Comparison Of ARIMA Model and Residual Autoregressive Model in Analyzing and Forecasting the Healthcare Consumer Price Index

Shuqiang Xu[1,a,*]

[1]School of Sociology and Political Science, Shanghai University, Shanghai, China

[a]806135661@qq.com

[*]corresponding author

Keywords: Medical care; Consumer Price Indices; ARIMA Model; Residual Autoregressive Model

Abstract: To study the change of residents' health care demand in China, through our research in January 2005 - December 2015 China health care consumer price index, using the SAS software first season index by calculation on the analysis of short-term season; Then in the use of X - 11 process, unit root test, such as white noise testing method for sequence preprocessing, on the basis of ARIMA model is set up respectively and residual autoregressive model to simulate the long-term trend of the sequence proposed merger forecast in January 2016 - December sequence; Finally, the fitting situation and prediction sequence are analyzed, and the advantages and disadvantages of the two models are compared.

1. Literature Review

With the continuous rise in the living standards and consumption levels of Chinese residents, the current academic research on medical care consumption is increasing. Some scholars have studied the relationship between the age structure of the Chinese population and residents ’health care consumption, and found that an increase in the old-age dependency coefficient will lead to an increase in the residents’ per capita health care consumption expenditure, while the child dependency coefficient has a small impact on residents ’health care consumption [6] . Some scholars have paid attention to the impact of the aging of the rural population on the health care consumption of residents and found that the aging of the rural population has a positive effect on the improvement of rural health care consumption [9]. Some scholars based on the analysis of Guangdong, Jiangsu, Zhejiang and Zhejiang provinces have shown that per capita income, government investment in health services, and aging are the main influencing factors of urban residents 'health care consumption. In recent years, the marginal consumption tendency of urban residents' health care, elasticity of demand income and medical treatment. The proportion of health consumption in per capita consumption expenditure is relatively low, indicating that there is still much room for development of health care consumption for middle- and high-income people [5]. Some scholars have researched and compared the inter-provincial differences of urban residents' health care consumption [4]. Some scholars have conducted non-parametric estimations of the Engel curve of medical and health care consumption in China's urban and rural areas, the marginal impact of total expenditures on the share of health care expenditures, and the elasticity of health care consumption expenditures, and analyzed and compared their urban and rural differences [10].

The consumer price index (CPI) is a general reflection of the household consumer goods and services purchased by macroeconomic indicators of changes in the price level. It is in particular within a period of relative price level amount of the number of a representative set of consumer goods and services change over time, and is used to reflect changes in household purchasing price level of consumer goods and services. The consumer price is one of the important contents of macroeconomic monitoring, and the relationship and choice between the indexes are of great significance [3]. Some scholars have performed functional data analysis on China's consumer price index and found that there is seasonality in China's consumer price index [7]. Some scholars have studied the cyclical fluctuations of the Chinese consumer price index and considered its internal and external driving factors [1]. Some scholars have put forward suggestions for poverty alleviation policies by studying the compilation of consumer price indices for urban and rural residents [2].

However, relatively few studies mainly focus on the health care consumer price index. This study compares two models that analyze the consumer consumption index, and provides support for future scholars to study related research on the healthcare consumer price index.

2. ARIMA Model

We obtained the data of China's healthcare consumer price index (HCCPI) from January 2005 to December 2015 from China Statistical Yearbook.

2.1. Model Ordering

Through stationary test, we can get the autocorrelation diagram of the first-order difference term sequence of HPCPI, have obtained that the sequence is censored to the first order. Further, we examine the partial correlation diagram of the first order difference term of the HCCPI as shown in Figure 1.

