Pacific International Journal, Vol. 6(1), 77-82; 2023
ISSN (Print) 2663-8991, ISSN (Online) 2616-4825
DOI: 10.55014/pij.v6i1.311
https://rclss.com/index.php/pij

The Forecast Analysis of Chinese Resident Consumption Based on the Residual Autoregressive Model

Shuchao Wang[1,2]

[1] Center for International Education, Philippine Christian University, Manila, 1004, Philippines
[2]School of Mathematics and Big Data, Chaohu University, Hefei 238000, China
Email: wangshuchao.sd@163.com

Abstract: This paper aims to improve the accuracy of resident consumption forecast by comparing different methods and selecting the residual autoregressive model. The model combines deterministic analysis method and autoregressive modeling to extract deterministic and random information, respectively, and improve the prediction accuracy. The paper introduces the modeling process of the residual autoregressive model and conducts empirical analysis on the consumption level data of Chinese residents to evaluate the fitting accuracy of the model. Finally, the model is used to forecast the consumption level of Chinese residents in 2021-2023, and the results indicate an average growth rate of about 7% over the next three years, assuming no major changes in the external environment. This paper contributes to the understanding of the residual autoregressive model and its application in the field of resident consumption forecasting, providing valuable insights for policymakers and researchers in this area.

Keywords: residual autoregressive model, deterministic analysis, random analysis, resident consumption, forecast

I. Introduction

There are many papers studied resident consumption, most of which are from the perspective of influencing factors of resident consumption. Feng Ying studied it from the perspective of the effect of consumer confidence index on consumer consumption, believing that consumer confidence index contains effective information that can explain consumption and is effective in forecasting consumer expenditure [1]. Wang Jinan established a fixed effect model to conduct an empirical study on the factors influencing Chinese resident consumption demand and found that disposable income is the most important factor influencing Chinese resident consumption [2]. Deng Yuxuan empirically analyzed the impact of fintech on the spatial structure of resident consumption by using urban panel data and spatial econometric model, and found that fintech was conducive to the growth of regional resident consumption [3]. Wang Lei studies the effect of gender structure difference on consumption scale and consumption upgrading of Chinese residents based on panel data of 31 provinces and regions of China, and found that the imbalance of gender structure in China has a positive effect on the expansion of resident consumption scale in general, but for the upgrading of resident consumption, there is a large regional heterogeneity under different age structures [4]. However, there are many factors that affect resident consumption, it is difficult to obtain all the data, and there are often multicollinearity and other problems among various factors. Therefore, it is time-consuming and laborious to analyze and forecast resident consumption from the perspective of influencing factors, and the prediction accuracy is not very high. Some papers change non-stationary sequence into stationary sequence by difference operation and then use an autoregressive integrated moving average model (ARIMA model) for prediction. Li Jie analyzed and predicted the consumption level of residents in Anhui Province by means of differential smooth processing and the establishment of ARIMA model [5]. Xiao Liang analyzed and predicted the CPI time series by establishing a seasonal ARIMA model, and explored the change rule of economic variable CPI over time [6]. Although differential processing can stabilize non-stationary data, part of the information is often lost, and the model is difficult to explain.

In the analysis of economic and financial problems, time series analysis is widely used. Time series data can be classified into stationary time series and non-stationary time series. For stationary non-white noise time series, different autoregressive moving average (ARMA) models can be selected for fitting according to the correlation between different moments of the series. For non-stationary time series, there are usually deterministic analysis methods and random analysis methods to analyze. The deterministic analysis method ignores the random fluctuation of the series and fits the deterministic trend. Although the method can fit the basic style, it is not the whole picture of time series changes, so the prediction results are often not accurate. Random analysis method firstly smooths the non-stationary sequence through difference operation, and then fits the ARMA model to the stationary sequence after difference. This method generally has high prediction accuracy, but the difference operation process will lose part of the data information [7], and it is difficult to intuitively explain the model.

The residual autoregressive model integrates deterministic and stochastic time series analysis methods. Firstly, the

[Received 25 Feb 2023; Accepted 07 Mar 2023; Published (online) 31 Mar 2023]

main deterministic information in the sequence is extracted by the deterministic analysis method, and then the random information is extracted by fitting the residual sequence with the autoregressive model. Deterministic analysis method can be used to fit the basic trend of data, and the established model is intuitive and easy to explain. The stochastic analysis method can improve the fitting accuracy of the model by extracting the relevant information contained in the residual error. This paper establishes a residual autoregressive model to predict the resident consumption.

