Journal of Infrastructure, Policy and Development 2024, 8(9), 5270.
https://doi.org/10.24294/jipd.v8i9.5270

Article

Modelling and predicting the Consumer Price Index in Saudi Arabia

Ayman Mahgoub[1,*], Teg Alam[2]

[1] Department of Management, College of Business Administration, Prince Sattam bin Abdulaziz University, Al Kharj 11942, Kingdom of Saudi Arabia

[2] Department of Industrial Engineering, College of Engineering, Prince Sattam bin Abdulaziz University, Al Kharj 11942, Kingdom of Saudi Arabia

*** Corresponding author:** Ayman Mahgoub, am.mohammed@psau.edu.sa

CITATION

Mahgoub A., Alam T. (2024). Modelling and predicting the Consumer Price Index in Saudi Arabia. Journal of Infrastructure, Policy and Development. 8(9): 5270. https://doi.org/10.24294/jipd.v8i9.5270

ARTICLE INFO

Received: 17 March 2024
Accepted: 31 July 2024
Available online: 11 September 2024

Abstract: The Consumer Price Index (CPI) is a vital gauge of economic performance, reflecting fluctuations in the costs of goods, services, and other commodities essential to consumers. It is a cornerstone measure used to evaluate inflationary trends within an economy. In Saudi Arabia, forecasting the Consumer Price Index (CPI) relies on analyzing CPI data from 2013 to 2020, structured as an annual time series. Through rigorous analysis, the SARMA (0,1,0) (12,0,12) model emerges as the most suitable approach for estimating this dataset. Notably, this model stands out for its ability to accurately capture seasonal variations and autocorrelation patterns inherent in the CPI data. An advantageous feature of the chosen SARMA model is its self-sufficiency, eliminating the need for supplementary models to address outliers or disruptions in the data. Moreover, the residuals produced by the model adhere closely to the fundamental assumptions of least squares principles, underscoring the precision of the estimation process. The fitted SARMA model demonstrates stability, exhibiting minimal deviations from expected trends. This stability enhances its utility in estimating the average prices of goods and services, thus providing valuable insights for policymakers and stakeholders. Utilizing the SARMA (0,1,0) (12,0,12) model enables the projection of future values of the Consumer Price Index (CPI) in Saudi Arabia for the period from June 2020 to June 2021. The model forecasts a consistent upward trajectory in monthly CPI values, reflecting ongoing economic inflationary pressures. In summary, the findings underscore the efficacy of the SARMA model in predicting CPI trends in Saudi Arabia. This model is a valuable tool for policymakers, enabling informed decision-making in response to evolving economic dynamics and facilitating effective policies to address inflationary challenges.

Keywords: Saudi Arabia; Consumer Price Index; forecasting; Box-Jenkins method; SARMA models; correlogram

1. Introduction

When determining the level of inflation, the Consumer Price Index (CPI) is widely used. When the inflation rate in an economy is kept to a minimum and doesn't undergo any substantial shifts, the economy is said to be strong. Low and stable inflation rates make it easier for businesses and investors to make investments, provide employment opportunities, and protect the national currency from being devalued. The consumer price index should remain steady to benefit the economy's medium to long-term development. On the other hand, fluctuations and unexpected inflation rates lead to uncertainty and reduce the level of confidence that businesses have, as well as economic activity and development initiatives.

When inflation rates are high, consumers' purchasing power, unemployment, the real exchange rate, and competitiveness are all negatively impacted. Additionally, unanticipated inflation may result in significant economic shocks with repercussions analogous to those of the previous example. The authorities in charge of monetary policy are always keen to get forecasts about the behavior of price indices in the economy.

It has been noted that the inflation rate in Saudi Arabia has increased in recent years, exceeding that of the European Union. This research investigates the issue of inflation in Saudi Arabia, which is the subject of this study. The rise in value-added tax (VAT) and administrative fees may be the critical factor responsible for this spike in inflation. Despite this, policymakers may get significant insights from this knowledge of the Consumer Price Index (CPI) behavior. It can generate a feeling of optimism and drive them to develop effective policies and methods to offset or reduce the unanticipated repercussions that inflation will have on the economy.

In this research, we evaluate monthly data on the consumer price index from January 2013 to May 2020, totaling 89 observations. We accomplish this by using the well-established Box-Jenkins approach based on the data. The General Authority for Statistics is the source of these particular facts. This investigation uses a seasonal autoregressive moving average model with the parameters (0,1,0) (12,0,12). Based on the findings, the fitted model is consistent, which offers a solid basis for formulating policies and forecasting the average pricing of products and amenities. Section 2 provides a concise analysis of the existing body of research. Following the presentation of the results in Section (4), Section (3) provides an overview of the approach. In the fifth and last Section, concluding remarks are presented.

2. Literature review

The existing research landscape within Saudi Arabia regarding the modeling and forecasting of the Consumer Price Index (CPI) using econometric models needs to be more extensive. More studies are required to address this specific aspect of economic analysis within the Saudi context.

Various econometric models have been employed globally for forecasting time series data, including Moving Averages (MA), Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ES), and Autoregressive Conditional Heteroscedasticity (ARCH). Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and Vector Autoregressive Conditional Heteroscedasticity (VARCH) models have been utilized as generalized variants of the traditional VAR models.

In a study by Adams et al. (2014), the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model was applied to quarterly data on Nigeria's CPI inflation rate from 1980 to 2010. Their findings identified the ARIMA (1,2,1) model as the most suitable, enabling the prediction of Nigeria's CPI for the subsequent five years.

The seasonal autoregressive integrated moving-average model by Ashuri and Lu (2010) is the most accurate time series technique for in-sample CCI forecasting, while the Holt-Winters exponential smoothing model is the most exact for out-of-sample forecasts. Time series models outperform the ENR's subject matter experts' CCI projection in out-of-sample forecasts.

Similarly, Mordi et al. (2012) proposed a short-term forecasting model by employing time series models for individual CPI components at specific disaggregation levels. The short-term forecasts were derived from the weighted sum of twelve CPI component projections.

In China, Zhang et al. (2013) constructed an ARMA model to accurately forecast CPI from 1995 to 2008, demonstrating the model's efficacy in capturing CPI dynamics over the specified period.

