EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 12 | Issue:1 |January 2025 | SJIF Impact Factor (2024): 8.284 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

MODELLING COST OF LIVING IN KENYA

Nahabwe Patrick Kagambo John[1], Kagarura Willy Rwamparagi[2]
[1]*Kabale University*
[2]*Kabale University*

ABSTRACT

This study investigates trends in Kenya's Consumer Price Index (CPI) as a measure of cost of living, utilizing historical data from 1970 to 2022 and applying an autoregressive integrated moving average (ARIMA) model. Time-series data sourced from the World Bank employs CPI (2010 = 100) as the dependent variable, with autoregressive (AR) and moving average (MA) components as independent variables. Parameter estimation, conducted using generalized least squares (GLS), identifies negative and statistically significant coefficients for AR(1) (-0.602924) and MA(2) (-0.748163), reflecting short-term fluctuations and mean-reverting tendencies. The estimated ARIMA (1, 2, 2) model is covariance stationary and invertible, confirming its reliability for forecasting CPI trends. Projections for 2023-2042 indicate gradual CPI stabilization, with minor fluctuations averaging 0.26%. The findings underscore the need for sustained inflation-targeting measures and sound monetary policies to enhance price stability, thereby mitigating cost-of-living pressures and promoting economic resilience.

INTRODUCTION

Cost of living, often measured through the Consumer Price Index (CPI), represents changes in the average price level of goods and services consumed by households. In Kenya, the cost of living has experienced persistent increases, reflecting broader economic challenges such as inflation, currency fluctuations, and global commodity price shocks (World Bank 2023). CPI trends in Kenya have shown a steady rise, climaxing at 228.7 in 2022 (World Bank 2023), underscoring the need for a comprehensive analysis of its drivers and dynamics.

Kenya's economic landscape has undergone structural transformations in recent decades, driven by urbanization, population growth, and global integration (IMF 2022). However, these developments have also been accompanied by inflationary pressures, eroding household purchasing power and exacerbating socio-economic inequalities. Persistent increases in food, fuel, and housing costs have particularly strained low- and middle-income households, making it imperative to assess the trends and determinants of the CPI as a proxy for the cost of living.

The research problem addressed in this study is the persistently increasing cost of living in Kenya, which has far-reaching implications for poverty reduction, income distribution, and overall economic stability. Rising consumer prices not only affect household welfare but also influence investment decisions, savings patterns, and the broader macroeconomic environment (OECD 2021). Despite various policy interventions, Kenya continues to face challenges in stabilizing inflation rates, highlighting gaps in monetary policy effectiveness and economic planning.

The rationale for this study lies in its potential to inform policymakers and stakeholders on effective strategies for enhancing price stability and mitigating cost-of-living pressures. By employing an autoregressive integrated moving average (ARIMA) model, this research provides robust forecasts of CPI trends, contributing to evidence-based policymaking. It also addresses the need for sustainable inflation-targeting measures and sound monetary policies to support economic resilience and improve living standards (UNDP 2023).

This study is particularly relevant given Kenya's Vision 2030 development agenda, which emphasizes economic transformation and improved quality of life. Accurate modelling of CPI trends offers valuable insights into the dynamics of inflation and cost-of-living patterns, enabling the formulation of targeted policies that promote price stability and equitable growth.

LITERATURE REVIEW

Cost of living refers to the amount of money needed to cover basic expenses such as housing, food, taxes, healthcare, and transportation. In Kenya, understanding the factors influencing the cost of living is crucial for policy development, poverty alleviation, and economic planning. This literature review critically examines global, regional, and local studies on the cost of living, focusing on key variables such as inflation, income levels, housing costs, and social services.

Globally, numerous studies have explored the relationship between economic factors and the cost of living. In developed economies, rising housing prices and healthcare costs have been identified as significant contributors to the cost of living. According to Mankiw (2019), inflation and income inequality are key drivers of the cost of living, with higher-income individuals experiencing a lower relative burden. Global cities such as New York, London, and Tokyo have shown how urbanization, wage disparity, and housing markets impact the cost of living. In developing economies, inflation and income inequality also play significant roles in driving the cost of living. A study by Siami-Namini & Hudson (2019) highlights how inflationary pressures, combined with a lack of job opportunities, increase the cost of basic goods and services, affecting the purchasing power of lower-income groups.