II. Theoretical approach: residual autoregressive model

The general forms of residual autoregression model are as follows:

$$y_t = \beta_0 + \beta_1 \cdot t + \cdots + \beta_k \cdot t^k + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \cdots + \varphi_p \varepsilon_{t-p} + v_t \tag{2}$$

Formula (1) is the fitting of the deterministic trend in the sequence. The variable y_t represents the consumption level of Chinese residents from 1991 to 2020, $\beta_i(i = 1, \cdots, k)$ is the regression coefficient of the model, ε_t is the residual of Formula (1), and t represents the time. To simplify the problem, we set $t = 1$ in 1991, $t = 2$ in 1992, and so on, $t = 30$ in 2020. By extracting the deterministic information of the sequence, the basic trend of the sequence can be fitted.

Formula (2) is the extraction of random information, it can further extract the remaining information that has not been sufficiently extracted and improve the fitting accuracy. ε_t is the residual of Formula (1), which contains the information not extracted by the regression model. Formula (2) further extracts the relevant information in ε_t by establishing an autoregressive model for it. $\varphi_j(j = 1, \cdots, p)$ is the regression coefficient of the autoregressive model, v_t is the residual of the autoregressive model, it is a white noise sequence.

The symbols and abbreviations used in this paper are summarized in Table 1.

Table 1 Description of symbols and abbreviations

abbreviation	
ARMA	autoregressive moving average model
ARIMA	autoregressive integrated moving average model
symbol	
$y_t(t = 1, \cdots, 30)$	the consumption level of Chinese residents from 1991 to 2020
t	the time, we set $t = 1$ in 1991, $t = 2$ in 1992, and so on, $t = 30$ in 2020
$\beta_i(i = 1, \cdots, k)$	the regression coefficient of the regression model (i.e. Formula (1))
ε_t	the residual of the regression model (i.e. Formula (1))
$\varphi_j(j = 1, \cdots, p)$	the regression coefficient of the autoregressive model (i.e. Formula (2))
v_t	the residual of the autoregressive model (i.e. Formula (2))

Source: Author

The residual autoregressive model requires that the deterministic trend in the time series can be extracted through Formula (1) until the residual sequence ε_t is stable, and then it tests the autocorrelation of the residual sequence ε_t. If the residual sequence ε_t is judged to be a white noise process, there is no need to continue the extraction of residual ε_t, indicating that Formula (1) is sufficient to extract information. If the residual sequence ε_t is significantly autocorrelated, indicating that Formula (1) is not sufficient to extract information, then an autoregression model should be established for the residual sequence ε_t until the information is fully extracted.

III. Empirical analysis of resident consumption level

To fit and predict the consumption level of Chinese residents, a residual autoregressive model is established in this paper. First, the regression model of variable y_t with respect to time t is established. Then the residual in regression model is evaluated to see if it is a white noise sequence. If the result is not, an autoregressive model of residual is established to fit the residual. Finally, a complete residual autoregressive model is estimated by combining these two models. The data are from the website of the National Bureau of Statistics of China from 1991 to 2020.

3.1 Fitting of the deterministic trend

To explore the variation trend of sequence y_t with time t, a time sequence diagram is drawn, as shown in Figure 1. From Figure 1, variable y_t is a quadratic function of time t, and a quadratic polynomial model of sequence y_t about time t can be established, as shown in Formula (3):

$$y_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \varepsilon_t \tag{3}$$

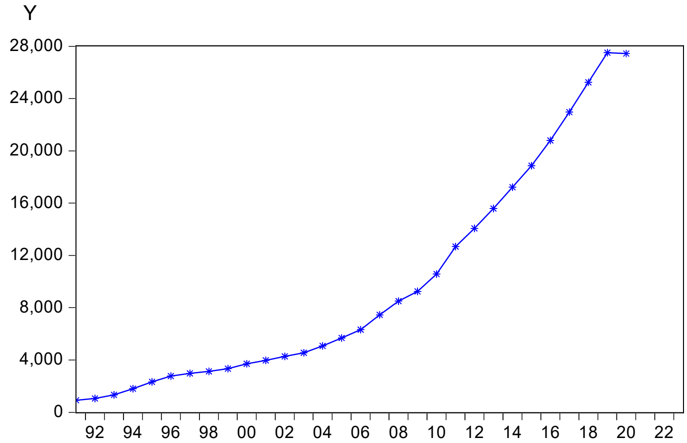

Figure 1 The trend of resident consumption in China from 1991 to 2020. Source: Author

