The Impact of Covid-19 Pandemic on China's CPI: Overall, Food and Healthcare

Yuntong Guo

Social Science Department, The Chinese University of Hong Kong,
Hong Kong, 999077, China

1155172884@link.cuhk.edu.hk

Abstract. This paper will look at how the Chinese epidemic affects the Consumer Price Index (CPI). The ARIMA model is used to predict the overall CPI and CPI in the food and medical fields utilizing time-series information spanning 2010 to 2020, therefore illuminating their anticipated patterns as well as drivers' significant shifts. Although the actual CPI values fluctuated beyond the model forecast values in certain periods, the study found that the COVID-19 outbreak generated considerable changes in the CPI and that the ARIMA model demonstrated good accuracy in CPI predictions before and during the epidemic. The significance of this study lies in providing a profound understanding of economic fluctuations in the context of a sudden public health event by analyzing the impact of the Chinese epidemic on CPI. This provides data support for government and enterprises in formulating economic policies and business strategies and guides residents' consumption decisions. For statistical authorities, it is suggested that they should fully consider the impact of unexpected events and external shocks when conducting economic forecasts and establish emergency response plans and adjustment mechanisms to enhance the adaptability and accuracy of forecast models. For consumers, it is recommended to treat price fluctuations rationally during emergencies such as the epidemic, avoid blind hoarding or excessive consumption, and adjust consumption behavior according to the actual economic environment.

Keywords: ARIMA Model, CPI, Covid-19 Pandemic.

1 Introduction

The Consumer Price Index (CPI), a crucial socioeconomic indicator, shows how costs have changed as time passes for a range of consumer goods and services. It is one of the most important instruments for assessing inflation or deflation and has a big influence on business decisions and operations, government economic policies, and consumer behavior. Changes in CPI are strongly correlated with changes in people's actual purchasing power and quality of life. Therefore, it is crucial to accurately project changes in the CPI. The global economy has been greatly impacted by the COVID-19 epidemic, and there have been notable fluctuations in the CPI. A deeper comprehen

Q. Wu et al. (eds.), *Proceedings of the 2024 3rd International Conference on Public Service, Economic Management and Sustainable Development (PESD 2024)*, Advances in Economics, Business and Management Research 309,
https://doi.org/10.2991/978-94-6463-598-0_19

sion of the extent of changes in the country's economy and the lives of its residents may be attained by looking at the number of changes in the CPI. People will be able to comprehend the scope of the recent coronavirus epidemic that is impacting us better as a result. In addition, analyzing changes in the CPI, which measures consumer prices, before and after the pandemic is crucial in recognizing how the market functions and developing sensible fiscal strategies.

In today's economic analysis, time series data analysis plays an important role. It not only helps to understand the data change patterns in the past, but also provides support for the prediction of future trends. Because of its excellent prediction capabilities, the ARIMA model is often used in time series research. Therefore, the goal of this research is to foresee and predict CPI in the wake of the current coronavirus epidemic using the ARIMA model. Along with the overall CPI, this will focus on the food and healthcare industries. This will assist in revealing potential CPI trends and patterns and examining the underlying reasons for the variances. According to a recent research, Autoregressive Integrated Moving Average (ARIMA) models are gaining traction as a substitute for CPI prediction. In 2009, Wang and colleagues published an ARIMA model that estimated the model using the Particle Swarm Optimization method. This approach yielded a more potent and efficient optimizing performance in contrast to conventional techniques [1].

A researcher used an ARIMA (12,1,12) model to anticipate a nation's monthly CPI in 2010, showcasing the model's ability to produce precise forecasts [2]. Based on deidentified CPI data, another group of researchers described the Swiss National Bank's ARIMA model and shown that it performed better in inflation predictions than pertinent benchmarks [3]. An ARIMA (1,1,1) model for Botswana's inflation was then found. The model was improved by adding the ARCH/GARCH model to account for series volatility [4]. To predict Zambia's inflation rate, researchers conducted a comparison between ARIMA models and Holt's Exponential Smoothing. They found that the ARIMA (12, 1, 0) model was best suited for CPI time series data [5]. In 2017, another scientist concentrated on utilizing ARIMA models to forecast CPI, stressing the need of comprehending the underlying workings of CPI to ensure precise forecasts [6]. In more recent research, utilizing CPI data from four major Chinese cities, some data scientists suggested the Generalized Space Time Autoregressive Integrated (GSTARI) model for CPI prediction and compared its accuracy with the traditional ARIMA model [7].