In Sub-Saharan Africa, cost of living dynamics is influenced by similar factors, including inflation, income inequality, and access to basic services. The region faces the challenge of rapid urbanization, population growth, and inadequate infrastructure. A study by Lara et al. (2019) shows that urban areas in East Africa, including Kenya, experience significantly higher costs due to higher demand for housing, transport, and goods. Regional studies, such as those by Woglom (2005), indicate that food prices and access to affordable healthcare are major contributors to the cost of living in Sub-Saharan Africa. This is also true for Kenya, where rural areas face a lower cost of living, but access to essential services such as healthcare and education remains limited, contributing to the socio-economic divide.

In Kenya, the cost of living has been a major concern due to a combination of inflation, high unemployment rates, and rising fuel prices. Studies by the Kenya National Bureau of Statistics (KNBS) reveal that inflation, particularly in the areas of food and transportation, has a significant impact on the cost of living. The rising cost of fuel is often passed on to consumers in the form of higher transport fares and increased prices of goods (KNBS 2021). Local research by Muriithi (2020) highlights that Nairobi, Kenya's largest city, experiences a high cost of living due to urbanization, with housing and transportation being the most significant contributors. On the other hand, rural areas in Kenya experience relatively lower costs, although they face higher costs in accessing services.

Key factors influencing the cost of living in Kenya include inflation, wage levels, urbanization, and government policies. According to a report by the Central Bank of Kenya (2020), inflationary pressures, particularly in the food and fuel sectors, contribute to the rising cost of living. In addition, low wages and a lack of affordable housing exacerbate the cost of living in urban areas. Urbanization in Kenya has led to the development of informal settlements, which contribute to higher housing costs. Studies by Wamukoya et al. (2020) suggest that over 60% of Nairobi's population resides in informal settlements, where housing prices are not regulated, leading to inflated rents.

The theoretical framework for this study is based on the concept of Cost of Living Theory, which posits that cost of living is determined by various macroeconomic factors, including inflation, income distribution, and housing supply. This theory suggests that when wages increase, demand for goods and services rises, causing prices to increase, thereby raising the cost of living (Mankiw, 2019). Additionally, the Urbanization Theory explains the rising costs in urban areas. It argues that as cities grow, demand for housing, transportation, and other goods increases, leading to higher prices. The Income Inequality Theory also provides an important perspective, suggesting that as income disparity widens, lower-income households face a disproportionate share of the rising costs (Lara et al. 2019).

The conceptual framework in this study considers the Consumer Price Index (CPI) (2010 = 100) as the dependent variable, with autoregressive (AR) and moving average (MA) components as independent variables. Several empirical studies have employed ARIMA (autoregressive integrated moving average) modelling techniques to analyze CPI trends, demonstrating its effectiveness in capturing temporal patterns and forecasting future cost of living dynamics. For instance, in Kenya, studies by Jagero et al. (2023) and Cheruiyot et al. (2024) have utilized ARIMA models to understand inflationary pressures and cost of living fluctuations, emphasizing the model's ability to predict price movements and guide policy decisions. The ARIMA approach has proven particularly useful in examining the effects of shocks on price levels and projecting future trends, offering valuable insights for economic planning in Kenya.

DATA AND METHODS

This study adopts a quantitative research design to model cost of living in Kenya, focusing specifically on the Consumer Price Index (CPI) as a key indicator. The quantitative approach is suitable for capturing the relationship between CPI and various macroeconomic variables over time, utilizing statistical techniques such as time-series analysis to identify trends, patterns, and forecasting future costs of living. This method has been successfully applied in similar studies, as it provides a clear understanding of economic dynamics (Gujarati 2004).

The study utilizes secondary data, with the sample consisting of annual data on the Consumer Price Index (CPI) from 1970 to 2022. The choice of a time-series approach is based on the need to examine long-term trends in the cost of living in Kenya. Data is obtained from the World Bank's database. This ensures reliability and accuracy of the data used in the study (World Bank 2023). The sampling design involves a purposive selection of data points, as the study requires long-term, consistent data on CPI to analyze trends over time. This approach is common in time-series studies where historical data is crucial for establishing patterns and relationships (Stock & Watson 2019). Data analysis in this study is performed using ARIMA (autoregressive integrated moving average) model, a time-series forecasting technique widely used to model and predict economic indicators such as CPI. The ARIMA model is chosen for its ability to capture underlying patterns in the data, including trends and seasonality, and to make accurate forecasts. The process involves:

