Forecast of Gross Output Value of Agriculture and Forestry in Guangxi Based on Holt-Winters Model

Run Li¹, Xiangmin Li¹, ZhiXiang Lu²,*

¹GuangXi International Business Vocational CollegeNaning, Guangxi, 530007, china
²School of Information and Engineering, Nanning University, Nanning, Guangxi, 530200, china

*luzhixiang2019@126.com

Abstract. Agriculture and forestry are the source of food, clothing and subsistence for human beings and also the primary conditions for all production. The forecast of the gross output value of agriculture and forestry directly affects the purchasing, pricing and storage of important plans in the agricultural sector. This paper collects quarterly data from 2012 to 2020 in Guangxi Zhuang Autonomous Region and uses the Holt-Winters prediction model based on the multiplication to fit the historical data. The model was validated by calculating the optimal smoothing coefficient through the minimum residual sum of squares, making predictions and comparing actual values for the data from 2018 to 2020. Finally, the model was used to derive future forecasts for agriculture and forestry from 2020 to 2022, and recommendations for agricultural development were made based on the forecasts.

Keywords: Holt-Winters, Guangxi agriculture, optimal smoothing coefficient, agricultural forecast.

1. Introduction

Guangxi has always been a large agricultural province, and its agriculture and forestry account for a large share of its gross product. With the development of agriculture and forestry, its gross output value is increasing, the information to be collected for data prediction is increasing, the influencing factors to be considered is also increasing, so the traditional method can no longer meeting the existing needs. Traditional methods are more difficult to model, and even if the model is established, it may not be possible to estimate the model parameters because of some missing data, thus leading to errors in forecasting. In general, for stationary or non-stationary time series, the one-parameter exponential, linear or curvilinear exponential smoothing models are more effective in forecasting, but for time series with both seasonality and trend, the above methods are basically ineffective. The gross output value of agricultural and forestry has a clear seasonal and trend pattern of variation.

TU Xing-yue(2014) [1] found a multiple linear regression forecasting model to predict soybean price trends from 1997 to 2011 to assist producers in planting and harvesting. However, soybean prices are annual aggregate data, which has a strong trend pattern and lacks seasonal patterns. Xiao Zheng(2017) [2] used gray prediction and multiple linear regression to develop a grain yield prediction model and a regression analysis model for pesticide utilization in Anhui province for the development of the agricultural circular economy. However, it has been mentioned by many scholars that gray prediction and multiple linear regression methods are not very accurate when faced with trend and seasonal data forecasts [3] [4]. Typical data with trend and seasonality are the number of tourists, Stanley Jere(2019) [5] used Holt-Winters additive model to accurately predict the number of tourists in Zambian attractions with a controlled error. However, he also mentioned that a larger sample data set is usually required for higher prediction accuracy. Jiang Weihe(2019) [6] pointed out that the Holt-
Winters multiplicative model has low requirements of training data and high prediction accuracy to accurately perform the prediction of electricity consumption data in rural areas. In recent times, this method has been continuously adopted to analyse data from different industries[7 – 9].

Therefore, in this paper, the Holt-Winters prediction model, which is very suitable for analysing data with obvious seasonality and trend, is used to fit the historical data and calculate the optimal smoothing factor through the minimum residual sum of squares to predict the gross output value of Guangxi agriculture and forestry in the coming years.

2. Data and methods

2.1. Model Introduction

The Holt-Winters model (also known as the cubic exponential smoothing model) is a forecasting model based on statistical time-series data proposed by the scholar Holt, which has the ability to better analyze data on forecast repeatability and seasonal trends. The Holt-Winters model contains two algorithmic models, the addition-based Holt-Winters and the multiplication-based Holt-Winters. Both of these two algorithms can predict time series that contain both trend and seasonality. The main difference is that the additive model is used if the trend and seasonal changes are superimposed in an additive form, and the multiplicative model is used if they are superimposed in a multiplicative form. After a preliminary exploration, it was found that the changes in the data are mainly superimposed in multiplicative form, so this paper focuses on the Holt-Winters multiplicative-based model.