ARIMA models were also employed by to forecast the CPI in Uganda and Somaliland respectively [8, 9]. This illustrates the adaptability of ARIMA in diverse economic environments. All things considered, the literature demonstrates how well ARIMA models work for CPI forecasting in a variety of locations and eras, demonstrating its versatility and precision in estimating inflation rates. The broad applicability of ARIMA in economic forecasting is demonstrated in another recent journal further emphasis on the use of ARIMA models in GDP and CPI forecasts for the Jordanian economy [10].

Although many studies have verified the effectiveness of the ARIMA model, the differences in different data sets and specific application scenarios make further research necessary. In this paper, it is still hoped to further verify the effectiveness of the ARIMA model in practical applications through empirical analysis and extend the discussion of its limitations and improvement directions.

2 Research Design

2.1 Data Sources and Pre-Processing

The time series data selected for this study comes from the National Bureau of Statistics. The data period covers 2010 to 2021. The reason for selecting this period of data is that it has good continuity and representativeness and includes the period before the outbreak of the new crown epidemic and the initial development stage. In comparison, this period has better observability and effectiveness.

For the data processing part, the original data firstly need to be integrated and selected by three main research CPI series for this study: overall CPI, food CPI and healthcare CPI, then re-arranged the datum points by the year 2010 according to the CPI calculation method and organized the data into a new CPI table. Finally, a data classification of selecting data before and after 2020 would be added.

2.2 Unit Root Test

Stationarity is an important aspect of time series analysis [11]. Thus, to verify the stationarity of the time series, the given data must undergo a unit root analysis before modeling. The augmented Dickey-Fuller (ADF) test will be used in this investigation to evaluate unit roots. If the test's p-value is less than the specified significant threshold of 0.05, the null hypothesis is invalidated, and the time series is regarded as stationary. The distinction stage may be skipped and the ARMA model of the components may be employed immediately if the series is already stationary. Alternatively, if the time series is not stationary, distinguish it until it becomes stationary. The goal of differencing is to identify the difference between two adjacent data components.

Table 1. ADF test

	t	p
	CPI-overall	
Ln value	-3.045	0.1200
1st order difference	-10.695	0.0000
	CPI-food	
Ln value	-2.189	0.4959
1st order difference	-11.365	0.0000
	CPI-healthcare	
Ln value	-2.927	0.1535
1st order difference	-12.219	0.0000

In Table 1, for the logarithmic values of all three CPI indicators, the original p-values are all greater than 0.05, indicating that these logarithmic value series are non-stationary, so differentiation is required. After differencing, it can be observed that for the logarithmic values after the first-order difference, the p-values become less than 0.05, indicating that these first-order difference sequences have become stationary sequences.

2.3 Arima Model

Data collected in time series evaluation and prediction may be accomplished with the help of the mathematical framework ARIMA. It is appropriate for the study of non-stationary time series data, incorporating the elements of autoregression (AR), difference (I), and moving average (MA).

The ARIMA model's expression is ARIMA (p, d, q), where p is the order of the AR portion, which reflects the order of the autoregressive component, or how many previous point values are utilized in the model to forecast the current value. The mathematical expression form of p is like this:

$$y_t = \phi 1 y_{t-1} + \phi 2 y_{t-2} + \cdots + \phi p y_{t-p} + \epsilon_t \tag{1}$$

In this formula, y_t is the time series value at time t, ϕ are the autoregressive coefficients, ϵ_t is the white noise error term.

