Prediction of ALRTI caused by HRSV infection in Zhengzhou area in seasonal ARIMA model

Guang-chao Zhang
Zhengzhou Children's Hospital
Xiao-min Sun (min6881@126.com)
Zhengzhou Children's Hospital
Dan Chen
Zhengzhou Children's Hospital
Qing Guo
Zhengzhou Children's Hospital
Yanhong Ren
Zhengzhou Children's Hospital

Article

Keywords: HRSV infection, ALRTI incidence rate, model of ALRTI, epidemiologic study

Posted Date: March 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1443843/v1

Abstract

Objective The seasonal trend of ARIMA product was used to predict the incidence of HRSV infection in Zhengzhou area and to explore the application of the model in predicting the incidence rate of ALRTI caused by HRSV infection in ARIMA.

Methods Based on the time series construction model of ALRTI incidence rate from January 2018 to January 2020, the incidence rate of ALRTI caused by HRSV infection from January 2021 to January 2022 was verified by HRSV, and the results were validated.

Results Model of ALRTI (1,0,1) (0,1,1) 12 is the best model, with *BIC* = 8.319, Ljung-Box = 20.787, *P* = 0.160. The average relative error of predicted and actual values of ALRTI incidence rate from March 2021 to January 2022 is 339.33%, and the actual value is within 95% confidence interval of the predicted value.

Conclusion In predicting ALRTI caused by HRSV infection, ARIMA model can better predict the trend of incidence, and its effect needs to be further optimized.

Full Text

Human Respiratory Syncytial Virus (HRSV) is the most important viral pathogen that causes acute lower respiratory tract infections(ALRTI)in children under 5 years of age.It is also the leading viral factor that causes ALRTI hospitalization in infants under 1 year of age worldwide, and also the leading viral factor leading to hospitalization of ALRTI in infants under 1 year old, seriously endangering children's health.The incidence of ALRTI caused by HRSV infection is increasing year by year in Zhengzhou, Henan Province, China[1].In order to effectively prevent and control the high incidence of ALRTI caused by HRSV infection in Zhengzhou, this study constructs a Autoregressive integral moving average model (ARIMA) to predict the incidence trend of ALRTI caused by HRSV infection in Zhengzhou, and discusses the application of this model in predicting its incidence.

1 Data And Methods

1.1 Source

Cases of ALRTI caused by HRSV infection in Children's Hospital of Henan Province (Zhengzhou City) .

1.2 The research methods

The ARIMA model analysis method in SPSS statistical software was used to establish the time series of ALRTI incidence caused by HRSV infection from January 2018 to January 2020 in Zhengzhou.HRSV infection incidence ALRTI month time series for the seasonal time series, the product season model is adopted, namely the ARIMA(p,d,q)×(P,D,Q)s.Where d is a stabilized difference order number, in the process of p, q for autoregressive and moving average order number.P and Q for seasonal autoregressive and moving average order, D for seasonal difference order, s for the seasonal cycle.Through data smoothing processing, model recognition, parameter estimation and testing, the model was explored and established. The relative error between actual incidence rate and predicted incidence rate from January 2018 to January 2020 was compared(relative error=|actual value - predicted value| /actual value) as extrapolation validation to evaluate the prediction effect of the model.

1.2.1 Sequence tranquilization

According to the Zhengzhou region from January 2018 to January 2020 HRSV infection ALRTI incidence sequence diagram, chart and partial autocorrelation coefficient.If sequence is nonstationary sequence, on the original sequence for the non-seasonal difference or seasonal difference, and eliminate the influence of long-term trend of the sequence and periodic changes and make sequence stable[2].

1.2.2 Model identification

According to the difference sequence after the autocorrelation function and partial autocorrelation function diagram, as a model for initial recognition and formulation.

1.2.3 Model parameter estimation and test

Using the nonlinear least square estimation model parameters,on the basic condition parameters have statistical significance,the goodness of fit is used to compare models.Goodness of fit of the model using standard bayesian criteria,the model with minimum standardized BIC value and LJung-Box Q statistic P value＞0.05 was optimal.

1.2.4 The evaluation model of prediction effect

Comparison in March 2021 to January 2022 the actual incidence of incidence and prediction relative error, verify the model prediction effect.