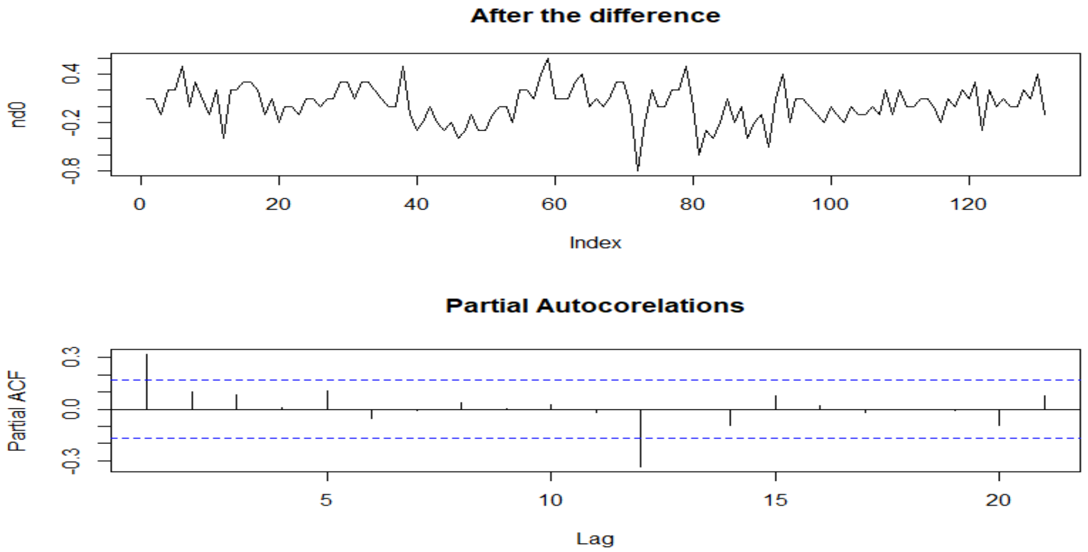

Figure 1. Partial correlation diagram.

As shown in Figure 1, the partial autocorrelation diagram of the first order of the HCCPI difference term sequence is more than twice the standard deviation range at the delay of 12 periods, and overall shows non-censoring (tailing). Therefore, it is considered to use the MA (1) model to fit the HPCPI first order difference term sequence, that is, to use the ARIMA (0,1,1) model to fit the HPCPI sequence.

2.2. Model Fitting

The ARIMA (0,1,1) model (no intercept term) was fitted to the HCCPI sequence from January 2005 to December 2015 using SAS software. As shown in Table 1, the parameter estimation value is -0.25786, and its p value is 0.0029, which is significant at the 5% confidence level, so the parameter estimation results are significant.

Table 1. Parameter estimation results.

Parameter	Estimate	Standard Error	t-value	Pr > |t|	lag
MA1,1	-0.25786	0.08507	-3.03	0.0029	1

We further use SAS software to test the residuals of the model through a white noise test to determine the overall interpretation of the model by the model. The test results are shown in Table 2.

Table 2. White noise test for residual term.

To lag	Chi-Square	DF	Pr>ChiSq	Autocorrelations					
6	9.20	5	0.1013	0.030	0.151	0.141	0.042	0.137	0.058
12	21.68	11	0.0270	0.002	0.067	0.035	0.001	0.053	-0.277
18	29.19	17	0.328	-0.046	-0.180	-0.000	-0.003	-0.125	-0.002
24	32.91	23	0.0827	-0.030	-0.112	0.060	-0.072	-0.013	-0.032

As shown in Table 2 , the sequence of residual terms with 6 and 24 delays is not significant at the 5% significance level, but the 12 and 18 delays are significant at the 5% significance level. Therefore, the sequence cannot be rejected as a white noise sequence, that is, the ARIMA (0,1,1) model does not sufficiently explain the HCCPI sequence, and it is necessary to consider improving the model.

2.3. Model Optimization

Considering the particularity of the partial correlation diagram of the HCCPI sequence, we try to introduce an AR model with a lag of n periods. First, we establish a ARIMA (1,1,1) model to fit the HCCPI sequence through a lag of 1. The parameters θ_1 and ϕ_1 and estimation results of the ARIMA (1,1,1) model are 0.53636 and 0.76750, which are significant at the 5% confidence level, so the parameter estimation results are significant; the white noise test of the residual term in this model In the results, the sequences of the residual terms delayed by periods 6, 12, 18, and 24 are not significant at the 5% confidence level, so the hypothesis that the sequence is a non-white noise sequence can be rejected, that is, ARIMA (1,1, 1) The model fully interprets the HCCPI sequence information; the model has an AIC value of -26.7934 and an SBC value of -21.043, which are sufficiently small, so it is considered that ARIMA (1,1,1) fits the HCCPI sequence well. Finally, the HCCPI sequence ARIMA (1,1,1) model is shown as follows:

$$(1-0.76750*B)(1-B)*x_t=(1-0.53636*B)*\varepsilon_t \qquad (1)$$

2.4. Result from Fitting and Forecasting

We finally use the ARIMA (1,1,1) model obtained by SAS software to fit the HCCPI sequence and make 12-period (12-month) predictions. The results are shown in Figure 2. Among them, the black star point is the actual value, the red curve is the fitted prediction value, the green curve is the 95% confidence interval upper bound fitting prediction value, and the blue curve is the 95% confidence interval lower bound fitting prediction value.