The order of the MA portion, represented by q, represents the sequence of the moving average part. This order represents the number of past error terms that are employed in the model to predict the current value. The mathematical expression form of q is like this:

$$y_t = \epsilon_t + \theta 1 \epsilon_{t-1} + \theta 2 \epsilon_{t-2} + \cdots + \theta q \epsilon_{t-q} \tag{2}$$

In this formula, y_t is the value of the time series at time t, θ are the moving average coefficients, are the error terms.

The integrated part, d, it is utilized when describing a set of modifications that will convert a non-stationary time series into a stationary time series.

3 Empirical Results and Analysis

3.1 Decision of Order

First, create the graph using the PACF and ACF techniques in this part; secondly, examine the ACF graph to identify the shortened lag order, which will serve as an estimate of q. Then search the PACF graph for the reduced lag order, which provides a p. estimation.

Observing the resulting ACF and PACF images helps one to ascertain the values of p and q; subsequently, the value of d may be found depending on the ADF test as illustrated in figure 1.

Therefore, it can be decided that the model for overall CPI is ARIMA (10, 1, 2), the model for food is also ARIMA (10, 1, 2), the model for healthcare is ARIMA (5, 1, 1).

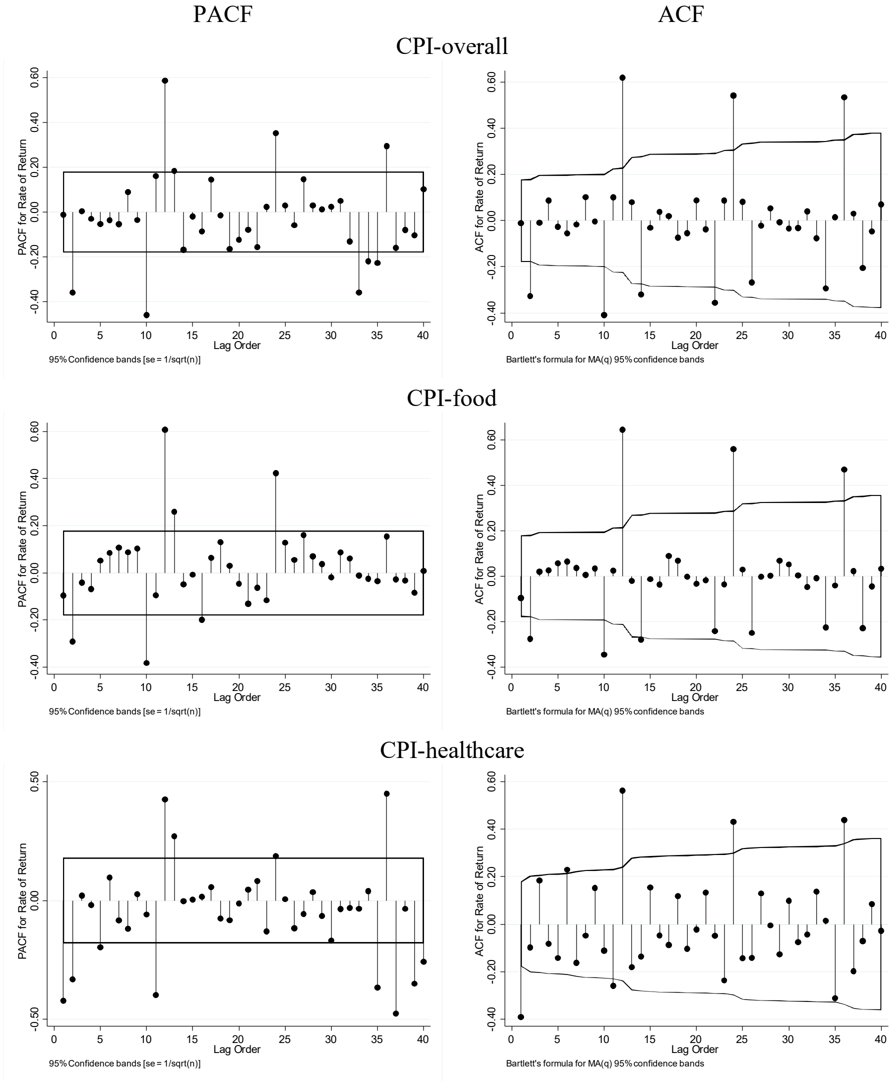

Fig. 1. ARMA (p, q) identification.
Photo credit: Original

Table 2. Residual test

Model	Portmanteau (Q) statistic	Prob > chi2
CPI-overall	58.8323	0.0000
CPI-food	180.8752	0.0000
CPI-healthcare	101.3535	0.0000