2.2. The Holt-Winters multiplicative model

Formula for calculating the initial value of the multiplication model is as follows.

\[ S_t = \sum_{i=1}^{L} \frac{X_t}{L}, I_t = \frac{X_t}{S_t}, b_t = 0(t = 1, 2, \cdots, L) \]  

(1)

Basic formulas for multiplication models are as follows.

\[
\begin{align*}
S_t &= a \left( \frac{X_t}{I_{t-L}} \right) + (1-a)(S_{t-1} - b_{t-1}) \\
I_t &= \beta \left( \frac{X_t}{S_t} \right) + (1-\beta)I_{t-L} \\
b_t &= \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \\
F_{t+m} &= (S_t + mb_t)I_{t-L+m}
\end{align*}
\]

(2)

\( X_t \) is the observed value at time \( t \). \( S_t \) is the stable component at time \( t \). \( I_t \) is the seasonal component at time \( t \). \( b_t \) is the trend component at time \( t \). \( F_t + m \) is the predicted value at time \( m \). \( m \) is the predicted the number of periods. \( L \) is the season length. \( \alpha, \beta, \gamma \in [0,1] \) is the smoothing parameter. \( \alpha, \beta, \gamma \) take the principle that the root-mean-square error (RMSE) between the predicted and measured values is minimal.

The basic RMSE formula is as follows.

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{X}_i)^2} \]  

(3)
In the above equation, $X_t$ is the observed value at time $t$, $\hat{X}_t$ is the predicted value at time $t$. $n$ is the number of periods predicted.

According to the optimization theory, the loss function is defined as the minimum sum of squares of predicted errors, and the formula is shown as follows.

$$\min E = \min \sum_{i=1}^{n} e^2 = \min \sum_{i=1}^{n} (y_t - f_t^P) = \min \sum_{i=1}^{n} (y_t^2 - 2y_t \sum_{i=1}^{n} \omega_i f_i t + (\sum_{i=1}^{n} \omega_i f_i t)^2)$$

And meet the conditions

$$\begin{align*}
\sum_{i=1}^{m} \omega_i &= 1, i = 1, 2, \ldots, m \\
\omega_i &\geq 0
\end{align*}$$

3. Case analysis

The data is from the National Bureau of Statistics of China. The quarterly data on the gross output value of Guangxi's agriculture and forestry from the 1st quarter of 2012 to 2nd quarter of 2020 were collected from the data published on the website under "indicators" and "agriculture" and "the gross output value of agriculture, forestry, animal husbandry and fisheries".

The 1st quarter of 2012 to 2nd quarter of 2020 as the original data for fitting, and analyses the data from the 3rd quarter of 2018 to 4th quarter of 2020, compared with the actual value to verify the feasibility of the algorithm. Finally, the data from the 1st quarter of 2012 to 2nd quarter of 2020 is used to fit and predict the gross output value of Guangxi agriculture and forestry in the next two years.

4. Data results

4.1. Data pre-processing

The quarterly data of the gross output value of Guangxi's agriculture and forestry were serialized and plotted in figure 1. It can be seen from the figure that there are obvious cyclical changes in the data of agriculture and forestry, and they have a trend of linear increase.

Fig. 1 New smart campus system architecture.
4.2. Data smoothing

As shown in figures 2 and 3 below, a simple moving average of the data shows that the interval is greater than 1 (i.e., there is a seasonal factor), and seasonal fluctuations and overall trends need to be further investigated through seasonal decomposition.

Fig. 2 Data smoothing

Fig. 3 Simple moving average

4.3. Seasonal decomposition

As shown in figures 4 and 5, the seasonal decomposition shows that both trends are additive models (A), and both horizontal term and seasonal term are multiplicative models (M), i.e., using the (A, M, M) model.
4.4. Holt-Winters Exponential Smoothing

Holt-Winters is exponential smoothing with parameters for the level term, slope, and the season term. From the above steps, we can conclude that MAM's triple exponential model can be used to fit agriculture and analyse data from 2012 Q1-2018 Q2 and use it to analyse the data from 2018 Q3-2020 Q2 and compare them with the real values.

The comparison of the Holt-Winters prediction model for agriculture and forestry with the actual values is shown in table 1, where prediction error = |predicted value - actual value|/actual value × 100%.

From the comparative analysis of the predicted value and the actual value, it can be concluded that the random fluctuation of agricultural shares in 2019 2nd, 3rd quarters and 2020 2nd quarter will lead to low forecast accuracy.

From the other quarters to see most of the forecast error within 10%, the forecast effect is better, that there is only random error in the model fitting process, and no systematic bias. For the 85% confidence interval, the true value has a very high probability to fall around the prediction results, indicating that the model can be applied to the prediction of the output value of Guangxi agriculture and forestry.