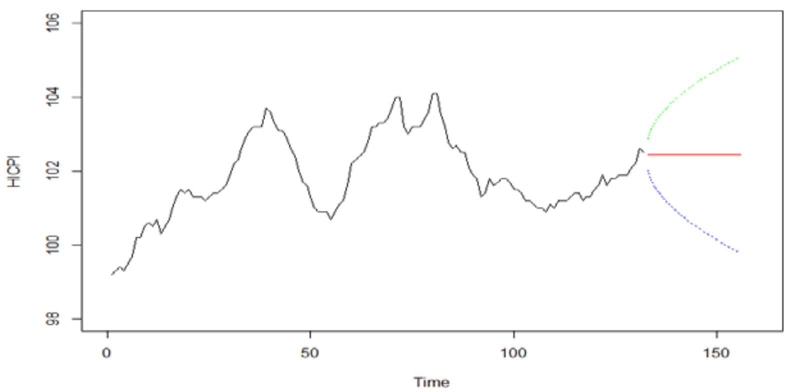

Figure 2. ARIMA (1,1,1) model fitted value and predicted value.

3. Residual Autoregressive Model

3.1. DW Test

We used SAS software to test the HCCPI sequence by Durbin- Waston test. The DW statistic is equal to 0.0474, and its p-value is significant at a 5% confidence level, indicating that the residual term of the HCCPI sequence is significantly positively correlated. Therefore, the residual sequence can be used to fit the autocorrelation model, that is, the residual autoregressive model is fitted to the HCCPI sequence.

3.2. Fitting Model

Residual autoregressive model fitting was performed on HCCPI sequences from January 2005 to December 2015 using SAS software, and the test results, screening process and parameter estimation results were shown in Figure 3 and Table 3, 4, 5.

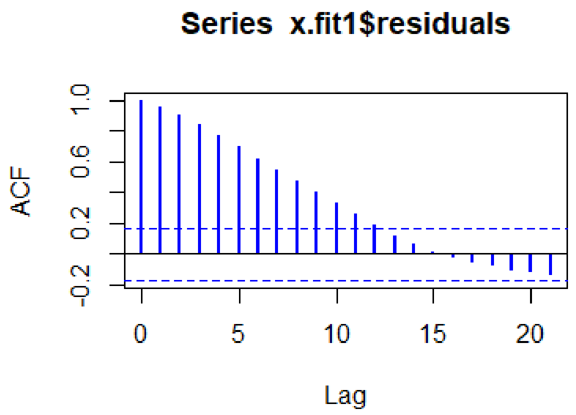

Figure 3. Residual sequence autocorrelation diagram.

Table 3. Report of insignificant items eliminated by stepwise regression.

Lag	Estimate	t-value	Pr > |t|
8	0.023305	0.17	0.8635
7	-0.043761	-0.39	0.6991
5	-0.063290	-0.47	0.6369
6	0.037280	0.46	0.6477
9	-0.036288	-0.37	0.7129
3	-0.063662	-0.48	0.6298
2	0.050826	0.46	0.6468
10	0.041482	1.29	0.2002

Table 4. Regression on remaining significant items.

Lag	Coefficient	Standard Error	t-value
1	-1.072010	0.045028	-23.81
4	0.130558	0.045028	2.90

Table 5. Fitting results of the residual autoregressive model.