Other studies have focused on specific countries such as Bangladesh. Akhter (2013) utilized the Seasonal Autoregressive Integrated Moving Average (SARIMA) model to forecast short-term inflation rates using monthly CPI data from 2000 to 2012.

Moreover, Norbert, et al. (2016) utilized the ARIMA model to fit CPI data from 1995 to 2015, highlighting the increasing price trend over the specified period.

Nyoni (2019) employed the Box-Jenkins ARIMA technique in Saudi Arabia to predict the CPI, indicating an expected price increase in the coming decades.

Ji et al. (2020) utilized the ARIMA model for CPI prediction but identified its limitation in considering time effects. To address this, they proposed an improved GSTARI (Generalized Space-Time Autoregressive Integrated) model, which demonstrated higher accuracy than ARIMA in CPI prediction.

Additionally, studies in other countries, such as Indonesia, have applied ARIMA models to forecast CPI, with Ahmar, Rahman and Mulbar (2018) identifying the ARIMA (1,0,0) model as suitable for CPI forecasting in Indonesia.

Zhang and Yang (2023) also examined the house value index and three possible indicators to confirm their predicting abilities pre- and post-COVID-19 and suggest time series research after the pandemic.

Several forecasting models, such as artificial neural networks (ANN) and other predictive models, have been used in predicting crude oil prices and other applications dependent on time series data (Alam, 2019; Alam and AlArjani, 2021).

Xu and Zhang (2022) examined how Chinese house costs changed from 2010 to 2019. The directed acyclic graph and monthly data from 99 of China's biggest cities show how house prices in different levels of cities are related to each other simultaneously. The PC algorithm finds the pattern of causes, and the LiNGAM algorithm finds the line of causes. This information is then used for innovation accounting analysis.

Furthermore, Mustapa and Ismail (2019) developed a hybrid ARIMA-GARCH model for forecasting S&P 500 stock prices, demonstrating its superiority in stock price prediction.

These studies underscore the significance of employing appropriate econometric models for CPI forecasting, considering each economy's unique characteristics and dynamics. They provide valuable insights into various modeling approaches' potential applications and limitations, contributing to advancing economic forecasting methodologies.

3. Materials and methods

3.1. Consumer price index time series data

The consumer price index data for the Kingdom of Saudi Arabia is obtained monthly from The General Authority for Statistics, covering the period from January 2013 to May 2020. The mean of this series is 97.08 index points, with a standard deviation of 2.10. In January 2013, the index had a low value of 92.70 points; in January 2018, it reached a high value of 100.90. This analysis suggests that the consumer price index experiences significant fluctuations and has a skewed curve to the left (skewness = −0.50) with a relatively flat peak (kurtosis = 2.55). The software programs used were MS Excel 2010 and EViews 8.1.

3.2. Testing stationarity

The first step in the analysis is to examine the data for stationarity. A time series data is considered stationary if its mean, variance, and auto-covance remain constant (do not fluctuate) over time. All applied studies that use time-series data assume that the series is stationary. However, it's important to remember that this is not just an assumption, but a fundamental requirement. Without this characteristic, the regression obtained between time series variables is often spurious, even though the coefficient of determination selection is high. This is a critical issue that the audience, as researchers, analysts, and students in the field of statistics, econometrics, or data analysis, are uniquely positioned to address. Time series data usually have a trend component, which may reflect certain conditions that affect variables in the same or opposite directions (Chatfield and Xing, 2019). Thus, if the series exhibits increasing or decreasing general trends, it is difficult to rely on its mean values in forecasting because one value of the average cannot be used to express all the series values, giving prediction values less (higher) than those obtained in the presence of increasing (declining) trend, meaning that the average ability to predict is weak (Box et al., 2011).

Further, since the variance expresses the degree of uncertainty in prediction, differences in its values from one group of the series to another make the average of the higher-variance group weaker than the average of the lower-variance group in the prediction, as the uncertainty in the high-variance group is more excellent. Thus, stability of the variance is also a property of stationarity or stillness. Graphical analysis, Box-Jenkins method, and unit root tests are among the most popular stationarity testing methods. The time series plot provides information on the nature of the data at its most basic level. If the graph exhibits any trend, the series is non-stationary; however, it could also be non-stationary without showing any trend. This paper adopts the Box-Jenkins method and the unit root or the augmented Dickey-Fuller (ADF) test.

3.2.1. Augmented Dickey–Fuller test

The augmented Dickey-Fuller (ADF) test is one of the most widely used testing methods for the stationarity of a time series. It is a modification of the Dickey-Fuller (DF) test to account for autocorrelated error terms of the test equation. Otherwise, the test will be invalid if these terms are correlated. To prevent this, lagged values of the differenced dependent variable were added to the test equation. For the series vt, the basic ADF test equation takes the following form, where p is the number of lags:

$$\Delta v_t = \mu v_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta v_{t-i} + u_t$$

This equation has three possible forms, depending on whether the series has a constant term (as above), continuous and trend terms, or without constant and trend terms. In each case, the lag lengths are determined either by the Akaike information criterion (AIC), the Schwartz Bayesian criterion (SBC), or the Lagrange multiplier (LM) test. One has to be careful when selecting lag lengths. Ideally, it should be relatively small to secure enough degrees of freedom and relatively large to prevent autocorrelation in the error terms (Harris, 1992). This equation tests the null hypothesis that the series is non-stationary ($\mu = 0$) against the alternative ($\mu < 0$). We accept the alternative hypothesis and conclude that the series is stationary and integrated into order one.

3.3. Box-Jenkins method

Box and Jenkins method employs the statistical Autoregressive Moving Average (ARMA) model in modeling, analyzing, and monitoring the behavior of time series variables. The ARMA model's stochastic process or sequence is a polynomial sum of Auto-Regressive (AR) and Moving Average (MA) terms. The general theoretical model appeared in 1951 in the thesis of the New Zealand statistician Peter Whittle, who tested hypotheses chronologically. This work was circulated in 1971 in a book by the statisticians George Box and Gwilym Jenkins. A Box-Jenkins methodology prediction model is built in four stages (Pandey and Basu, 2020).