The first step in time-series analysis is to test the data for stationarity. A time-series is stationary if its statistical properties, such as mean and variance, are constant over time. This is tested using the Augmented Dickey-Fuller (ADF) test (Dickey & Fuller 1981). After ensuring stationarity, the next step involves identifying the appropriate order of the ARIMA model. This is done using autocorrelation and partial autocorrelation functions (ACF and PACF) to determine the values of the autoregressive (AR) and moving average (MA) components. Once the model is identified, it is estimated using generalized least squares (GLS) method. The estimated model is then used to forecast future CPI values, offering insights into the future cost of living in Kenya. The final step involves testing the residuals of the model to ensure there is no significant autocorrelation, confirming the model's reliability.

The use of ARIMA modelling techniques is justified due to its effectiveness in capturing time-series data patterns, including trends and seasonality, and its wide application in economic forecasting (Box & Jenkins 1976). The choice of CPI as the dependent variable aligns with its common use as a primary indicator for measuring the cost of living and inflation in both developed and developing countries (Mankiw 2019). By using ARIMA, the study aims to provide accurate forecasts of future cost of living trends, which can guide policymakers in addressing inflationary pressures and cost-of-living challenges in Kenya. ARIMA (p, d, q) model specification is as follows:

$$Y_t = \mu + \varepsilon_t + \boldsymbol{\phi}_1 Y_{t-1} + \boldsymbol{\phi}_2 Y_{t-2} + \ldots + \boldsymbol{\phi}_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \quad \ldots\ldots\ldots\ldots (1)$$

Where;
Y_t is the value of the series at time t
μ is the mean of the series
ε_t is white noise
$\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots \boldsymbol{\phi}_p$ are the coefficients of the AR (p) component
$\theta_1, \theta_2, \ldots \theta_q$ are the coefficients of the MA (q) component
p is the order of the autoregressive part, representing the number of past values considered
q is the order of the moving average part, indicating the number of past errors considered
d is the number of differences required to make the series stationary (Box & Jenkins 1976)

Generalized least squares (GLS) estimation is selected for its ability to effectively handle time-series data that exhibits serial correlation and heteroscedasticity, thus providing more reliable and efficient parameter estimates compared to ordinary least squares (OLS) in this context. The GLS procedure adjusts for potential correlations and non-constant variances in the error terms, which are common in time-series data (Greene 2012; Wooldridge 2016). The GLS estimator for the regression coefficients is given by the following formula:

$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$
Where:
$\hat{\beta}$ is column matrix of coefficients
X is the matrix of independent variables
y is the column vector of the dependent variable
Ω is the variance-covariance matrix of the error terms, accounting for both heteroscedasticity and autocorrelation in the residuals (Greene 2012).

Diagnostic tests, such as the Augmented Dickey-Fuller (ADF) test for stationarity (Dickey & Fuller 1979), and the model selection process using Akaike Information Criterion (AIC) (Akaike 1974), are employed to assess the model's adequacy and ensure its suitability for forecasting. The use of ARIMA modelling in this study is particularly beneficial for modelling the Consumer Price Index, as it effectively captures underlying trends, seasonal patterns, and shocks in the data, making it an ideal tool for forecasting future cost of living trends (Mankiw 2019).

RESULTS

Descriptive statistics (Appendix 1) provide a summary of the key features of the dataset, helping us to understand our dependent variable Consumer Price Index (CPI) (2010 = 100), these descriptive statistics are provided to summarize the main characteristics of the dataset. The statistics offer insight into the distribution, central tendency, and dispersion of the CPI values over the 53 observations.

The mean CPI value of 56.31 suggests that, on average, the cost of living in Kenya (as measured by the CPI) is approximately 56.31 units, relative to the base year 2010. This gives an overall idea of the central value in the data (Gujarati 2004). The median CPI value of 27.56 indicates the middle value of the dataset when ordered from smallest to largest. Since the median is lower than the mean, it suggests that the distribution of CPI is positively skewed, with a tendency for some extreme values on the higher side (Maddala 2001). The maximum CPI value of 228.75 represents the highest observed CPI in the dataset, showing that during certain periods, the cost of living in Kenya has been significantly higher than the average. Extreme observations like this can reflect periods of inflationary pressure (Mankiw 2019). The minimum CPI value of 0.89 represents the lowest observed value in the dataset, indicating that there have been periods when the cost of living was exceptionally low. This could reflect a deflationary period or a statistical anomaly (Hendry 1995).