According to the explanation in Table 2, it can be clearly understood that all models have not passed the residual test, which means that the existence of these parameters is not statistically significantly supported, in other words, the parameters involved in the model are not very significant, therefore, they do not have a significant impact on explaining the data. So, the models do not fit white noise. This means that the model may be underfitting or overfitting.

3.2 Forecast Results Explanations

Table 3, 4, and 5 show the forecasts for CPI after applying the model, and with the predicted values, differences and different percentages together after the COVID-19. Figure 2, 3, 4 visualize the differences between predicted values and real values.

Based on a review of the first overall CPI chart, it can be observed that, while both movements are downward, the anticipated trend is declining more than the real pattern. In comparison, the true CPI values have declined more than what was expected. But what is different is that, contrary to the predicted trend, the actual overall CPI value continued to decline after August 2020, especially after October 2020, when it experienced a sharp drop. From the chart data, people can clearly observe that the overall CPI reached its overall peak in February 2020, which is 132.468, and then fell directly to 128.983 in November of the same year; unlike the forecast results, the predicted CPI value will rebound from 131 to 132 in July 2020.

Table 3. Overall CPI Empirical Results ARIMA (10, 1, 2)

	Overall CPI	PV	Differences	%
Apr-19	126.043			
May-19	126.531			
Jun-19	127.119			
Jul-19	127.607			
Aug-19	127.871			
Sep-19	128.368			
Oct-19	129.751			
Nov-19	129.631			
Dec-19	128.643			
Jan-20	131.152			
Feb-20	132.468	132.0144	0.45361	0.34%
Mar-20	131.326	131.5574	-0.23137	-0.18%
Apr-20	130.202	131.4035	-1.20153	-0.91%
May-20	129.568	131.6695	-2.10146	-1.60%
Jun-20	130.297	131.9311	-1.6341	-1.24%
Jul-20	131.053	132.3329	-1.27992	-0.97%
Aug-20	130.94	132.1377	-1.19769	-0.91%
Sep-20	130.551	132.3811	-1.83008	-1.38%
Oct-20	130.4	133.63	-3.23004	-2.42%
Nov-20	128.983	132.9883	-4.00526	-3.01%

Early in the pandemic, from February to July 2020, the trends and approximative values of the expected CPI values and the actual values were rather similar, according to the visualization chart of food CPI. The model does exhibit some degree of accuracy since the expected values coincided with the actual values even between March and August 2020. But the expected trend and the actual trend diverged significantly following August of the same year. The upward expected trend and the declining actual trend highlight this difference most.

Overall, the actual data shows a pattern with significant fluctuations and trends. At the end of 2019 and early 2020, CPI rose rapidly, then reached a peak in February 2020, and then fluctuated downward; while the forecast data started in February 2020, showing a similar trend to CPI, but with a lag in some periods and smaller fluctuations.