3.3.1. Model specifications

Once the data passes the stationarity test, a mathematical model is specified to describe the variable's data-generating process. This model is based on some statistical measures that distinguish it from another. Experience and knowledge gained from other studies can help in that direction.

ARIMA models, along with correlograms, are pivotal tools in the analysis of time series data. These tools, along with plots of the autocorrelation function (ACF) or the partial autocorrelation function (PACF), are used to identify the behavior of the data. If the summary mean is stationary, the correlogram spikes will decrease quickly toward zero (Liu, Liu and Shi, 2020). The following **Table 1** shows the behavior of the (ACF) and (PACF) as model identification tools:

Table 1. Behavior of the ACF and PACF for seasonal AR, MA, and ARMA processes.

	ARMA (P, Q) s	**SMA (Q)** s	**SAR (P)** s
ACF	Tails off	Cuts off lag QS	Tails off
PACF	Cuts off lag QP	Tails off	Tails off

Identification refers to the determining of the optimum lags *p*, *d*, *q*, and *P*, *D*, and *Q* of the SARIMA model, which is then denoted SARIMA (*p*,*d*,*q*) (*P*,*D*,*Q*), where *p*, *q*, and *d* are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average components of the non-seasonal model, and *P*, *D* and *Q* refer to the seasonal autoregressive, differencing (integrated), and moving average orders, respectively. To obtain the values p, d, q, P, D, and Q, we use the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) (Box, et al., 2015).

The equation for a p-th order autoregressive (AR) model, denoted AR (p), is written as:

$$v_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i v_{t-i} + u_t$$

where v_t is the original series, α_0, α_1 and α_p are the (AR) parameters to be estimated, and u_t is a random error.

The equation for a q-th order moving average model, denoted MA(q), is written as:

$$v_t = \beta_0 + \sum_{i=1}^{q} \beta_i u_{t-i} + u_t$$

where:v_t is the original series, β_0, β_1 and β_p are the MA(q) average parameters to be estimated, and u_t is series of unknown random errors or residuals.

The Equation for an ARMA(p,q) model takes the form:

$$v_t = \mu + \alpha_1 v_{t-1} + \alpha_2 v_{t-2} + \ldots + \alpha_p v_{t-p} - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \beta_q u_{t-q} + u_t$$

where $\{v_t\}$, α_1, α_2, …, β_1, β_2, …, are as previously defined.

The (seasonal) SARMA (p,q) (P,Q) model equation is:

$$v_t = \mu + \sum_{i=1}^{p} \alpha_i v_{t-i} + u_t + \sum_{i=1}^{P} \Phi_i V_{t-is} - \sum_{i=1}^{q} \beta_i u_{t-i} + \sum_{i=1}^{Q} \Theta_i u_{t-is}$$

where $\{v_t\}$, $\{\alpha_i\}$, and $\{\beta_i\}$are as previously defined, $\{\Phi\}$ and $\{\Theta\}$ are their seasonal counterparts.

The equation for an ARIMA (p,d,q) model is:

$$\alpha_p(B)(1-B)^d v_t = \beta_q(B) u_t$$

When there is a seasonal influence in a time series, the ARIMA model is sometimes referred to as the SARIMA model. The above equation represents the multiplicative SARIMA model, which may be written as follows:

$$\Phi(B^S)\alpha_p(B)\nabla_S^D \nabla_{v_t}^d = \mu + \Theta(B^S)\beta_q(B)u_t$$

where:

$\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p$ (the p order of AR operator)

$\beta(B) = 1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q$ (the q order of MA operator)

$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_P B^{PS}$(the P order of seasonal AR operator)

$\Theta(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{2S} + \cdots + \Theta_Q B^{qQS}$ (the Q order of seasonal MA operator)

$\nabla_{v_t}^d = (1-B)^d$and $\nabla_S^D = (1-B)^D$ refers nonseasonal and seasonal operator.

μ is a constant, v_t is the time series, s is the season, and u_t is the random error (Beran, 2017).

3.3.2. Model estimation

After nominating one or more suitable models to describe the observed time series, we estimate the model (s) parameters using observed data and special statistical estimation methods. We estimate $c, \varphi, \theta, \Phi, and\ \Theta$ of the *SARIMA (p, d, q) (P, D, Q)* model in the above equation (Shmueli and Lichtendahl, 2016).

3.3.3. Model diagnostic tests

After estimating the SARIMA model, diagnostic tests are conducted to ensure its reliability and validity. Firstly, we assess the significance of the predicted parameters and residuals derived from the fitted SARIMA model. This involves examining whether the estimated coefficients and residuals exhibit statistical importance, indicating their reliability in capturing the underlying patterns in the data.

Secondly, we validate the model's assumptions by comparing the actual observations with the values predicted by the candidate model. This validation process helps determine the model's accuracy and adequacy in representing the observed data. If the model passes these diagnostic tests, it is considered suitable for generating future forecasts.

However, if the model fails to meet the criteria set by these tests, indicating potential inadequacies or deficiencies in its specification, we initiate a corrective process. This involves setting up a new model and repeating the estimation process, diagnostic tests, and validation steps outlined above. By iteratively refining the model based on diagnostic feedback, our unwavering goal is to develop a robust and reliable forecasting framework that accurately captures the underlying dynamics of the data (Montgomery et al., 2015).

3.3.4. Forecasting the study variable

Forecasting depends on assuming and fitting a stationary model to ensure the residuals don't show any trends or valuable information. Comparing the model's results to real-world data requires this step. Time series datasets are usually split into two parts to make this review easier. The first part is used to fit the model. For example, the second part tests how well the model can predict the future. We usually test how well a model, or even several types of models, reflects accurate data by lowering a measure like the root mean square error (Hyndman and Athanasopoulos, 2018). Using the difference between actual and projected numbers as a standard, this statistic measures how well the model performs. This error measure can be minimized to make models more dependable and accurate in their predictions, which increases their usefulness in real life.

4. Data and model selection

A total of 89 monthly observations of the Consumer Price Index (CPI) for Saudi Arabia are included in the dataset. These observations can be found from January 2013 through May 2020. A graph that illustrates the Consumer Price Index (CPI) series at its initial level is shown in **Figure 1**, which allows for a visual examination of the CPI series. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the data are shown in **Figure 2**, which also illustrates the series' correlogram.