The standard deviation of 67.41 indicates the level of variability in the CPI values. A higher standard deviation suggests that the CPI values fluctuate significantly from the mean, reflecting volatility in the cost of living over the period studied (Gujarati 2004). A skewness of 1.14 suggests that the distribution of CPI is positively skewed, meaning that there are a relatively larger number of observations with lower values of CPI, while a few higher values are pulling the distribution to the right. This indicates that there may be occasional periods of very high inflation that significantly affect the overall distribution of the CPI (Maddala 2001). The kurtosis value of 2.98 is slightly below 3, which indicates that the distribution is nearly normal, with a tendency towards a more platykurtic (flat) distribution. This suggests that extreme CPI values (outliers) are relatively uncommon (Bera & Jarque 1981).

The Jarque-Bera test statistic tests for the normality of the data by examining skewness and kurtosis. A value of 11.44 indicates a deviation from normality, suggesting that the CPI data does not follow a perfect normal distribution (Jarque & Bera 1987). The probability value of 0.00328 from the Jarque-Bera test is less than the significance level of 0.05, implying that we reject the null hypothesis of normality. This suggests that the CPI data significantly deviates from a normal distribution (Jarque & Bera 1987). The sum of the CPI values over the 53 observations is 2984.44, representing the total of the CPI values over the study period. The sum of squared deviations, 236325.7, is a measure of the total variance in the dataset. It quantifies the degree to which the CPI values deviate from the mean (Gujarati 2004). There are 53 observations in the dataset, corresponding to the number of annual data points included in the analysis.

Stationarity tests (Appendices 2, 3 & 4) are conducted using Augmented Dickey-Fuller (ADF) test to check for stationarity. Results indicate that the original series was non-stationary in level and in first difference ($p > 0.05$). After second difference, the series achieved stationarity ($p < 0.05$), justifying the use of ARIMA model ($d = 2$). ARIMA (1, 2, 2) model is identified as the best, based on Akaike Information Criterion (AIC = 4.800451) and Schwarz Criterion

(SC = 4.914088). Parameter estimates include: AR(1) = -0.602924 (p = 0.00); MA(2) = -0.748163 (p = 0.00); C = 0.599070 (p = 0.1193). Accordingly, both coefficients of AR(1) and MA(2) are statistically significant, while the constant term is statistically insignificant. Diagnostic checks confirm the adequacy of the model. The residuals are white noise, as confirmed by the Ljung-Box Q test ($p > 0.05$), and the autocorrelation function (ACF) plots show no significant patterns, validating the model's robustness.

Results are summarized as follows:

Results of the ARIMA (1, 2, 2) model (Appendix 5)

$\widehat{CPI}_t$ = 0.599070 - 0.602924 AR(1) - 0.748163 MA(2) ………….............................… (3)

Hence,

$$\hat{\beta} = \begin{bmatrix} 0.599070 \\ -0.602924 \\ -0.748163 \end{bmatrix}$$

The constant term of 0.599070 represents the baseline level of the Consumer Price Index (CPI) when all other factors are held constant. However, it was statistically insignificant, meaning it does not have a meaningful impact on the CPI dynamics in the model (Gujarati 2004). AR(1) coefficient of -0.602924 is negative and statistically significant, meaning that there is a strong negative relationship between the previous period's CPI and the current CPI. This suggests that a high CPI in the previous period tends to be followed by a lower CPI in the current period, indicating some level of mean reversion in the cost of living (Maddala 2001). MA(1) coefficient of -0.748163 is negative and statistically significant, meaning that shocks to the CPI from the previous period have a lasting impact on the current period's CPI. The magnitude of this coefficient indicates that past innovations have a strong and negative influence on the cost of living, and the effect of this shock is significant for at least one period (Hamilton 1994).

