assumptions.

3. Conclusion

This paper introduced the history of time series, on the basis of time series analysis theory, the use of time series model on the annual sales of a store data analysis, and on the basis of the later sales forecast, through the test. Similar to the direct result analogy, it can be said that the analysis of time series on this data is successful, which can provide decision-making information for the store's later operation and sales plan.

According to the information and my own understanding about time series, the future research on time series may develop in the following directions:

1. Network is more and more developed to enter 5 g era, everyone is the producer of the data time series data will also increase a lot, but the most massive amounts of time series data is steady, the characteristic parameters and mathematical distribution is change over time, the model must track the change to adapt to the current data and predict the future;

2. As people require more and more accurate and in-depth discussion of practical problems, it has become inevitable to adopt the modern time series method of time-

ISSN 2616-5767 Vol. 2, Issue 1: 13-20, DOI: 10.25236/AJETS.020010

varying parameter model and adaptive prediction technology in the actual prediction work;

3. In order to study the dynamic relationship between different variables and the development of computer hardware, although ARMA model is the basis, more and more models will be built based on it, and the application of multiple models in vector ARMA model or state space model will increase day by day;

4. BP neural network model and a series of other intelligent algorithms will also increasingly apply to the time series analysis, and the synergy of various models can better reflect the change trend of the time series.

References

[1] Nie shuyuan. Early development of time series analysis [D]. Northwest university, 2012.
[2] Li xiaoling. Prediction based on time series analysis [D]. Sichuan University, 2006.
[3] Nie shuyuan. Historical development of time series analysis [J]. Journal of Guangxi university for nationalities (natural science edition), 2012 (01): 30-34.
[4] Luo fangqiong, Wu chunmei. Theory and application review of time series analysis [J]. Journal of Liuzhou normal college, 2009 (03): 119-123.