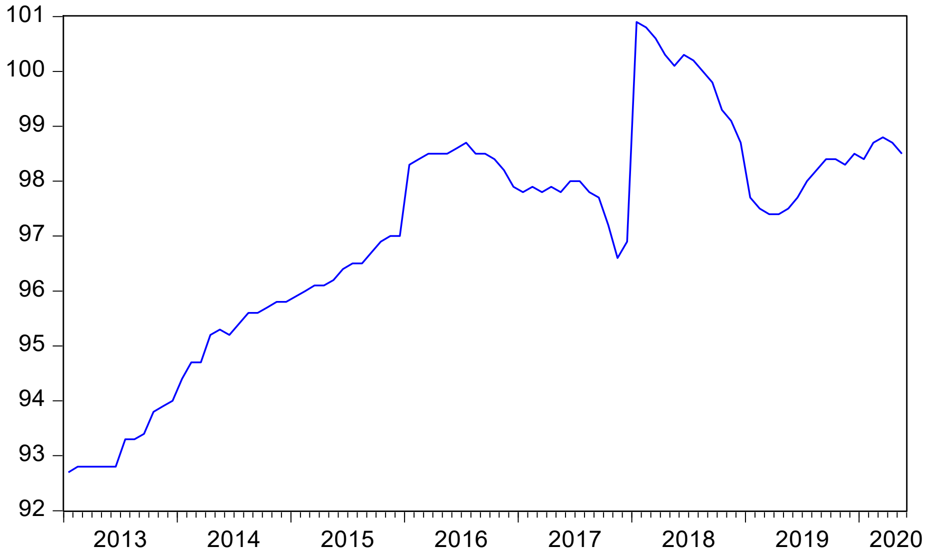
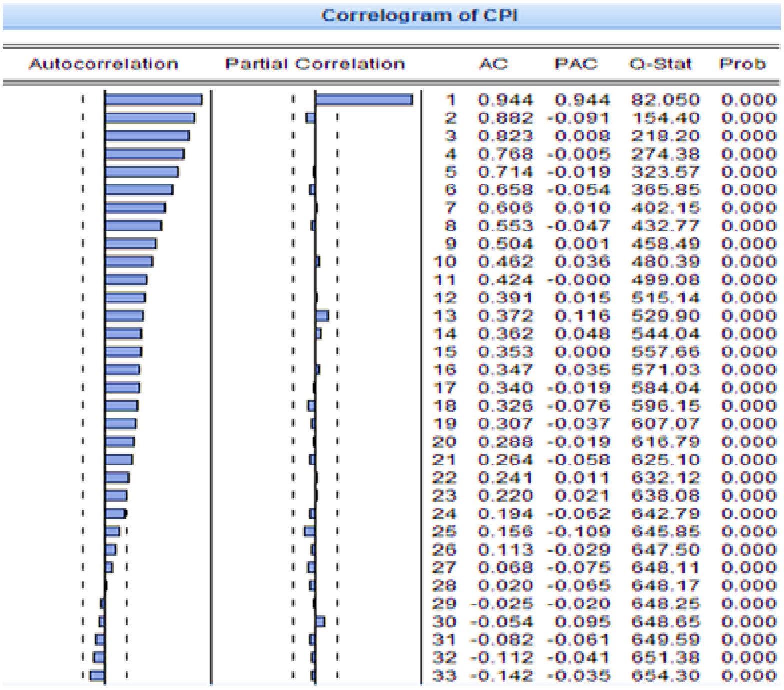


Figure 1. CPI graph.

Source: Author's construction based on data from general authority for statistics, Saudi Arabia (General Authority for Statistics, 2024).



	AC	PAC	Q-Stat	Prob
1	0.944	0.944	82.050	0.000
2	0.882	-0.091	154.40	0.000
3	0.823	0.008	218.20	0.000
4	0.768	-0.005	274.38	0.000
5	0.714	-0.019	323.57	0.000
6	0.658	-0.054	365.85	0.000
7	0.606	0.010	402.15	0.000
8	0.553	-0.047	432.77	0.000
9	0.504	0.001	458.49	0.000
10	0.462	0.036	480.39	0.000
11	0.424	-0.000	499.08	0.000
12	0.391	0.015	515.14	0.000
13	0.372	0.116	529.90	0.000
14	0.362	0.048	544.04	0.000
15	0.353	0.000	557.66	0.000
16	0.347	0.035	571.03	0.000
17	0.340	-0.019	584.04	0.000
18	0.326	-0.076	596.15	0.000
19	0.307	-0.037	607.07	0.000
20	0.288	-0.019	616.79	0.000
21	0.264	-0.058	625.10	0.000
22	0.241	0.011	632.12	0.000
23	0.220	0.021	638.08	0.000
24	0.194	-0.062	642.79	0.000
25	0.156	-0.109	645.85	0.000
26	0.113	-0.029	647.50	0.000
27	0.068	-0.075	648.11	0.000
28	0.020	-0.065	648.17	0.000
29	-0.025	-0.020	648.25	0.000
30	-0.054	0.095	648.65	0.000
31	-0.082	-0.061	649.59	0.000
32	-0.112	-0.041	651.38	0.000
33	-0.142	-0.035	654.30	0.000

Figure 2. Correlogram of the CPI.

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

The Consumer Price Index (CPI) series, as meticulously examined in **Figure 1** and **Figure 2**, exhibits nonstationary behavior at its initial level. This conclusion is further supported by the comprehensive findings of the Augmented Dickey-Fuller (ADF) test, as detailed in **Table 2**. Across all levels of significance, the absolute value of the ADF test statistic consistently falls below the critical values, providing robust and compelling evidence that the series is not stationary.

Table 2. Augmented Dickey-Fuller test for the CPI.

Augmented Dickey-Fuller Test Equation for CPI			
The CPI has a unit root, according to the null hypothesis.			
Exogenous: Constant in nature			
(Automatic, SIC-based, max lag = 11) Length of Lag: 0			
		t-test	*P*-value
The Dickey-Fuller test statistic has been improved.		−2.02	0.28
Test critical values:	0.01 level	−3.51	
	0.05 level	−2.89	
	0.10 level	−2.58	

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

Consequently, the ADF test is carried out on the first contrast of the series, and the results are shown in **Table 3** in more detail. Concurrently, the correlogram of the first differenced CPI series is shown in **Figure 3**. The data show a seasonal influence, as shown by the fact that the ACF and the PACF show periodic rises each year.