2 The Results

2.1 Draw the sequence diagram

Draw the Zhengzhou region from January 2018 to January 2020 HRSV infection incidence of ALRTI time sequence diagram.The Fig. 1 shown in Zhengzhou region from January 2018 to January 2020 HRSV infection ALRTI has obvious seasonal disease (12 months for fashion cycles).Every popular cycle there is a popular boom,peaked in November to February of the next year.This sequence has the characteristics of both the seasonal cyclical fluctuations and has a rising trend year by year,so, we adopt the product of ARIMA model.

2.2 Sequence tranquilization

Zhengzhou region from January 2018 to January 2020 HRSV infection incidence of ALRTI time sequence diagram showed a trend of cyclical fluctuations,it can't meet the requirements of a stabilized.According to the characteristics of the seasonal fluctuation sequence diagram to natural logarithm conversion and the sequence of the primary season one order difference.After the difference of time series autocorrelation function and partial autocorrelation function has no obvious truncation and trailing phenomenon(Fig. 2,3),is also not a linear attenuation trend.Difference after time sequence diagram (Fig. 4) is close to a smooth, after the difference sequence suitable for time series model.

Figure2

2.3 The parameters of the model estimation and diagnosis

After determining model type, need to determine the P, d, q and P, d, q value,and to formulate stage for the model.According to the sequence of season change characteristics and smooth processing d = 0, D = 1.Based on the autocorrelation function and partial autocorrelation function diagram, P = 1, q = 1.Season model P, Q value is difficult to determine, in accordance with the relevant research [3], combined with the goodness of fit of the model and residual error and coefficient of correlation between estimates.Using Ljung-Box tests of residual white noise, eliminate the non-white noise model.After the test,model ARIMA(1,0,1)(0,1,1)12 had the minimum standardized BIC (1.457),smooth R2 = 0.216.Residual error sequence of autocorrelation coefficient and partial correlation coefficients within the 95% confidence interval (Fig. 5),Ljung—Box = 20.787, P = 0.160.Therefor,model ARIMA(1, 0, 1)(0, 1, 1)12 was selected as the optimal model.

2.4 Model to predict

According to ARIMA modeling method, the time series of ALRTI incidence caused by HRSV infection from January 2018 to January 2020 in Zhengzhou were modeled.Then, the monthly incidence of ALRTI caused by HRSV infection from March 2021 to January 2022 was used as validation data, and the sequence diagram of actual and predicted values was drawn, as shown in Fig. 6.According to the predicted value and actual value relative error to the assessment of the effect of forecast model (table 1).

3. Discussion

Respiratory tract infection is the most common disease in children, and HRSV is the most important viral cause of ALRTI in children under 5 years of age[4].Research has shown that 90% of children before the age of 2 are infected HRSV [5].At present, there is no vaccine or effective antiviral drug used in HRSV ALRTI prevention or treatment of infection.Therefore,through the incidence trend forecast reflects its importance.The time series model integrates the comprehensive effects of complex factors into time variables, which overcomes the problems of complicated factors affecting the onset of disease or problems that relevant data are not available[6],it has good applicability in the prediction of data with typical trend characteristics[7].

In this study, the monthly incidence data of ALRTI caused by HRSV infection from January 2018 to January 2020 in Zhengzhou were studied,through sequence stabilization, model recognition, parameter estimation and diagnosis, model prediction effect evaluation and other steps, the model ARIMA(1, 0, 1)(0, 1, 1)12 is established.This model fitted the incidence of ALRTI caused by HRSV infection in Zhengzhou, indicating that ARIMA model could predict the incidence of ALRTI in a short period of time with no significant fluctuation in the actual incidence trend, especially the seasonal model could fit the seasonal characteristics of ALRTI caused by HRSV infection. To judge the trend of the epidemic in advance and provide scientific basis for the formulation of prevention and treatment strategies [8].

This study forecasts from the interval estimation are consistent with the actual incidence interval estimation, all fall into the actual incidence forecast 95% CI.In terms of prediction accuracy, the minimum and maximum relative errors were 6.66%, 1316.96% and 339.33%, respectively, for the monthly incidence of ALRTI caused by HRSV infection from March 2021 to January 2022, but the 95%CI width of the predicted value was too large.In the ALRTI caused by HRSV infection, the actual prevention and control work is affected by social factors, climate factors and population immunity level, etc.[9]. The incidence of ALRTI caused by HRSV infection is complicated and changeable. Zhengzhou, as a transportation hub in central China with dense population and high mobility,it is necessary to further explore its onset of forecast model research work, make more practical guiding significance and value.