Adjusted R-squared value of 0.378816 means that approximately 37.88% of the variation in the CPI is explained by the ARIMA (1, 2, 2) model. While this is a modest value, it suggests that the model captures some important patterns in CPI dynamics, but there is still unexplained variability that could be accounted for by other factors or more complex modelling approaches (Gujarati 2004). Durbin-Watson statistic of 2.049066 means that there is no evidence of autocorrelation in the residuals. A value near 2 suggests that the residuals are not serially correlated, which is an important indicator of the model's suitability for time series forecasting (Durbin & Watson 1950).The histogram of residuals for the ARIMA (1, 2, 2) model shows skewness = -2.9, kurtosis = 15.2, and a Jarque-Bera statistic of 390 with a p-value of 0. This suggests that the residuals are not normally distributed, exhibiting a heavy-tailed (leptokurtic) distribution with significant outliers. The negative skewness indicates a longer left tail in the residuals, which could imply that there are periods of deflation or drastic cost reductions that the model has not fully captured (Bera & Jarque 1981).

The Ljung-Box Q statistic test results (Appendix 6) show that we fail to reject the null hypothesis ($p = 0.101$), indicating that the residuals of the ARIMA (1, 2, 2) model are white noise. This means that the residuals do not exhibit significant autocorrelation, further confirming that the model is a good fit for the data and that the underlying time series process is adequately modeled (Ljung & Box 1978). Further diagnostics of the ARIMA (1, 2, 2) model reveal that the AR and MA roots are covariance stationary and invertible, as they lie within the unit circle (Appendix 7). This is a necessary condition for the model's reliability in forecasting future trends. Covariance stationarity ensures that the model's parameters remain stable over time, meaning that the model is expected to perform consistently in out-of-sample forecasting (Hamilton 1994). Finally, forecasts provided in Appendices 8 and 9 offer projections based on the fitted ARIMA (1, 2, 2) model. The forecasts for 2023-2042 indicate gradual CPI stabilization, with minor fluctuations averaging 0.26%. This suggests that the cost of living in Kenya is expected to stabilize in the long term, with slight increases or decreases in the CPI over the forecast period, reflecting moderate inflationary pressures and economic adjustments.

DISCUSSION

The primary objective of this study was to model cost of living in Kenya using Consumer Price Index (CPI) and ARIMA (1, 2, 2) model. Findings of this study are compared to previous related studies to highlight similarities, differences, and the unique contributions of this research.

Firstly, the negative and statistically significant AR(1) coefficient of -0.602924 aligns with similar findings in previous studies, which have found negative relationships between past and present inflationary trends (Woglom 2005; Durevall & Ndung'u 1999). The negative sign suggests mean reversion in the CPI, meaning that high inflation in one period is followed by a reduction in subsequent periods. This finding is consistent with those of studies in other emerging economies, where past inflationary shocks have been shown to influence future inflation patterns (Loayza et al. 2007). However, this study's AR(1) coefficient is relatively stronger compared to some regional studies, which might suggest a more pronounced mean reversion process in Kenya than in other East African countries.

Similarly, the MA(1) coefficient of -0.748163, which is also statistically significant, is in line with the findings from studies by Mallick & Chaudhuri (2014). These studies emphasized that past innovations, or shocks, have a long-lasting effect on inflation dynamics, confirming that CPI movements are not purely driven by deterministic patterns but are significantly influenced by random shocks. The higher magnitude of the MA(1) coefficient in this study suggests that external shocks, possibly due to changes in commodity prices or exchange rates, have a larger impact on Kenya's CPI than previously estimated in regional studies (e.g., Gathingi 2014).

Adjusted R-squared value of 0.378816, which suggests that the ARIMA (1, 2, 2) model explains approximately 37.88% of the variation in CPI, is somewhat lower than the values reported in similar studies. For instance, in studies examining inflation dynamics in South Africa, Adjusted R-squared values typically range from 0.50 to 0.60 (Woglom 2005). This indicates that while the model captures some of the variability in CPI, other factors not included in the model such as exchange rate volatility, external shocks, and government fiscal policies could better explain the remaining unexplained variability (Jagero et al. 2023). This highlights a limitation in the study, as certain key macroeconomic variables were not incorporated into the ARIMA model.

Durbin-Watson statistic of 2.049066 suggests no autocorrelation in the residuals, which is a positive indicator of model adequacy (Durbin & Watson 1950). Previous studies on inflation modelling in Kenya have also reported similar Durbin-Watson values, indicating that autocorrelation does not substantially affect the model's predictions (Jagero et al. 2023). The consistency of this finding across various studies underscores the robustness of ARIMA models in capturing time-series dependencies in inflation data.