Table 3. Augmented Dickey-Fuller test for the first difference of the CPI.

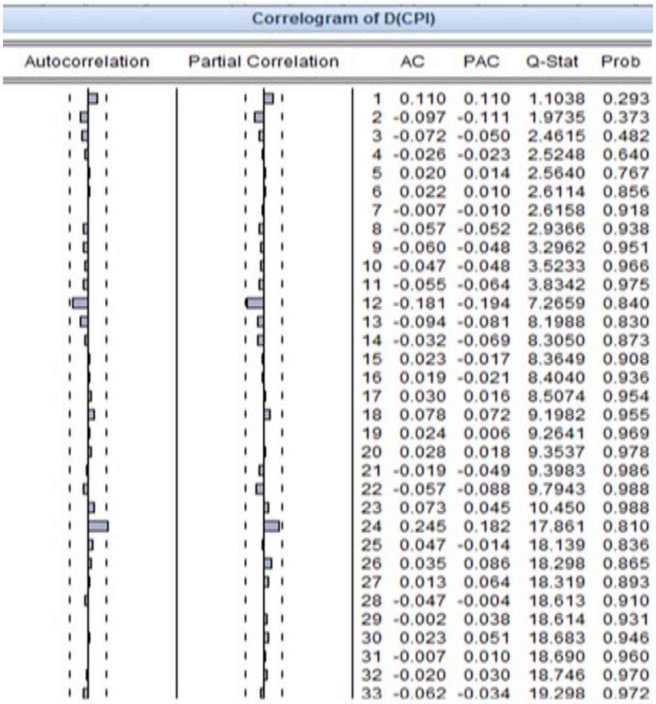
Augmented Dickey-Fuller Test Equation for D(CPI)			
The null hypothesis is that there is a unit root in the first differential CPI.			
Exogenous: Constant in nature			
(Automatic, SIC-based, max lag = 11) Length of Lag: 0			
		t-test	*P*-value
Augmented Dickey-Fuller test statistic		−8.24	0.00
Test critical values:	0.01 level	−3.51	
	0.05level	−2.90	
	0.10 level	−2.58	

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

The research unequivocally demonstrates that the Consumer Price Index (CPI) series exhibits nonstationary behavior at its initial level. This necessitates the transformation of the data into its first difference. Furthermore, the presence of seasonal trends, as indicated by both the ACF and the PACF, strongly underscores the need for a mixed model that incorporates a seasonal influence. This model could be a crucial tool for more accurate and insightful statistical analysis and modeling.

The Dickey-Fuller test statistic for the CPI is reported as −2.02, and the associated *p*-value is 0.28. The test statistic is not significantly lower than the critical values at any conventional significance level (0.01,0.05,0.10), and the *p*-value is higher than the typical significance level of 0.05. Therefore, insufficient evidence exists to reject the null hypothesis of non-stationarity for the CPI data.

The Augmented Dickey-Fuller test statistic is reported as −8.24, indicating a substantial negative value. This suggests a significant deviation from the null hypothesis of non-stationarity.



Correlogram of D(CPI)

	AC	PAC	Q-Stat	Prob
1	0.110	0.110	1.1038	0.293
2	-0.097	-0.111	1.9735	0.373
3	-0.072	-0.050	2.4615	0.482
4	-0.026	-0.023	2.5248	0.640
5	0.020	0.014	2.5640	0.767
6	0.022	0.010	2.6114	0.856
7	-0.007	-0.010	2.6158	0.918
8	-0.057	-0.052	2.9366	0.938
9	-0.060	-0.048	3.2962	0.951
10	-0.047	-0.048	3.5233	0.966
11	-0.055	-0.064	3.8342	0.975
12	-0.181	-0.194	7.2659	0.840
13	-0.094	-0.081	8.1988	0.830
14	-0.032	-0.069	8.3050	0.873
15	0.023	-0.017	8.3649	0.908
16	0.019	-0.021	8.4040	0.936
17	0.030	0.016	8.5074	0.954
18	0.078	0.072	9.1982	0.955
19	0.024	0.006	9.2641	0.969
20	0.028	0.018	9.3537	0.978
21	-0.019	-0.049	9.3983	0.986
22	-0.057	-0.088	9.7943	0.988
23	0.073	0.045	10.450	0.988
24	0.245	0.182	17.861	0.810
25	0.047	-0.014	18.139	0.836
26	0.035	0.086	18.298	0.865
27	0.013	0.064	18.319	0.893
28	-0.047	-0.004	18.613	0.910
29	-0.002	0.038	18.614	0.931
30	0.023	0.051	18.683	0.946
31	-0.007	0.010	18.690	0.960
32	-0.020	0.030	18.746	0.970
33	-0.062	-0.034	19.298	0.972

Figure 3. The Correlogram of the first difference of the CPI.

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

The *p*-value associated with the test statistic is reported as 0.00, below any conventional significance level (e.g., 0.01, 0.05, 0.10). This indicates strong evidence against the null hypothesis, further supporting the rejection of non-stationarity.

Additionally, the critical values provided at significance levels of 0.01, 0.05, and 0.10 are negative, with the test statistic significantly lower than these vital values. This reinforces the rejection of the null hypothesis and confirms the data's stationarity.

Table 4 presents the results of model selection criteria for time series analysis, focusing on various SARMA and ARMA models. The aim is to identify the most suitable model for forecasting based on several evaluation metrics, including the Sum of Squared Residuals (SQR), Akaike Information Criterion (AIC), Schwarz Bayesian Criterion (SBC), Theil's Inequality Coefficient (Theil I C), and Root Mean Square Error (RMSE).

Table 4. Model selection criteria.