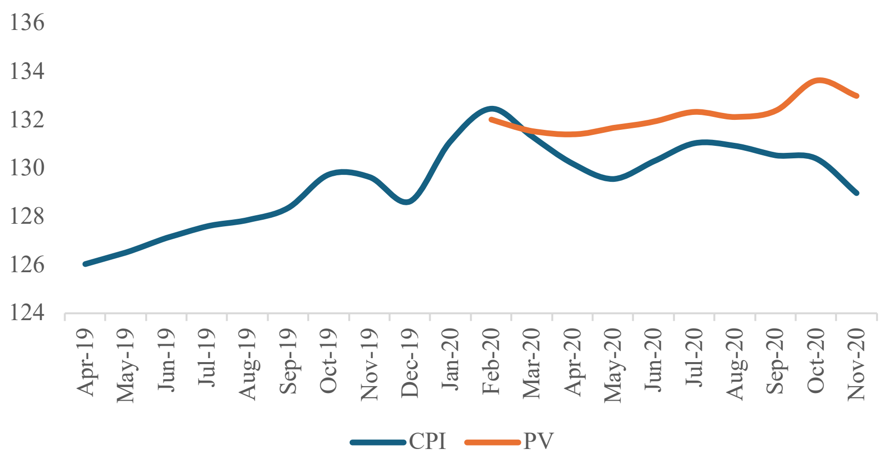

Fig. 2. CPI-overall before and after Covid-19.
Photo credit: Original

The forecast data of food CPI also differed greatly from the actual data in the later period of the COVID-19. From February to June 2020, both sets of data showed a downward trend, from 143 and 142 to 141 respectively. In July of the same year, as predicted, CPI data began to rise, but in August, there was a significant difference. The actual food CPI index stopped rising after reaching 142.555 in July, and began to decline instead, falling to 140.311 in November. The forecast data tended to be flat from August to November and remained at around 144.

Table 4. Food CPI Empirical Results ARIMA (10, 1, 2)

	food	PV	Differences	%
Apr-19	140.874			
May-19	140.315			
Jun-19	141.979			

Jul-19	143.271			
Aug-19	143.565			
Sep-19	142.235			
Oct-19	141.528			
Nov-19	140.733			
Dec-19	138.499			
Jan-20	140.391			
Feb-20	143.404	142.1139	1.29009	0.91%
Mar-20	140.837	141.6203	-0.78329	-0.55%
Apr-20	140.31	140.3176	-0.00759	-0.01%
May-20	139.753	140.0287	-0.27567	-0.20%
Jun-20	141.411	141.4219	-0.01093	-0.01%
Jul-20	142.555	143.4709	-0.91586	-0.64%
Aug-20	142.847	144.5191	-1.67214	-1.16%
Sep-20	141.666	144.4515	-2.78553	-1.93%
Oct-20	141.103	144.7477	-3.64468	-2.52%
Nov-20	140.311	144.6545	-4.34345	-3.00%

According to the above description, the image can show the above differences more clearly. The actual data demonstrates a trend of decreasing, rising, and subsequently decreasing, whereas the predicted data almost exhibits a trend of decreasing, growing, and ultimately stabilizing.

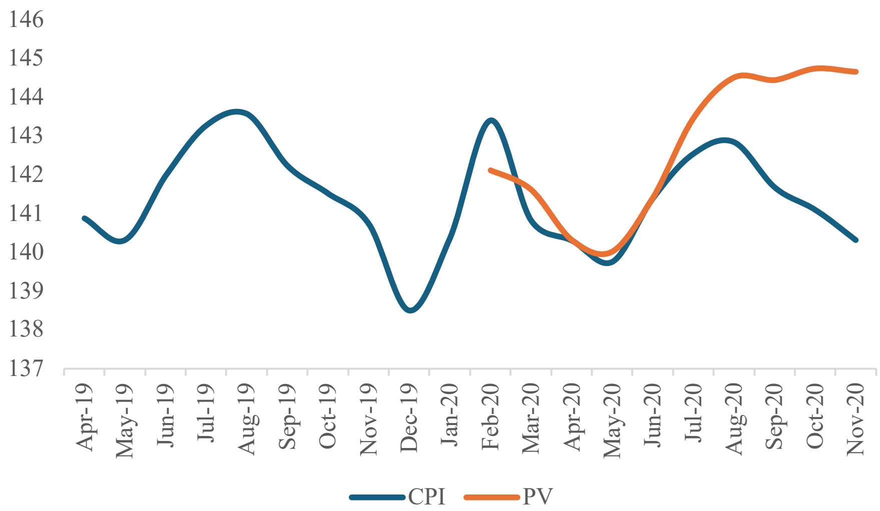

Fig. 3. CPI-food before and after Covid-19.
Photo credit: Original

The discrepancy between the predicted and actual CPIs for healthcare is not particularly substantial; however, particular gaps are more apparent during specific periods.