Declarations

Personal contribution:

The first author(Guangchao Zhang):Design, data analysis, literature search.

Corresponding author(Xiaomin Sun):Technical guidance.

Other authors: Data collection.

References

1. Zijin Shao, Mingxiu Liang.Action Initiative on prevention and Treatment of Human Respiratory syncytial Virus infection in Children in China.Chinese Medical Journal,2021,101(36): 2861–2866.
2. Tong Wang.Medical statistics and SPSS software.Beijing: 7/5000 Peking University Press,2008.
3. Zhenqiu Sun.Medical statistics.3/5000 3rd edition.Beijing: People's Medical Publishing House,2010:390–403.
4. Kyu HH, Pinho C,Wagner JA, et al. Global and national burden of diseases and injuries among children and adolescents between 1990 and 2013: findings from the global burden of disease 2013 study[J]. JAMA Pediatr, 2016, 170(3):267–287.DOI: 10.1001/jamapediatrics.2015.4276.
5. Wennergren G, Kristj a nsson S. Relationship between respiratory syncytial vincs bronchiolitis and future obstructive airway diseases[J]. Eur Respir J, 2001, 18(6):1044–1058. DOI:10.1183/09031936.01.00254101.

6. Meng Fandong Wu Di Sui Chengguang.Human Rabies Incidence in China:Trends and Predictions from a Time Series Analysis from 2004 Through 2015.Chinese Health Statistics,2016,33(3):389–391 + 395.
7. Chen Chun, Xiao Xin-cai.Application of seasonal ARIMA model on the prediction of incidence of hand foot and mouth disease in Guangzhou.Chinese Journal of Preventive Medicine, 2016, 17(2): 90–94.
8. National Clinical Research Center for Respiratory Diseases, Respiratory Group of Pediatrics Branch of Chinese Medical Association, Pediatric Respiratory Working Committee of Respiratory Branch of Chinese Medical Doctor Association, etc. Expert consensus on diagnosis, treatment and prevention of respiratory syncytial virus infection in children[J].Chinese Journal of Clinical Pediatrics,2020,35(4):241-250.DOI: 10.3760/cma.j.cn101070-20200225-00243.
9. Meng Xiangfei,Wang Xiaona.Risk factors for wheezing in children with respiratory syncytial virus infection[J].Chinese Journal of Nosocomiology,2020,30(13): 2053–2057.

Tables

Table 1
Comparison of predicted and actual monthly incidence of ALRTI due to HRSV infection from March 2021 to January 2022

month	actual value	estimate	actual value 95% CI	relative error(%)
3	1104.00	1312.02	1941.56-4565.61	208.02
4	684.00	1141.01	711.41-2993.42	457.01
5	423.00	664.96	1012.54-2342.46	241.96
6	182.00	577.61	1039.04-2194.26	395.61
7	193.00	304.03	1285.98-1894.03	111.03
8	143.00	501.48	1075.72-2078.69	358.48
9	306.00	291.35	1279.44-1862.13	14.65
10	1104.00	663.75	903.75-2231.25	440.25
11	2810.00	1493.04	72.75-3058.84	1316.96
12	3062.00	3244.02	1679.12-4808.93	182.02
1	4009.00	4015.66	2451.22-5580.11	6.66
average relative error			339.33	