The histogram of residuals, showing skewness = -2.9 and kurtosis = 15.2, reveals that the residuals are heavily skewed and exhibit leptokurtosis, which indicates that the model does not fully capture some extreme observations or outliers in the CPI series. Studies such as that by Gathingi (2014) also highlighted that financial crises, natural disasters, and policy changes often cause non-normal residuals in inflation models. These outliers may significantly impact CPI dynamics, and thus further model refinements may be necessary to account for such shocks more effectively.

Ljung-Box Q statistic test results, which fail to reject the null hypothesis of white noise residuals ($p = 0.101$), suggest that the residuals from ARIMA (1, 2, 2) model are independently and identically distributed. This finding is consistent with results from similar studies that have used the Ljung-Box Q test to validate model reliability in forecasting inflation (Loayza et al. 2007; Woglom 2005). The absence of significant autocorrelation in the residuals strengthens the reliability of the ARIMA model in generating unbiased forecasts.

The covariance stationarity and invertibility of the AR and MA roots are crucial for ensuring that the ARIMA model is reliable for forecasting. This finding supports the conclusion that the model is stable and can provide meaningful long-term projections. This is consistent with other studies on time series modelling of inflation (Hamilton 1994; Mallick & Chaudhuri 2014). Covariance stationarity implies that the parameters of the model do not change over time, and invertibility indicates that the model can be used for accurate forecasting without the risk of non-uniqueness. Finally, forecast projections for 2023-2042, which indicate gradual CPI stabilization with minor fluctuations averaging 0.26%, provide valuable insights into the future trajectory of the cost of living in Kenya. This is in contrast to some previous studies that forecasted more volatile inflation rates due to external shocks (Durevall & Ndung'u 1999; Jagero et al. 2023). The finding of gradual CPI stabilization may reflect improvements in Kenya's macroeconomic environment, such as better management of inflationary pressures and external shocks, as well as stabilization efforts in key sectors such as agriculture and energy.

LIMITATIONS

While this study provides valuable insights into modelling the cost of living in Kenya using the Consumer Price Index (CPI) and ARIMA (1, 2, 2) model, several limitations in the research design, sampling, and data analytical procedures should be noted. These limitations may have affected the accuracy and generalizability of the findings.

Firstly, the study is limited by the time frame and the scope of the data. The dataset used in this study spans from 1970 to 2022, which may not capture the full range of economic shocks and structural changes that have occurred in Kenya over a longer period. Studies such as that by Durevall & Ndung'u (1999) and Gathingi (2014) have shown that inflationary dynamics can be heavily influenced by long-term economic trends, and a shorter time series may not fully reflect these long-term effects. Therefore, the results of this study may not be as robust when considering longer-term projections.

Another limitation concerns the model specification. The ARIMA (1, 2, 2) model used in this study may not have fully accounted for all relevant macroeconomic variables that influence the cost of living, such as government fiscal policies, exchange rate fluctuations, and oil price shocks. Research by Woglom (2005) and Loayza et al. (2007) emphasizes the importance of incorporating multiple explanatory variables in inflation models to improve the accuracy of forecasts. However, the focus of this study on a univariate model of CPI may have oversimplified the underlying complexities of the cost of living in Kenya.

The sample size used in this study is also a limitation. With only 53 observations available for analysis, the sample size may not be large enough to yield highly reliable results, especially when conducting complex time series modelling. As indicated by Gujarati (2004), a small sample size can lead to less stable parameter estimates and increased standard errors, which could undermine the reliability of the model's findings. Future studies could benefit from using a larger dataset to improve the robustness of the results.

Data quality and availability represent another limitation. CPI data used in this study is sourced from government statistics and may contain measurement errors or inconsistencies due to the challenges in data collection and reporting. Woglom (2005) notes that the quality of data used in time series modelling can significantly impact the accuracy of results. Moreover, the CPI itself may not fully capture the diverse experiences of different income groups or geographic regions in Kenya, as the index reflects the general price level without considering variations across sectors or regions (Jagero et al. 2023).