Variable	Estimate	Standard Error	t-value
Intercept	100.7280	0.8049	125.15
t	0.0149	0.0100	1.49
AR1	-1.1240	0.0335	-28.45
AR4	0.1620	0.0395	4.10
SBC	-11.3594		
AIC	-22.8906		
R^2	0.9611		

As shown in Figure 3, the residual sequence autocorrelation plot shows that the residual sequence has a significant first-order positive correlation. As shown in Table 3 and 4, the reports of insignificant items that are eliminated gradually and the remaining significant items that are gradually returned are shown in the report. Except for the residual sequence values of the delayed periods 1 and 4, the sequence values of the remaining periods are all delayed. There is no significant autocorrelation; as shown in Table 5, the parameter estimation result of the residual autoregressive model is that the intercept term Intercept is 100.7260, the time t is 0.0149, the delay term residual term is -1.1240, and the delay term residual is -1.1240. The difference term is 0.1620, where the intercept term, the delayed 1-term residual term, and the delayed 4-term residual term are all significant at the 5% confidence level and the time is significant at the 15% confidence level; the SBC of this model is -11.359431, and the AIC is - 22.890639, both sufficiently small, Total R-Square is 0.9611, indicating a good model fit. Finally, the HCCPI sequence residual autoregressive model is obtained as follows:

$$\begin{cases} x_t = 100.7260 + 0.0149 * t + u_t \\ u_t = 1.1240 * u_{t-1} - 0.1620 * u_{t-4} + \varepsilon_t, \varepsilon_t \sim N(0,0.04656) \end{cases} \quad (2)$$

3.3. Result from Fitting and Forecasting

We finally used the SAS software to fit the HCCPI sequence through the obtained residual autoregressive model, and made 12-period (12-month) predictions backward. The results are shown in Figure 4. Among them, the black star point is the actual value, the red curve is the overall fit prediction value, and the green curve is the trend fit prediction value.

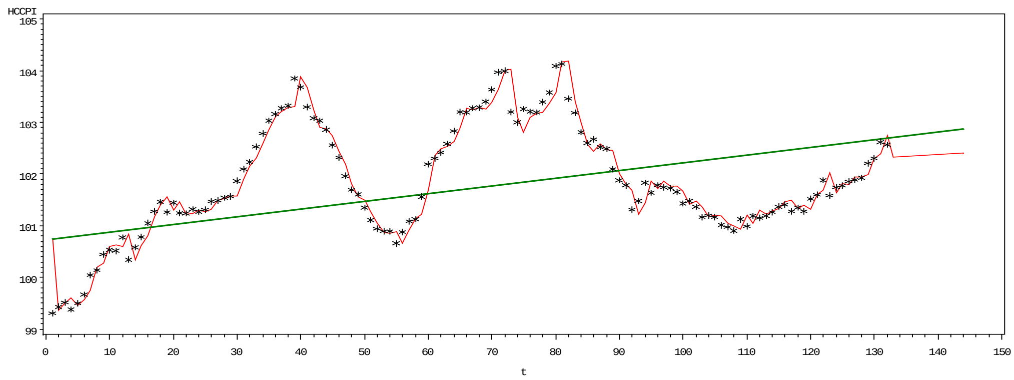

Figure 4. Fitting results of the residual autoregressive model.

After establishing the ARIMA (1,1,1) model and the residual autoregressive model respectively to fit the HCCPI sequence from January 2005 to December 2015 and to predict the January to December 2016 HCCPI sequence, we use SAS software to combine the two The model's fitted prediction results and actual values are put together for comparison and analysis, as shown in Figure 5.