Figures

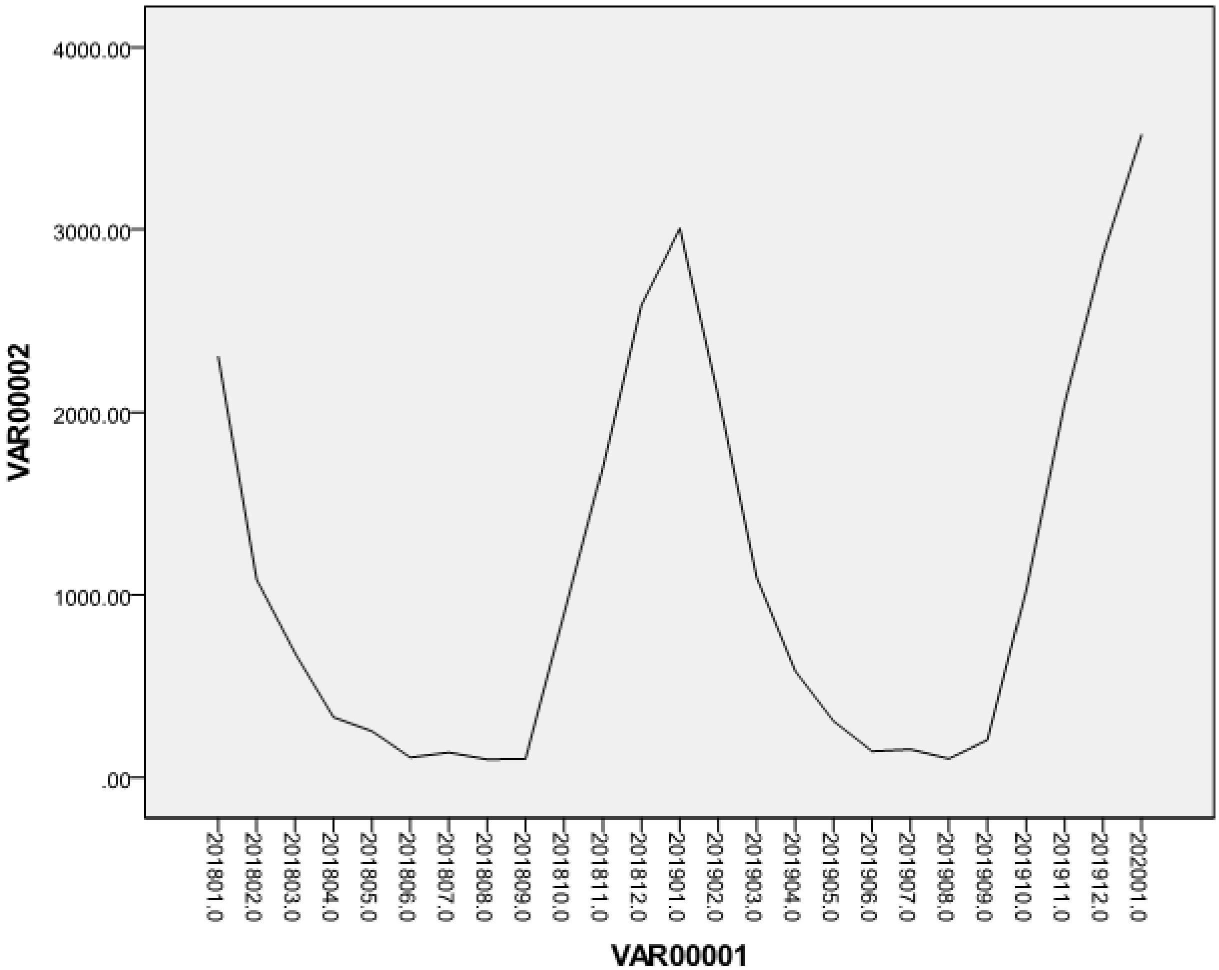

Figure 1

Legend not included with this version.