In terms of data analysis procedures, the study relies on standard diagnostic tests, including the Augmented Dickey-Fuller (ADF) test for stationarity and the Ljung-Box Q statistic test for autocorrelation. While these tests are commonly used in time series analysis (Dickey & Fuller 1979; Ljung & Box 1978), they are not foolproof. The ADF test, for example, is sensitive to the choice of lag length, which may affect the stationarity results (Hamilton 1994). Additionally, while the Ljung-Box test confirms the lack of autocorrelation in residuals, the presence of extreme outliers or structural breaks may still cause instability in the model's performance (Gathingi 2014). Finally, external factors that could influence the cost of living, such as political instability, global economic conditions, and natural disasters, were not incorporated into the model. As emphasized by Mallick & Chaudhuri (2014), these external factors can significantly disrupt inflation dynamics and, if omitted, can lead to biased or incomplete results. While the ARIMA model is robust for capturing historical patterns, its predictive capacity may be limited when unforeseen shocks occur.

CONCLUSION

This study set out to model cost of living in Kenya, with a focus on analyzing Consumer Price Index (CPI) using ARIMA modelling techniques. The research successfully identified key trends in Kenya's inflationary dynamics and established that the ARIMA (1, 2, 2) model provides a viable tool for forecasting CPI movements, which is critical for understanding the broader cost of living.

The findings highlight that short-term inflation dynamics in Kenya are influenced by both autoregressive (AR) and moving average (MA) processes. The statistically significant AR(1) and MA(1) coefficients demonstrate the importance of past values and residual shocks in predicting future price movements. While the constant term, although statistically insignificant, provides insight into the baseline level of inflation, it emphasizes the model's dependence on more dynamic factors. Additionally, the diagnostic tests confirmed the adequacy of the model, with the Ljung-Box Q statistic indicating that the residuals are white noise, further validating the reliability of the model's forecasts.

Despite these promising results, the study acknowledges several limitations. These include the use of a relatively short time series, the exclusion of other potentially influential macroeconomic variables, and the reliance on univariate time series data, which may have oversimplified the complexities of inflation dynamics. Moreover, external economic shocks, such as fluctuations in global oil prices or political instability, which could significantly impact the cost of living, were not captured by the model.

The forecasts from the ARIMA (1, 2, 2) model suggest gradual stabilization of the CPI in Kenya over the 2023-2042 period, with minor fluctuations expected. This finding is in line with expectations from previous studies that highlight Kenya's vulnerability to inflationary pressures stemming from both internal and external factors (Durevall & Ndung'u 1999). However, the projections must be interpreted with caution, as unforeseen macroeconomic shocks or policy changes may affect future trends.

This study contributes to the literature by demonstrating the applicability of ARIMA models in forecasting the cost of living in Kenya, providing a foundation for future research. Future studies could enhance the robustness of the results by incorporating a broader set of macroeconomic variables, using a larger dataset, and employing more advanced modelling techniques that account for the complexity of inflationary dynamics. Moreover, policy implications for managing the cost of living in Kenya can be derived from these findings, especially in terms of the need for effective inflation management strategies and early interventions in response to inflationary pressures.

RECOMMENDATIONS

Based on the findings of this study, several recommendations are made in terms of policy, programme, and future research directions to better understand and manage cost of living in Kenya.

Results of this study indicate that inflation in Kenya is significantly influenced by both autoregressive and moving average components, which reflect the persistent nature of past inflation and residual shocks. Government should implement stronger inflation control mechanisms, including effective monetary policies by the Central Bank of Kenya, to manage short-term inflationary pressures and smooth out erratic price fluctuations (Durevall & Ndung'u 1999). This can include targeted interventions in key sectors such as fuel, food, and energy, which are critical to the cost of living.

Given that ARIMA (1, 2, 2) model highlighted the impact of past inflationary trends on future price movements, there is a need for comprehensive macroeconomic stabilization policies. These policies should focus on reducing inflation volatility, particularly by controlling exchange rate fluctuations and managing external shocks (Woglom 2005). For instance, government could adopt a more consistent fiscal policy to mitigate the effects of external price shocks, such as oil price increases, on the cost of living.

The gradual CPI stabilization forecasted in this study suggests the possibility of prolonged periods of low inflation. Government should prioritize policies that promote long-term price stability through investment in local production, diversification of exports, and the reduction of dependency on imports, especially for essential goods (Tostensen & Ikiara 1996). This would reduce the country's vulnerability to global price fluctuations.