Take the 1st quarter of 2012 to 2nd quarter of 2020 as the original data for fitting, and forecast the output value of Guangxi agriculture and forestry in the next 8 quarters, the forecast results are as follows tab. 2.

| Date     | Actual | Agricultural Forecast | Error   | Actual | Forestry Forecast | Error |
|----------|--------|-----------------------|---------|--------|-------------------|-------|
| 2018 Q3  | 1192.21| 1172.41               | 1.66%   | 213.69 | 208.20            | 2.57% |
| 2018 Q4  | 2717.48| 2654.31               | 2.32%   | 379.86 | 379.76            | 0.03% |
| 2019 Q1  | 230.11 | 217.99                | 5.27%   | 43.09  | 44.37             | 2.96% |
| 2019 Q2  | 586.29 | 480.98                | 17.96%  | 125.53 | 123.12            | 1.92% |
| 2019 Q3  | 1457.71| 1240.48               | 14.90%  | 223.53 | 222.15            | 0.62% |
| 2019 Q4  | 3102.27| 2806.21               | 9.54%   | 410.54 | 404.79            | 1.40% |
| 2020 Q1  | 248.99 | 230.29                | 7.51%   | 45.27  | 47.24             | 4.36% |
| 2020 Q2  | 585.18 | 507.73                | 13.23%  | 137.49 | 130.98            | 4.74% |

Tab.1 Prediction errors in agriculture and forestry
Date/Forecast | Agricultural | Forestry |
---|---|---|
2020 Q3 | 1490.10 | 242.55 |
2020 Q4 | 3337.91 | 437.29 |
2021 Q1 | 274.42 | 49.63 |
2021 Q2 | 621.54 | 146.09 |
2021 Q3 | 1592.36 | 257.87 |
2021 Q4 | 3563.11 | 464.67 |
2022 Q1 | 292.62 | 52.67 |
2022 Q2 | 585.18 | 154.96 |

**Tab.2** Forecast results

5. Conclusion

According to the forecasting model and results, it shows that the gross output value of agriculture and forestry in Guangxi in the 1st quarter of 2012 to 2nd quarter of 2020 tends to increase. After predicting the output value of Guangxi agriculture and forestry data, it is found that the output value of agriculture and forestry in the next eight quarters in general shows a growth trend. Based on the multiplicative Holt-Winters prediction model, this paper presents an empirical analysis of Guangxi agriculture and forestry output value data, which fully illustrates that the model is applicable to Guangxi agriculture and forestry output value prediction, and also verifies the practicality and validity of the model. It can provide an efficient and concise method for relevant prediction, and also provide a strong reference basis for the plan making of Guangxi agriculture department.

Acknowledgements

Teaching reform research project of Guangxi vocational education in 2017: The Application Research On Information Life Cycle Theory Of Digital orientation in Higher Vocational Colleges (No.: GXHZJG2017B07).
Research Project of GuangXi International Business Vocational College in 2019: Construction of a Information platform for Diagnosis and Improvement of Internal Quality Assurance System (No.: GSZX2019012).

References
[1] Xing-Yue T U , Jia-Ni X , Cheng-Kun G and et al 2014 Short-term forecast of agricultural products price based on time series and RBF J. Guangdong Agricultural Sciences
[2] Zheng X , Mengyang Q , Kang L and et al 2017 Development Forecast of Agricultural Circular Economy in Anhui Province in China Journal of Jixi University
[3] Elaoud A , Hassen H B , Salah N B and et al 2017 Modeling of soil penetration resistance using multiple linear regression (MLR) J. Arabian Journal of Geosciences p442
[4] Taki M , Ajabshirchi Y , Ranjbar S F and et al 2017 Evaluation of heat transfer mathematical models and multiple linear regression to predict the inside variables in semi-solar greenhouse J. Journal of Agricultural Machinery
[5] Jere S , Banda A , Kasense B and et al 2019 Forecasting Annual International Tourist Arrivals in Zambia Using Holt-Winters Exponential Smoothing J. Open Journal of Statistics pp258-267
[6] Jiang W , Wu X , Gong Y and et al 2019 Monthly electricity consumption forecasting by the fruit fly optimization algorithm enhanced Holt-Winters smoothing method
[7] Akolo I R 2019 Perbandingan exponential smoothing holt-winters dan arima pada peramalan produksi padi di provinsi gorontalo
[8] Zhu Y , Zhao Y , Zhang J and et al 2019 Spring onion seed demand forecasting using a hybrid Holt-Winters and support vector machine model plos one 14
[9] Zger Y E, Akpnar M, Musayev Z and et al 2019 Electrical Load Forecasting Using Genetic Algorithm Based Holt-Winters Exponential Smoothing Method J. Sakarya University Journal of Computer and Information Sciences pp108-123