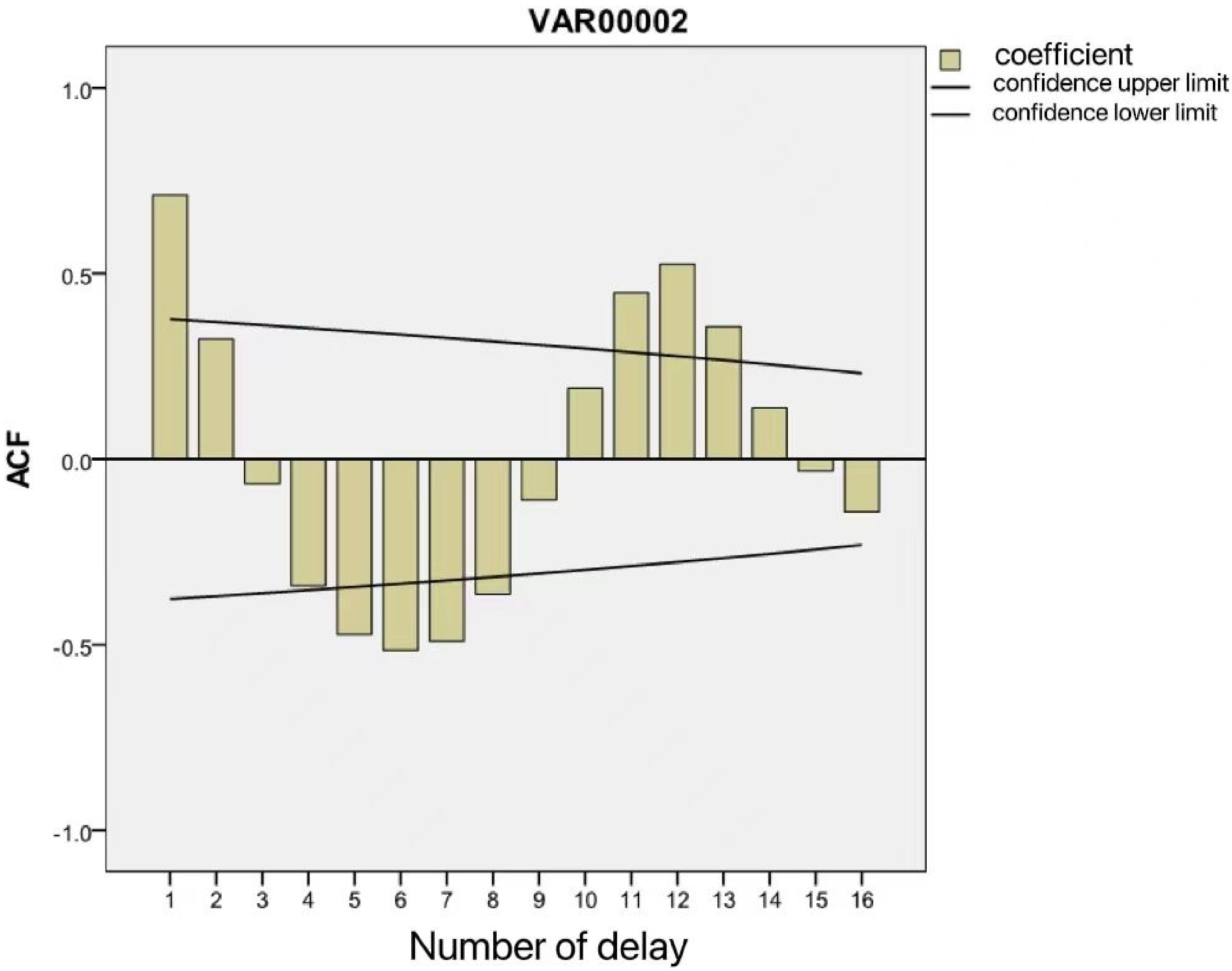

Figure 2

Legend not included with this version.