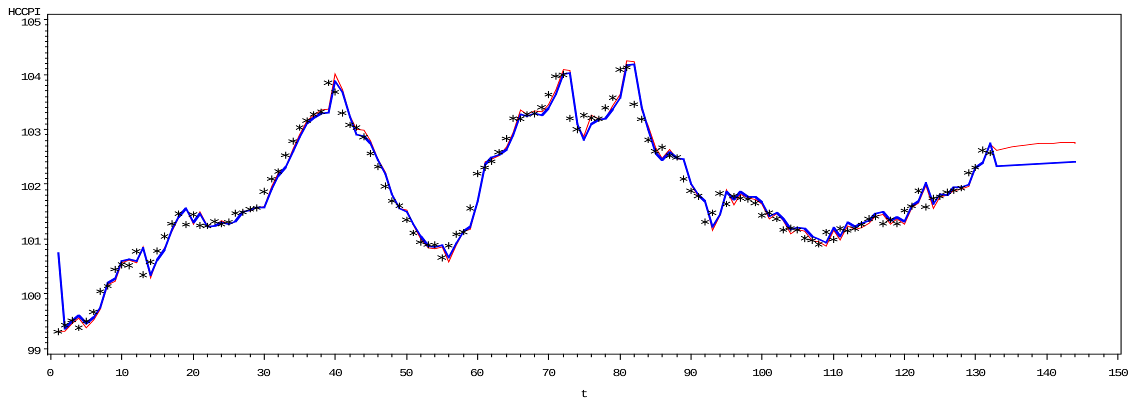

Figure 5. ARIMA (1,1,1) model, residual autoregressive model fitted predicted and actual values.

First compare the overall fit of the two models. In Figure 5, star point in black is the actual value, the red curve is the ARIMA (1,1,1) model fitted predicted value, and the blue curve is the residual autoregressive model fitted predicted value. The overall fitted values of the ARIMA (1,1,1) model and the residual autoregressive model for January 2005 to December 2015 are not significantly different, and the predicted values for January to December 2016 are significantly different. Although the predicted values of both remained basically stable after 2 months of continuous decline in 2016, it is clear that the residual autoregressive model has greater fluctuations. Then compare the AIC and BSC criteria of the ARIMA (1,1,1) mode. It can be seen that the AIC value and the BSC value of the residual autoregressive model are greater than ARIMA (1,1, 1) The model, so the residual autoregressive model has lower fitting accuracy than the ARIMA (1,1,1) model. The main reason is that it has a lower accuracy in extracting deterministic information.

Table 6. ARIMA (1,1,1) model and residual autoregression model AIC, SBC.

Modeling	AIC	SBC
ARIMA(1,1,1) Model $(1-0.76750B)(1-B)x_t = (1-0.53636B)\varepsilon_t$	-26.7934	-21.043
Residual autoregressive model $\begin{cases} x_t = 100.7260 + 0.0149t + u_t \\ u_t = 1.1240u_{t-1} - 0.1620u_{t-4} + \varepsilon_t, \varepsilon_t \sim N(0{,}0.04656) \end{cases}$	-11.359431	-22.890639

Secondly, the interpretation of the data by the two model equations is compared. As shown in Table 6 , the fitting equation of the ARIMA (1,1,1) model can be simplified as $x_t = 1.76750\nabla x_t - 0.76750\nabla^2 x_t + \varepsilon_t - 0.53636\nabla\varepsilon_t$, which indicates that the HCCPI has a positive correlation with its current change, and a negative correlation with its current change. The adjustment coefficients are 1.76750. And -0.76750, which is positively correlated with random fluctuations and negatively correlated with the current fluctuations of random fluctuations, with adjustment coefficients of 1 and -0.53636, respectively. From this, we can hardly see the long-term trend of the changes in the HCPPI sequence. demand trained to experience the impact of this rough conclusion.

The fitting equation of the residual autoregression model indicates that the base value of HCCPI is 100.7260, indicating that people's demand for medical care is generally on the rise; HCCPI is positively correlated with time, with a positive correlation coefficient of 0.0149, which means that each time is delayed by one period (1 month), HCCPI increased by 0.0149, indicating that people's demand for medical care is increasingly urgent; HCCPI is positively correlated with random fluctuations, while random fluctuations are positively correlated with previous period fluctuations, and negatively correlated with upward fourth period fluctuations. The adjustment coefficients are 1.1240 and -0.1620, indicating that people's demand for medical care will be adjusted according to their short-term experience and longer-term experience. Generally speaking, the increase in medical care demand in the short-term will promote people's further consumption, but higher long-term health care Demand will curb people's enthusiasm for their consumption.

By comparing the conclusions of the two models, it is concluded that although the ARIMA (1,1,1) model has high fitting accuracy, it is not easy to explain the conduction mechanism behind the model. On the contrary, the residual autoregressive model clearly explains the HCCPI Law of change.