In this paper, EViews 8.0 is used to estimate Formula (3), and the results are shown in Formula (4):

$$y_t = 2620.91 - 388.24t + 41.59t^2 + \varepsilon_t$$
$$(6.01) \quad (-5.98) \quad (20.49) \tag{4}$$
$$R^2 = 0.993, \quad F = 1860.09, \quad D.W. = 0.49$$

In Formula (4), the value in each parenthesis is the T-statistic value of each parameter (same below). According to the results, the goodness of fit is as high as 0.993, indicating that the quadratic polynomial curve model can explain most of the information about the change of resident consumption. $F = 1860.09$, $p = 0.000$ indicating that the model is significant at the significance level of 5%. The $p = 0.000$ in T-test for each parameter indicating that each parameter has passed the significance test at the significance level of 5%. $D.W. = 0.49$, while at the 5% significance level $d_L = 1.28$ and $d_U = 1.57$, so $D.W. < d_L$, indicating that the residual sequence of the model has high positive autocorrelation.

The fitting effect diagram of deterministic trend drawn by EViews8.0 is shown in Figure 2. From Figure 2, we know that although the quadratic polynomial curve model can fit the basic trend of the sequence y_t, there is still a certain gap between the fitted value and the actual value. The residual sequence ε_t has an obvious highly positive autocorrelation problem, and some residuals exceed two standard deviations, such as 1995-1999 and 1998-1999.

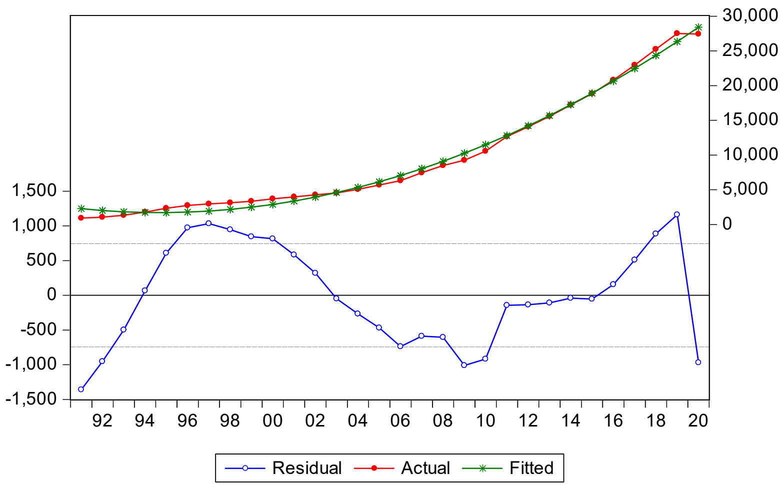

Figure 2 Fitting effect diagram of deterministic trend. Source: Author

3.2 Fitting of the remaining information of the regression model

The quadratic polynomial curve model is used to fit the deterministic trend and the results show that $D.W. = 0.49$, indicating that the residual sequence ε_t is highly autocorrelated. Therefore, ADF unit root test is performed on the residual sequence ε_t, and the test results are shown in Table 2. It is clear that at the significance level of 5%, the residual sequence is stable. Then, autocorrelation and partial autocorrelation graph are made for residual sequences ε_t, as shown in Figure 3.

Table 2 ADF test of residual sequence

	significance level	t-statistic	prob.*
ADF test statistics		-2.3494	0.0206
test critical values	1%	-2.6471	
	5%	-1.9529	
	10%	-1.6100	

Source: Author

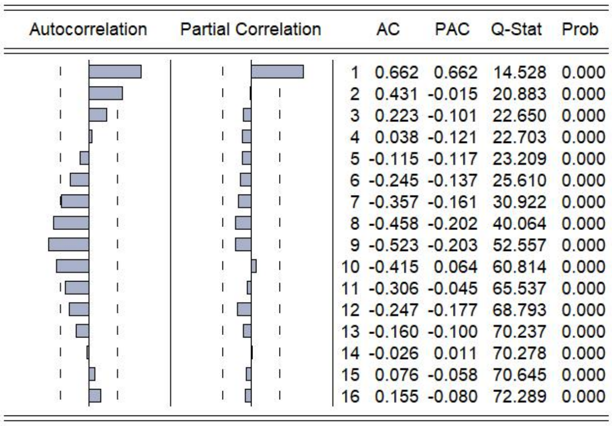

Figure 3 Autocorrelation and partial autocorrelation of residual sequence. Source: Author