Model	SQR	AIC	SBC	Theil I C	RMSE
SARMA (0,1,0) (12,0,12)	15.79699	1.345911	1.43791	0.00233	0.45591
ARMA (1,1,1) (12,0,12)	15.61594	1.402014	1.556513	0.0053	1.04584
SARMA (1,1,0) (12,0,12)	15.69560	1.380436	1.504036	0.01196	2.3145
SARMA (0,1,1) (12,0,12)	15.64521	1.362572	1.485242	0.012176	2.4560
SARMA (0,1,0) (12,0,0)	20.67671	1.588783	1.650118	0.008886	1.7249
SARMA (0,1,0) (0,0,12)	21.33977	1.466567	1.522870	0.01007	1.9415

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia Source (General Authority for Statistics, 2024).

Based on the model selection criteria, the SARMA (0,1,0) (12,0,12) model performs the best, as it has the lowest SQR, AIC, SBC, Theil I C, and RMSE values among the models evaluated. This analysis indicates that the SARMA (0,1,0) (12,0,12) model provides the data's most accurate and economical representation, making it the preferred choice for forecasting based on the given evaluation metrics.

Table 5. Estimated model.

Dependent Variable: D(CPI)				
Method: Least Squares				
Date: 07/12/20 Time: 23:53				
Sample (adjusted): 2014M02 2020M05				
Included observations: 76 after adjustments				
Convergence achieved after 24 iterations				
MA Backcast: 2013M02 2014M01				
Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	0.062020	0.051312	1.208683	0.0207
SAR (12)	−0.775836	0.059655	−13.00537	0.0000
SMA (12)	0.957902	0.042473	22.55309	0.0000
R-squared	0.262127	Mean dependent var		0.053947
Adjusted R-squared	0.241911	S.D. dependent var		0.534276
S.E. of regression	0.465185	Akaike info criterion		1.345911
Sum squared resid	15.79699	Schwarz criterion		1.437913
Log likelihood	−48.14460	Hannan-Quinn criter		1.382679
F-statistic	12.96650	Durbin-Watson stat		1.811752
Prob(F-statistic)	0.000015			
Inverted AR Roots	0.95 − 0.25i	0.95 + 0.25i	0.69 − 0.69i	0.69 − 0.69i
	0.25 + 0.95i	0.25 − 0.95i	−0.25 − 0.95i	−0.25 + 0.95i
	−0.69 + 0.69i	−0.69 + 0.69i	−0.95 + 0.25i	−0.95 − 0.25i
Inverted MA Roots	0.96 + 0.26i	0.96 − 0.26i	0.70 + 0.70i	0.70 − 0.70i
	0.26 + 0.96i	0.26 − 0.96i	−0.26 − 0.96i	−0.26 + 0.96i
	−0.70 − 0.70i	−0.70 − 0.70i	−0.96 + 0.26i	−0.96 − 0.26i

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

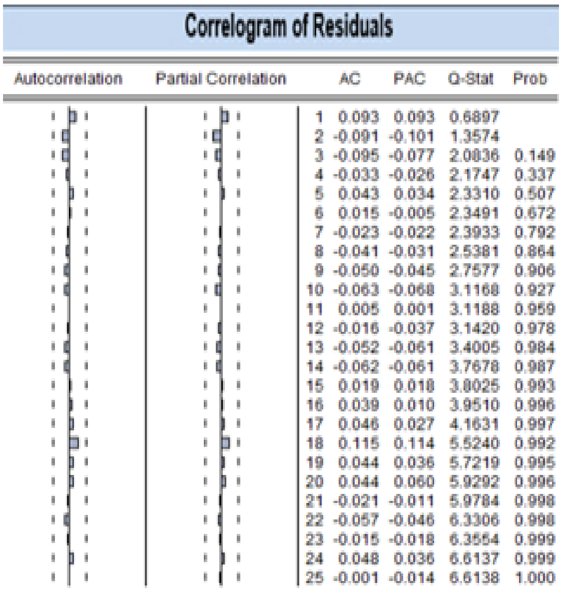
The estimated model in **Table 5** provides valuable insights into the relationship between the CPI's first difference and its determinants. The significant coefficients and model fit statistics suggest that the model adequately captures the dynamics of the CPI data and can be used for forecasting and analysis purposes.

Figure 4 shows the estimated models fitted, residual, and actual values. It shows a tight relationship between the precise and fitted values. However, we need to check the model's usefulness by looking at the white noise and stability of the residual series. There must be room for improvement in the model, and further information extraction is required if this series does not conform to the white noise hypothesis. Predictions based on an unstable series are useless.

Figure 4. Actual, fitted, and residual series.

Source: Author's construction based on data from general authority for statistics, Saudi Arabia. (General Authority for Statistics, 2024).

Figure 5 displays the correlogram and Q-statistic for the model residuals. We observe that the correlation and partial correlation values fall between ± 2SE of their values, suggesting that the stability and white-noise conditions hold for these residuals. Furthermore, the *P*-values of the Q-statistics are all significantly different from the 1% level.



Correlogram of Residuals

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.093	0.093	0.6897	
		2	-0.091	-0.101	1.3574	
		3	-0.095	-0.077	2.0836	0.149
		4	-0.033	-0.026	2.1747	0.337
		5	0.043	0.034	2.3310	0.507
		6	0.015	-0.005	2.3491	0.672
		7	-0.023	-0.022	2.3933	0.792
		8	-0.041	-0.031	2.5381	0.864
		9	-0.050	-0.045	2.7577	0.906
		10	-0.063	-0.068	3.1168	0.927
		11	0.005	0.001	3.1188	0.959
		12	-0.016	-0.037	3.1420	0.978
		13	-0.052	-0.061	3.4005	0.984
		14	-0.062	-0.061	3.7678	0.987
		15	0.019	0.018	3.8025	0.993
		16	0.039	0.010	3.9510	0.996
		17	0.046	0.027	4.1631	0.997
		18	0.115	0.114	5.5240	0.992
		19	0.044	0.036	5.7219	0.995
		20	0.044	0.060	5.9292	0.996
		21	-0.021	-0.011	5.9784	0.998
		22	-0.057	-0.046	6.3306	0.998
		23	-0.015	-0.018	6.3554	0.999
		24	0.048	0.036	6.6137	0.999
		25	-0.001	-0.014	6.6138	1.000

Figure 5. Correlogram and Q-statistic for the residuals.

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia (General Authority for Statistics, 2024).