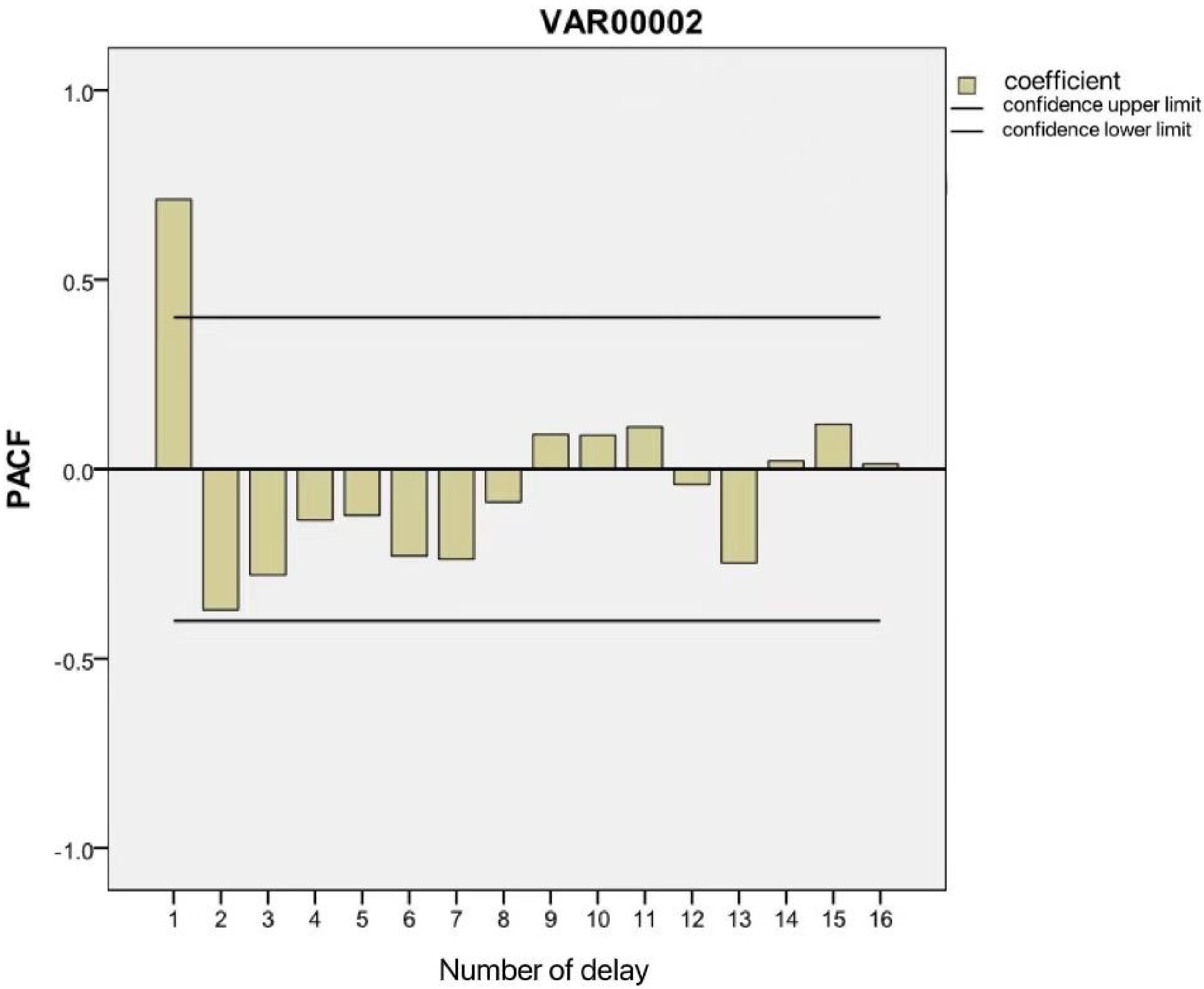

Figure 3

Legend not included with this version.