By observing the autocorrelation and partial autocorrelation of the residual sequence ε_t, it is found that the autocorrelation coefficient of the residual sequence shows damping sinusoidal attenuation, and the partial autocorrelation shows first-order truncation characteristics. Based on the tailing of the autocorrelation coefficient and the truncation characteristic of the partial autocorrelation coefficient, the AR(1) model can be initially selected and established. The paper also experiments with several AR(p) models of various orders and compares the goodness of fit, D.W. and the significance of each variable. Finally, the centralized AR(1) model is selected to extract the residual sequence ε_t. The estimation results of AR(1) model are shown in Formula (5):

$$\varepsilon_t = 0.71\varepsilon_{t-1} + v_t$$

$$(5.67) \qquad (5)$$

$$R^2 = 0.5321, \qquad D.W. = 1.15$$

It is evident from the estimation outcomes of Formula (5) that: $R^2 = 0.5321$ indicating the model fits well. The parameter has passed the significance test at the significance level of 5%; $D.W. = 1.15$, indicating that the autocorrelation of residual sequence v_t is not obvious. Autocorrelation and partial autocorrelation graphs are drawn for the residual sequence v_t of Formula (5), as shown in Figure 4.

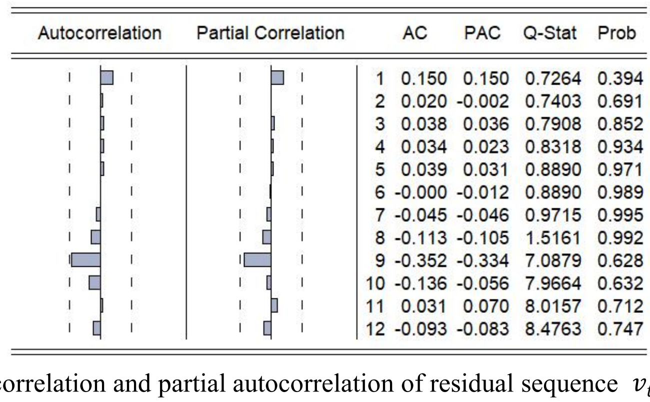

Figure 4 Autocorrelation and partial autocorrelation of residual sequence v_t. Source: Author

From Figure 4, the autocorrelation coefficient and partial autocorrelation coefficient of the residual sequence v_t are within two standard deviations and passes the white noise test. Therefore, it is believed that model AR(1) could better extract the relevant information in the process.

3.3 The residual autoregression model

The separate estimates of deterministic information and random information based on the aforementioned methods neglect the influence of correlation, which will impair the precision of model fitting because there is a correlation between the residual sequence ε_t and sequence fitting value $\hat{y}_t$. The model represented in Formula (6) can be constructed by merging the deterministic trend extraction model and the random information extraction model.

$$\begin{cases} y_t = 4554.07 - 557.67t + 44.76t^2 + \varepsilon_t \\ \varepsilon_t = 0.71\varepsilon_{t-1} + v_t \\ E(v_t) = 0, VAR(v_t) = \sigma^2, Cov(v_t, v_{t-i}) = 0, \forall i \geq 1 \end{cases} \quad (6)$$

In the estimation process of Formula (6), we can also get the following results: $R^2 = 0.997$, which shows that the model fits well; $F = 3000.11, p = 0.000$ indicates that the model passes the F-test at the significance level of 5%; In the T-test for each parameter, $p = 0.000$ indicates that all the parameters pass the T-test at the significance level of 5%; $D.W. = 1.29$, while $d_L = 1.20$ and $d_U = 1.65$, indicates that the autocorrelation of residual v_t is not obvious.

The fitting effect diagram of residual autoregressive model drawn by EViews8.0 is shown in Figure 5. The result shows that the residual autoregressive model can fit the sequence well. In general, there is a small gap between the fitting value and the actual value. However, there is a large fitting gap in 2020. Affected by the COVID-19 pandemic, people are spending less, leading to a decline in real consumption, which deviates from the long-term trend.