4. Conclusion and Recommendation

According to the above analysis, the following conclusions can be drawn: First, from the perspective of model comparison: the ARIMA model and the residual autoregressive model are common models for fitting non-stationary white noise time series. Among them, the fitting accuracy of ARIMA model is higher than that of residual autoregressive model. The fitting equation of residual autoregressive model is more intuitive and easier to explain than the fitting equation of

ARIMA model. Second, from the perspective of analysis results: First, from seasonal analysis It can be seen that people's demand for medical care encountered a trough in March, ushered in the highest peak of the year in July, and a small peak in October, which can be explained by the peak summer heat demand for people's health Autumn is a period of high incidence of infectious diseases. In addition, since the average seasonal index in January-December is concentrated around 100, the seasonal effect can be considered to have no significant effect on people's long-term medical care needs. Secondly, because the base value of HCCPI is about 100.7 greater than 100, with the continuous improvement of living conditions, people's demand for medical care is generally on the rise; HCCPI is positively correlated with time, and each time the time is delayed by 1 month, HCCPI increases by about 1.5%. , Indicating a growing enthusiasm for health care. Finally, people 's demand for healthcare is significantly affected by past experience. Generally speaking, higher-than-expected healthcare demand in the past month will prompt people to increase their own consumption of healthcare, with each 1% increase in such high demand The current demand increased by about 1.1%; on the contrary, going back to the fourth month, higher-than-expected medical demand will cause people to reduce their consumption. For every 1% increase in such high consumption, the current consumption decreases by about 0.2%.

Based on the above conclusions, we also propose two suggestions: (1) Model selection: For time series that have a deterministic trend and need to explain the underlying conductive mechanism, the residual autoregressive model is recommended; for those that do not have a deterministic law , Or time series for accurate prediction, the ARIMA model is recommended. (2) Health care: The growth of HCCPI reflects the increase in residents' demand for health care. This requires the government to improve the health care related system, carry out more scientific planning, and play a role in controlling medical expenses and medical insurance funds in accordance with the actual conditions of the country. The leading role is to formulate corresponding policies to regulate prices in the healthcare market.

"Expensive in medical treatment" is a social hot issue for many years in our country, and the current medical care price index does not fully reflect this kind of phenomenon, this makes when calculating the medical care price index to give all medical expenses, patients discharged from hospital outpatient times per capita medical indicators such as the specific weight has more practical significance, it is also a direction of further research in the future.

References

[1] Chen Chengzhong, Lin Zhenshan. (2009). A study on the periodicity and driving factors of CPI volatility in China. Inquiry into Economic Issues (8), 77-84(in Chinese).

[2] Fan Neng, Wang Yong, Du Qing. (2017). Compilation of stratified consumer price index and enlightenment of poverty alleviation policies. China Economic Studies (2), 125-135(in Chinese).

[3] Gui Wenlin, Han Zhaozhou. (2012). Relationship and choice among China's consumer price index. Statistical Research, 29(9), 6-13(in Chinese).

[4] Han Xuemei, Wang Zengfu. (2015). Analysis of inter-provincial differences and influencing factors of urban residents' health care consumption. Gansu Social Sciences (2), 183-186(in Chinese).

[5] Nie Ling, Li Bin. (2010). Research on medical care consumption of urban residents -- based on the analysis of the three provinces of guangdong, jiangsu and zhejiang. Journal of Guangdong University of Finance and Economics, 25(6), 64-70(in Chinese).

[6] Wang Xueyi, Zhang Chong. (2013). Age structure of Chinese population and residents' health care consumption. Statistical Research, 30(3), 59-63(in Chinese).

[7] Yan Mingyi, Du Peng. (2010). Functional data analysis of seasonal changes in China's consumer price index. Statistics & Information Forum, 25(8), 100-106(in Chinese).

[8] Ye Xiangming, Xu Yun, Xu Shengxin, Hu Xijia. (2012). Practical significance of expanding

medical care consumption. Soft Science of Health (12), 1012-1014(in Chinese).

[9] Zhang Chong, Wang Xueyi, Sun Weihong. (2015). Effects of aging of rural population on residents' health care consumption -- based on China's provincial panel data from 2002 to 2012. Collected Essays on Finance and Economics(1), 32-38(in Chinese).

[10] Zhou Xianbo, Tian Fengping. (2009). Differential analysis of health care consumption of urban and rural residents in China -- nonparametric estimation based on the engel curve model of panel data. Statistical Research, 26(3), 51-58(in Chinese).