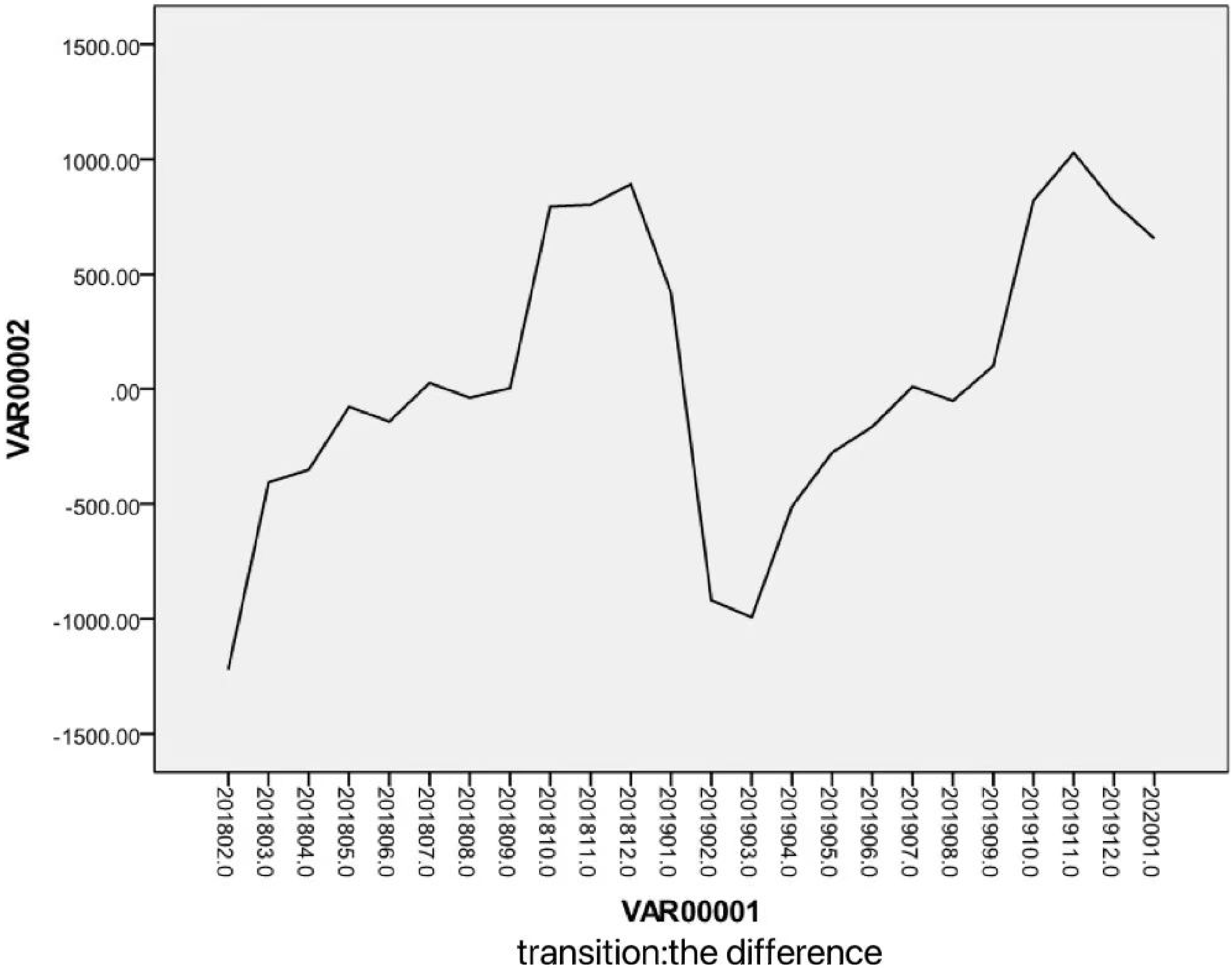

Figure 4

Legend not included with this version.