Moreover, we test for stationarity of the residuals with different exogenous terms, namely with a constant, with a constant and trend, and without constant and trend, respectively. **Table 6** shows these results which confirm stationarity of the errors.

Table 6. Tests for residuals' stationarity.

ERRORS has a unit root (null hypothesis). **Exogenous: Constant** **(Automatic, based on SIC, maxlag = 11) Lag Length: 0 (Automatic, based on SIC)**			
		t-test	*p*-value
Augmented Dickey-Fuller test statistic		−8.64	0.000
Values to be tested:	0.01 level	−3.53	
	0.05 level	−2.90	
	0.10 level	−2.59	
ERRORS has a unit root (null hypothesis). Exogenous: Constant, Linear Trend (Automatic, based on SIC, maxlag = 11) Lag Length: 0			
		t-test	*p*-value
Augmented Dickey-Fuller test statistic		−8.58	0.000
Values to be tested:	0.01 level	−4.09	
	0.05 level	−3.48	
	0.10 level	−3.17	
ERRORS has a unit root (null hypothesis). Exogenous: None (Automatic, based on SIC, maxlag = 11) Lag Length: 0			
		t-test	*p*-value
Augmented Dickey-Fuller test statistic		−8.699547	0.000
Values to be tested:	0.01 level	−2.598416	
	0.05 level	−1.945525	
	0.10 level	−1.613760	

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

The results of the tests for residuals' stationarity indicate strong evidence against the presence of a unit root, suggesting that the residuals are stationary. This is consistent across different specifications of exogenous variables, including constant and linear trend terms. These findings support the adequacy of the model and provide assurance for further statistical analyses and inference based on the residuals.

Finally, to assess the forecasting power of the selected model, we have used the root mean square error (RMSE) and Theil Inequality coefficient, which takes values between 0 and 1. The closer this value to 0, the better the forecast model. As shown in **Table 4**, the selected model SARMA (0,1,0) (12,0,12) outperforms other models as it gives the lowest values for both RMSE and Theil inequality coefficient, estimated at 0.4559 and 0.0023, respectively. The selected model SARMA (0,1,0) (12,0,12) is used to generate forecast values for a year ahead. **Table 7** reports the forecast values from June 2020 to June 2021.

Table 7. Forecast values for the next 12 month.

Month	**Forecast**
2020M06	98.7
2020M07	98.6

Table 7. (*Continued*).

Month	Forecast
2020M08	98.6
2020M09	98.6
2020M10	98.7
2020M11	98.8
2020M12	98.9
2021M01	99.0
2021M02	99.0
2021M03	99.0
2021M04	99.1
2021M05	99.3
2021M06	99.4

Source: Author's calculations based on data from general authority for statistics, Saudi Arabia source (General Authority for Statistics, 2024).

Obtaining significant insights into the anticipated path that the Consumer Price Index will take is made possible by the predicted values for the following twelve months. Even though predictions are susceptible to uncertainty and may be impacted by unanticipated occurrences or shifts in economic circumstances, they are a valuable instrument for planning and decision-making in a variety of economic sectors throughout the world. These projections, which are based on rigorous statistical research, support a better understanding of inflation dynamics. They also help the process of making informed decisions.

5. Conclusion

Even though our time series analysis and forecasting provide vital insights into the proposed behavior of the Consumer Price Index (CPI) over the next twelve months (June 2020 to June 2021), it is essential to keep in mind that there are some restrictions to take into consideration. Notably, the results of our research showed that the Consumer Price Index (CPI) series needs to display stationarity, which challenges the assumptions often made in time series analysis.

The SARMA (0,1,0) (12,0,12) model emerged as the best option for modeling the CPI data, proving its capacity to estimate and represent the data effectively despite this constraint. Having said that, it is of the utmost importance to understand that no model is flawless and that various methods or adjustments may improve forecasting accuracy.

Looking ahead, there is significant potential for future research to explore more intricate modeling approaches or incorporate additional factors to address the non-stationarity of the CPI series and enhance forecasting accuracy. Moreover, there is a clear opportunity for research to focus on identifying and mitigating the impact of outliers and structural changes in the data, which would enhance the reliability of CPI estimates.

In conclusion, our research provides a robust foundation for forecasting average price patterns and offering insights that can inform policy decisions. However, it is evident that further study is necessary to address the limitations we have identified and to explore new avenues. By taking this approach, we can significantly enhance the robustness and effectiveness of CPI forecasting models.

Author contributions: Conceptualization, AM and TA; methodology, AM and TA; software, AM; validation, AM and TA; formal analysis, AM; investigation, AM and TA; resources, AM; data curation, AM; writing—original draft preparation, AM; writing—review and editing, TA; visualization, AM and TA; supervision, AM; project administration, AM; funding acquisition, AM. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1445).

Conflict of interest: The authors declare no conflict of interest.

References

Adams, S. O., Awujola, A., Alumgudu, A. I. (2014). Modeling Nigeria's Consumer Price Index Using ARIMA Model. International Journal of Development and Economic Sustainability, 2(2): 37-47.

Ahmar, A. S., GS, A. D., Listyorini, T., et al. (2018). Implementation of the ARIMA(p,d,q) method to forecasting CPI Data usingforecast packagein R Software. Journal of Physics: Conference Series, 1028, 012189. https://doi.org/10.1088/1742-6596/1028/1/012189

Ahmar, A. S., Rahman, A., Mulbar, U. (2018). α-Sutte Indicator: a new method for time series forecasting. Journal of Physics: Conference Series, 1040, 012018. https://doi.org/10.1088/1742-6596/1040/1/012018

Akhter, T., (2013). Short-term forecasting of inflation in Bangladesh with seasonal ARIMA processes. MPRA Paper, Munich University Library.