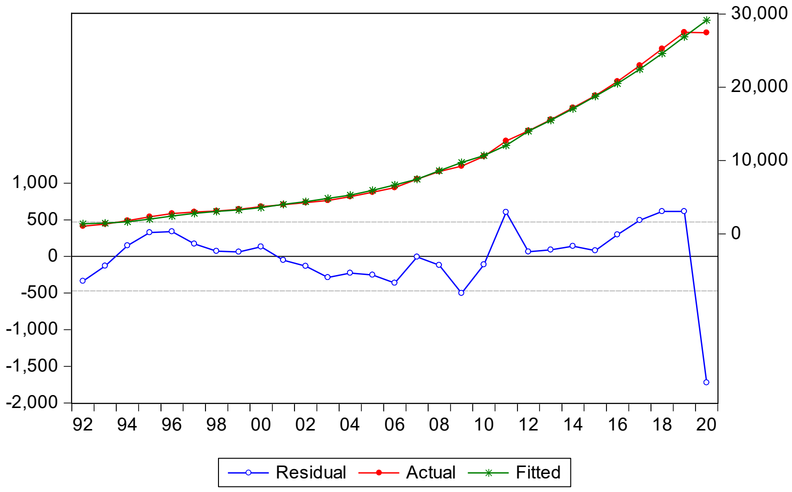

Figure 5 Fitting effect diagram of residual autoregressive model. Source: Author

3.4 Evaluation of the model

The absolute value of average relative error measures the prediction effect by calculating the ratio of the deviation between the actual value and the predicted value to the actual value. It can not only avoid the problem of inconsistent weights and measures of mean absolute error, but also solve the problem of relative error offset, and has good evaluation performance. The calculation formula is shown in Formula (7):

$$MAPE = \frac{1}{n}\sum\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad (7)$$

From Formula (7), $MAPE = 5.66\%$ shows that the fitting effect of residual autoregression model meets the requirements.

IV. Prediction results and analysis

The absolute value of average relative error $MAPE = 5.66\%$ shows that the residual autoregressive model established in this paper has good fitting effect. Table 3 shows the prediction results of China's resident consumption level from 2021 to 2023 adapting this model. According to the prediction, China's resident consumption level in 2021 will exceed 30,000 yuan, with an annual growth rate of about 7%.

Table 3 Forecast results of resident consumption in China in 2021-2023

year	predicted value (unit: yuan)
2021	30282
2022	32544
2023	34896

Source: Author

V. Conclusions

This paper compares different consumption forecasting methods and their characteristics, and finally chooses residual autoregressive model to forecast resident consumption. Residual autoregressive model combines the advantages of deterministic analysis method and stochastic analysis method, it preserves complete data information and improves the prediction accuracy of the model. The estimation results of the residual autoregressive model in this paper show that the model has good fitting effect, and the model passes goodness of fit test, F-test, T-test and D.W. test, and the absolute relative error is 5.66%. Finally, this model is used to forecast the consumption level of Chinese residents in 2021-2023. The results show that the average growth rate of Chinese resident consumption in the next three years is about 7% under the condition that the external environment does not change greatly.

***Acknowledgments*:** Key Project of Universities Humanities and Social Science Research in Anhui Province (SK2021A0613), the Key Project of Chaohu University (XLZ-202104), Social Science Innovation and Development Research Project in Anhui Province (2018CX119).

REFERENCES

[1] Feng Ying. *Empirical test of consumer confidence index on consumer consumption forecasting.* Statistics and Decision, 2013(15):15-18.
[2] Wang Jinan. *Empirical research on influencing Factors of Household Consumption*. Fujian Forum, 2014(9):39-43.
[3] Deng Yuxuan. *A study on the spatial structure of financial technology influencing household consumption growth*. Accounting Learning, 2021(21):151-153.
[4] Wang L. *The impact of gender structure difference on Household consumption in China*. Commercial Economics Research, 2021(14):54-57.
[5] Li Jie. *Analysis modeling and prediction of anhui province residents' consumption level based on ARIMA model.* Modern Business, 2017(01):195-196.
[6] Xiao Liang. *Forecasting household consumption level based on seasonal ARIMA model.* Statistics and Decision,2016 (8):83-86.
[7] Lu Erpo. *Research on energy demand forecasting model in China.* Statistics and Decision, 2005(10):29-31.
[8] Wang Yan. *Applied time Series Analysis*. Renmin University of China Press, 2015:160-166.
[9] Xu Guoxiang. *Statistical Prediction and decision.* Shanghai University of Finance and Economics Press, 2016:195-196.
[10] Li Yanyi et al. *EViews Statistical Analysis and Application*. Publishing House of Electronics Industry, 2013:147-160.