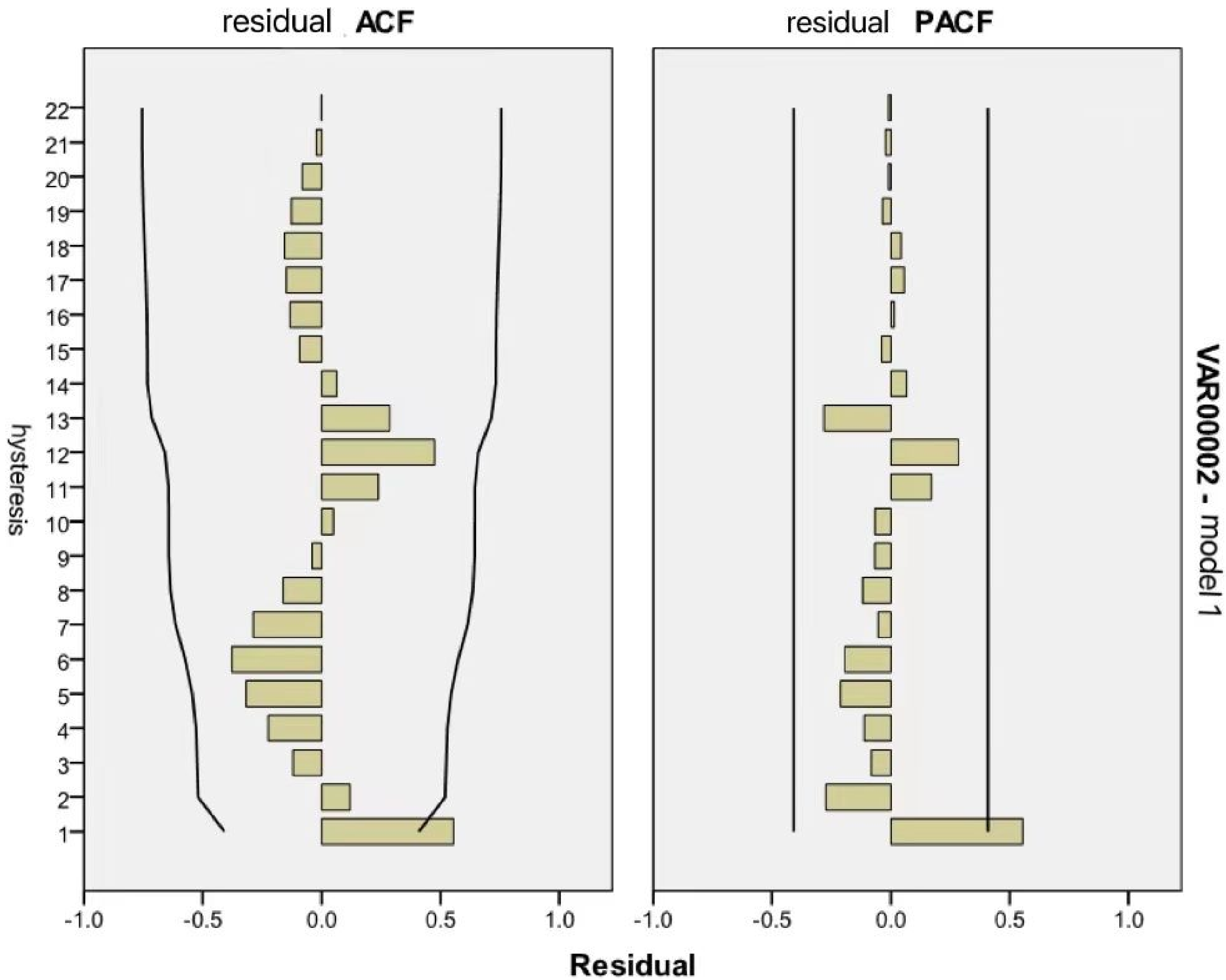

Figure 5

Legend not included with this version.

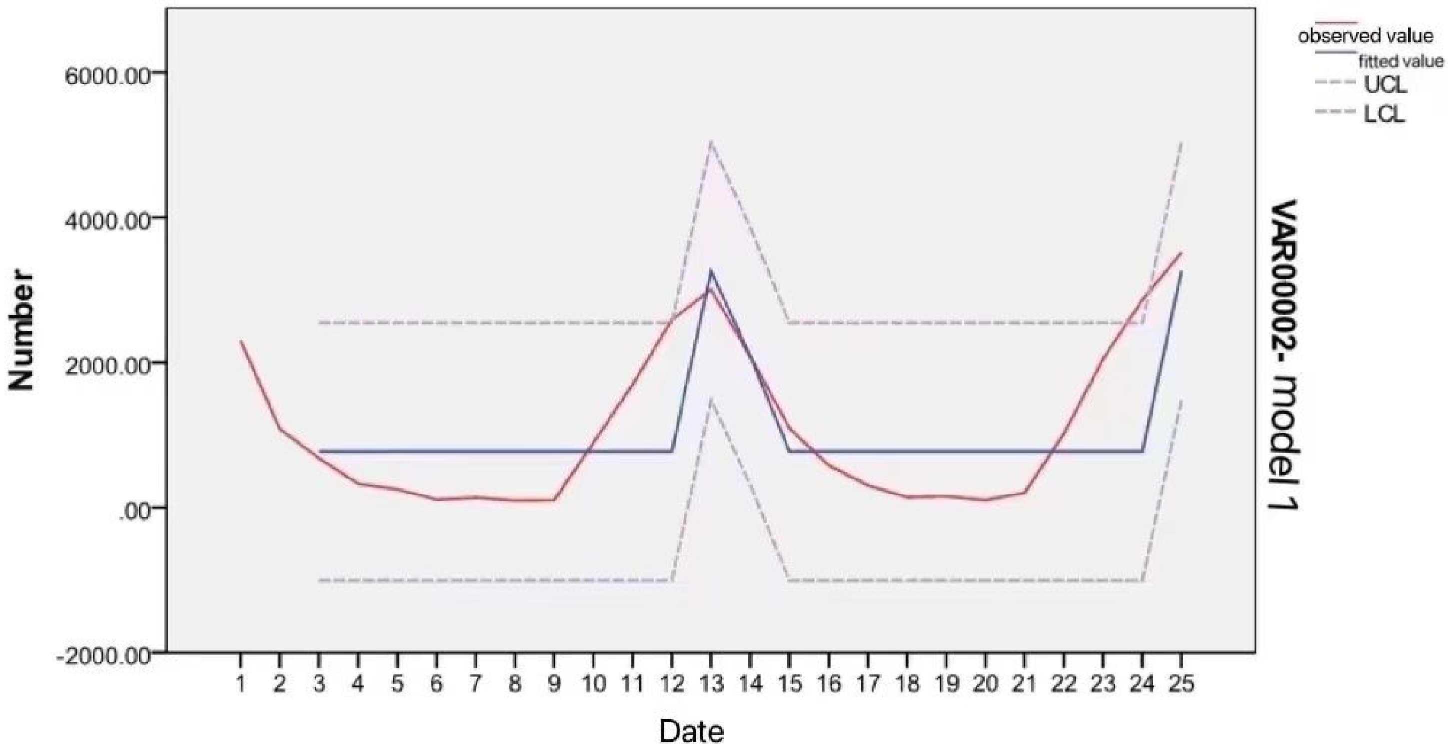

Figure 6

Legend not included with this version.