Alam, T. (2019). Forecasting exports and imports through artificial neural network and autoregressive integrated moving average. Decision Science Letters, 249–260. https://doi.org/10.5267/j.dsl.2019.2.001

Alam, T., AlArjani, A. (2021). A Comparative Study of CO2 Emission Forecasting in the Gulf Countries Using Autoregressive Integrated Moving Average, Artificial Neural Network, and Holt-Winters Exponential Smoothing Models. Advances in Meteorology, 2021, 1–9. https://doi.org/10.1155/2021/8322590

Almanjahie, I. M., Chikr-Elmezouar, Z., &Bachir, A. (2019). Modelling and forecasting the household water consumption in Saudi Arabia. Applied ecology and environmental research, 17(1), 1299-1309. http://dx.doi.org/10.15666/aeer/1701_12991309

Ashuri, B., Lu, J. (2010). Time series analysis of ENR construction cost index. Journal of Construction Engineering and Management, 136(11), 1227-1237. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000231

Beran, J. (2017). Mathematical Foundations of Time Series Analysis. Springer International Publishing. https://doi.org/10.1007/978-3-319-74380-6

Boniface, A., Martin, A. (2019). Time Series Modeling and Forecasting of Consumer Price Index in Ghana. Journal of Advances in Mathematics and Computer Science, 1–11. https://doi.org/10.9734/jamcs/2019/v32i130134

Box, G. E., Jenkins, G. M., Reinsel, G. C. (2011). Time series analysis: forecasting and control. John Wiley & Sons.

Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Chatfield, C., Xing, H. (2019). The Analysis of Time Series. Chapman and Hall/CRC. https://doi.org/10.1201/9781351259446

Du, Y., Cai, Y., Chen, M., et al. (2014). A Novel Divide-and-Conquer Model for CPI Prediction Using ARIMA, Gray Model and BPNN. Procedia Computer Science, 31, 842–851. https://doi.org/10.1016/j.procs.2014.05.335

Eissa, N. (2020). Forecasting the GDP per Capita for Egypt and Saudi Arabia Using ARIMA Models. Research in World Economy, 11(1), 247. https://doi.org/10.5430/rwe.v11n1p247

General Authority for Statistics. (2024). Construction and Building survey. Available online: https://www.stats.gov.sa/en/awareness-surveys/cb (accessed on 2 June 2023).

Gjika Dhamo, E., Puka, L., Zaçaj, O. (2018). Forecasting consumer price index (cpi) using time series models and multi regression models (albania case study). In: Proceedings of the 10th International Scientific Conference “Business and Management 2018.” https://doi.org/10.3846/bm.2018.51

Harris, R. I. (1992). Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. Economics letters, 38(4), 381-386. https://doi.org/10.1016/0165-1765(92)90022-Q

Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

Jere, S., Banda, A., Chilyabanyama, R., et al. (2019). Modeling Consumer Price Index in Zambia: A Comparative Study between Multicointegration and Arima Approach. Open Journal of Statistics, 09(02), 245–257. https://doi.org/10.4236/ojs.2019.92018

Ji, S., Dong, J., Wang, Y., et al. (2020). Research on CPI Prediction Based on Space-Time Model. In: Proceedings of the 2019 6th International Conference on Dependable Systems and Their Applications (DSA). https://doi.org/10.1109/dsa.2019.00058

Kharimah, F., Usman, M., Widiarti, W., Elfaki, F. A. M. (2015). Time series modeling and forecasting of the consumer price index Bandar Lampung. Science International Lahore, 27(5), 4619-4624.

Kuhe, D. A., Egemba, R. C. (2016). Modelling and Forecasting CPI Inflation in Nigeria: Application of Autoregressive Integrated Moving average Homoskedastic Model. Journal of Scientific and engineering Research, 3(2), 57-66.

Liu, T., Liu, S., Shi, L. (2020). Time Series Analysis Using SAS Enterprise Guide. Springer Singapore. https://doi.org/10.1007/978-981-15-0321-4

Masood, A., Bahrawi, J., Elfeki, A. (2019). Modeling annual rainfall time series in Saudi Arabia using first-order autoregressive AR (1) model. Arabian Journal of Geosciences, 12(6). https://doi.org/10.1007/s12517-019-4330-3

Montgomery, D. C., Jennings, C. L., Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.

Mordi, C. N. O, Adeby, M. A, Adamgbe, E. T (2012). Short-term inflation forecasting for monetary policy in Nigeria. Central Bank of Nigeria Occasion Paper No. 42

Mustapa, F. H., Ismail, M. T. (2019). Modelling and forecasting S&P 500 stock prices using hybrid Arima-Garch Model. Journal of Physics: Conference Series, 1366(1), 012130. https://doi.org/10.1088/1742-6596/1366/1/012130

Norbert, H. (2016). Modeling and Forecasting Consumer Price Index (Case of Rwanda). American Journal of Theoretical and Applied Statistics, 5(3), 101. https://doi.org/10.11648/j.ajtas.20160503.14

Nyoni, T. (2019). Predicting consumer price index in Saudi Arabia. available online: https://www.researchgate.net/publication/331546302_Predicting_Consumer_Price_Index_In_Saudi_Arabia (accessed on 2 February 2024).

Pandey, K., Basu, B. (2020). Mathematical modeling for short term indoor room temperature forecasting using Box-Jenkins models. Journal of Modelling in Management, 15(3), 1105–1136. https://doi.org/10.1108/jm2-08-2019-0182

Riofrío, J., Chang, O., Revelo-Fuelagán, E. J., et al. (2020). Forecasting the Consumer Price Index (CPI) of Ecuador: A Comparative Study of Predictive Models. International Journal on Advanced Science, Engineering and Information Technology, 10(3), 1078–1084. https://doi.org/10.18517/ijaseit.10.3.10813

Shmueli, G., Lichtendahl Jr, K. C. (2016). Practical time series forecasting with r: A hands-on guide. Axelrod Schnall Publishers.

Ülke, V., Sahin, A., Subasi, A. (2016). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. Neural Computing and Applications, 30(5), 1519–1527. https://doi.org/10.1007/s00521-016-2766-x

Xu, X., Zhang, Y. (2022). Contemporaneous causality among one hundred Chinese cities. Empirical Economics, 63(4), 2315–2329. https://doi.org/10.1007/s00181-021-02190-5

Zhang, F., Che, W., Xu, B., et al. (2013). The Research of ARMA Model in CPI Time Series. In: Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation. https://doi.org/10.2991/isccca.2013.8

Zhang, X., Yang, E. (2023). Have housing value indicators changed during COVID? Housing value prediction based on unemployment, construction spending, and housing consumer price index. International Journal of Housing Markets and Analysis, 17(1), 242–260. https://doi.org/10.1108/ijhma-01-2023-0015