The starting point of the predicted and actual values, that is, February 2020, was around 122; however, in March 2020, the actual healthcare CPI rose to 123, and this data was maintained until May of the same year, and then showed a downward trend in the following month; the predicted CPI results first maintained the predicted value of 122 as the starting point for two months, and then increased to 123, and there was no more obvious change afterwards.

Table 5. healthcare CPI Empirical Results ARIMA (5, 1, 5)

	healthcare	PV	Differences	%
Sep-19	120.876			
Oct-19	121.835			
Nov-19	121.592			
Dec-19	121.361			
Jan-20	123.424			
Feb-20	122.212	122.7503	-0.5383	-0.44%
Mar-20	123.077	122.7652	0.31179	0.25%
Apr-20	123.074	123.4468	-0.37284	-0.30%
May-20	123.312	123.3387	-0.02665	-0.02%
Jun-20	122.83	123.3573	-0.52728	-0.43%

The visualization of this data allows the public to clearly understand the differences and similarities between the predicted and actual groups. The actual CPI value trend is to rise sharply first, then fall slowly, and finally fall sharply again; unlike the actual situation, the predicted CPI first rises slowly and then falls slowly.

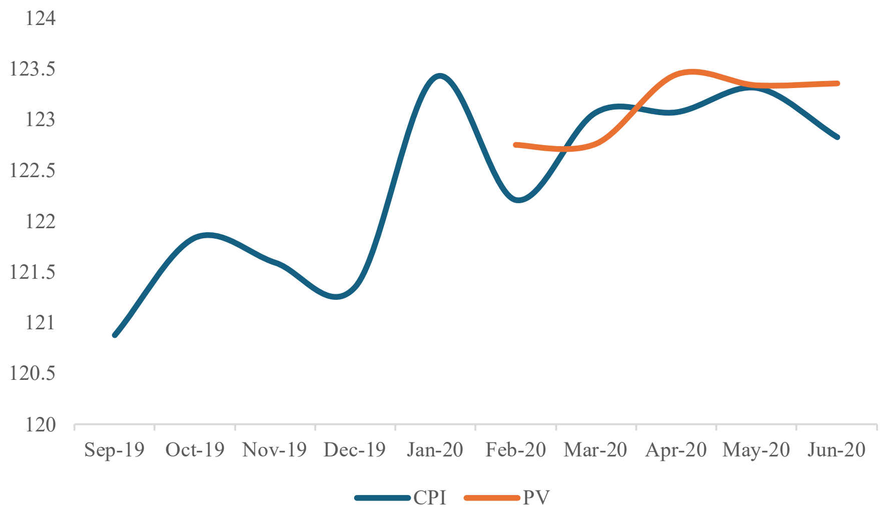

Fig. 4. CPI-healthcare before and after Covid-19.
Photo credit: Original

In general, the public's perception of price is significantly distinct from the results of such forecasts. The primary source of individuals' perceptions of the Consumer Price Index (CPI) is high-frequency consumer products that are used in daily life, such as food and masks. What are the potential causes of these outcomes?

The main possible reason is that in CPI surveys, statisticians usually exclude outliers and only count prices within a reasonable range. For example, if the prices of certain commodities fluctuate wildly due to short-term supply and demand imbalances, these abnormal prices may not be included in CPI calculations. This approach helps maintain the stability of the index but may not fully reflect the true fluctuations of the market during abnormal times (such as epidemics). In other words, during the data selection process, statisticians may filter abnormal data based on the requirements of statistical methods and data quality control principles. Although this approach can maintain the stability and reliability of the data, it may also result in some actual price fluctuations not being included in the statistics, especially during the epidemic period when prices fluctuate violently. In addition, due to social distancing measures and market lockdowns, statisticians are unable to conduct on-site surveys and instead rely on online platforms and telephone surveys, which may suffer from sample bias and insufficient data collection. For example, some regions and commodities may be underestimated or ignored in statistics, resulting in the CPI not fully reflecting the actual market price.

Other reasons include that in the early stages of the epidemic, supply chain disruptions caused the prices of certain commodities (such as masks, disinfectants, etc.) to rise sharply, but as supply recovered, prices stabilized or declined. At the same time, insufficient demand has caused the prices of other commodities (such as clothing, luxury goods, etc.) to fall. These interacting forces might cause the CPI to drop generally. CPI calculations also rely on a set basket of goods and services, which is typically revised every few years to reflect variations in consumer expenditure patterns. During the epidemic, despite significant changes in consumer behavior, the commodity basket calculated by CPI may not be adjusted in time. It may also be because, during the pandemic, traditional price survey methods (such as on-site price collection) may be restricted and replaced by new methods such as telephone surveys and online price collection. These changes may cause the accuracy and consistency of data collection to be affected.

4 Conclusion

Like previous literature, this paper once again verifies the effectiveness and applicability of the ARIMA model in predicting data, especially in the data analysis of CPI before and after the COVID-19 pandemic. The empirical results of this paper further support this view. Therefore, there are relatively obvious limitations in the evaluation of the effectiveness and applicability of the model.

Through time series analysis of CPI data, this article demonstrates the powerful tool of the ARIMA model in economic forecasting and emphasizes the important role of time series analysis in understanding and predicting economic trends. In addition, this article also reveals the significant impact of the COVID-19 epidemic on CPI, prompting researchers to consider the impact of emergencies and external shocks when making economic forecasts.

This study's focus is CPI, with a particular emphasis on the COVID-19 outbreak's effects. A comparison of the Consumer Price Index (CPI) before and after the outbreak shows the enormous effect of the pandemic on the economy and people's lives. This point of view offers a fresh study direction and is somewhat rare in current literature. In addition to the overall CPI, this article also conducts a separate analysis of the CPI in the food and healthcare sectors, further refining the ARIMA model's predictive capabilities in different fields.

This article verifies the effectiveness of the ARIMA model in time series forecasting by analyzing the time series data of an economic indicator. Empirical results show that the ARIMA model can fit the data well and make accurate predictions. However, the limitations of models in processing different types of data also deserve further research and exploration. Future research can combine other statistical models to improve prediction accuracy and broad application.

References

1. Wang, H., & Zhao, W. (2009). Arima model estimated by particle swarm optimization algorithm for consumer price index forecasting. Artificial Intelligence and Computational Intelligence.
2. Dongdong, W. (2010). The consumer price index forecast based on Arima Model. 2010 WASE International Conference on Information Engineering.
3. Huwiler, M., & Kaufmann, D. (2013, January 1). Combining disaggregates forecasts for inflation: The SNB's Arima Model: The SNB's Arima Model. Swiss National Bank Economic Study.
4. Molebatsi, K., & Raboloko, M. (2016). Time series modelling of inflation in Botswana using monthly consumer price indices. International Journal of Economics and Finance.
5. Jere, S., & Siyanga, M. (2016). Forecasting inflation rate of Zambia using Holt's exponential smoothing. Open Journal of Statistics, 06(02), 363-372.
6. Xiao, Z. (2017). CPI prediction based on Arima Model. Proceedings of the 7th International Conference on Education, Management, Information and Computer Science (ICEMC 2017).
7. Ji, S., Dong, J., Wang, Y., & Liu, Y. (2020). Research on CPI prediction based on space-time model. 2019 6th International Conference on Dependable Systems and Their Applications (DSA).
8. Mwanga, Y. (2020). Arima forecasting model for Uganda's consumer price index. American Journal of Theoretical and Applied Statistics, 9(5), 238.
9. Mohamed, J. (2020). Time series modeling and Forecasting of Somaliland Consumer Price index: A comparison of Arima and regression with Arima errors. American Journal of Theoretical and Applied Statistics, 9(4), 143.
10. Ghazo, A. (2021). Applying the Arima model to the process of forecasting GDP and CPI in the Jordanian economy. International Journal of Financial Research, 12(3), 70.
11. Mushtaq, R. (2011, August 17). Augmented dickey fuller test. SSRN.