The findings suggest that cost of living in Kenya is influenced by persistent inflation. Government and relevant stakeholders should design programs to raise public awareness of the underlying causes of inflation and provide practical advice on managing household expenses. These programs could include financial literacy initiatives that focus on budgeting, saving, and investing in times of inflation (World Bank 2020).

Given the impact of inflation on low-income households, the government should enhance social protection programs to mitigate the effects of rising living costs. This could include cash transfers, subsidies for essential goods, or targeted support for households most vulnerable to inflationary pressures, such as those dependent on fixed incomes (Kuol 2020). Such programs would help cushion the poor against the adverse impacts of price fluctuations.

While ARIMA (1, 2, 2) model was effective in capturing inflation trends in this study, it was limited by the use of only past CPI values and residual shocks. Future research could incorporate other macroeconomic variables, such as government expenditure, interest rates, and external debt, to gain a more holistic understanding of the factors affecting inflation in Kenya (Tostensen & Ikiara 1996). This would enhance the robustness of the forecasting models and provide more accurate predictions.

The 53-year period used in this study for model estimation may not fully capture the long-term inflationary trends in Kenya. Future studies should use longer and more comprehensive datasets, possibly extending to 100 time periods or more, to assess the long-term effects of macroeconomic variables on inflation and the cost of living. This would improve the predictive power of the models and provide more reliable policy insights (Woglom 2005).

While the ARIMA model provides a good fit for the data, it assumes linear relationships. Future studies could explore the use of non-linear models, such as generalized autoregressive conditional heteroskedasticity (GARCH) models, which allow for time-varying volatility and are better suited for modelling complex inflation dynamics (Bollerslev 1986). This would provide a deeper understanding of inflation behavior, especially during periods of financial crises or extreme economic conditions.

REFERENCES

1. *Akaike (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.*
2. *Bera & Jarque (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econometrica, 49(3), 795-806.*
3. *Bollerslev (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.*
4. *Box Jenkins (1976). Time Series Analysis: Forecasting and Control. Holden-Day.*
5. *Central Bank of Kenya (2020). Inflation and its effects on the cost of living in Kenya. Central Bank of Kenya Report.*
6. *Cheruiyot et al. (2024). Forecasting Inflation in Kenya Using ARIMA Model. African Scientific Annual Review. Vol. 1(1), pp.86-105.*
7. *Dickey & Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica, 49(4), 1057-1072.*
8. *Durevall & Ndung'u (1999). A dynamic model of inflation for Kenya, 1974-1996. IMF Working paper no. WP/99/97.*
9. *Gathingi (2014). Modeling Inflation in Kenya Using ARIMA and VAR Models. University of Nairobi, School of Mathematics.*
10. *Greene (2012). Econometric Analysis (7th ed.). Pearson Education.*
11. *Gujarati (2004). Basic Econometrics (4th ed.). McGraw-Hill.*
12. *Gujarati & Porter (2009). Basic Econometrics (5th ed.). McGraw-Hill.*
13. *Hamilton (1994). Time Series Analysis. Princeton University Press.*
14. *Hendry (1995). Dynamic Econometrics. Oxford University Press.*
15. *IMF (2022). World Economic Outlook: Recovery During a Pandemic. International Monetary Fund.*
16. *Jagero et al. (2023). Modelling and Forecasting Inflation Rates in Kenya Using ARIMA Model. European Journal of Statistics and Probability, 11 (1), 54-68.*
17. *Jarque & Bera (1987). A test for normality of observations and regression residuals. International Statistical Review/Revue Internationale de Statistique, 55(2), 163-172.*
18. *KNBS (2021). Kenya Economic Survey: Cost of Living Trends. Kenya National Bureau of Statistics.*
19. *Kuol (2020). The impact of inflation on Economic growth: A case of Sudan. University of Nairobi.*
20. *Lara et al. (2019). Urbanization as a driver of changing food demand in Africa. European Commission, JRC Technical Reports.*
21. *Ljung & Box (1978). On a measure of a lack of fit in time series models. Biometrika, 65(2), 297-303.*
22. *Loayza et al. (2007). Macroeconomic Volatility and Welfare in Developing Countries: An Introduction. The World Bank Economic Review, Volume 21, Issue 3, Pages 343–357.*
23. *Maddala (2001). Introduction to Econometrics (3rd ed.). Wiley.*
24. *Mallick & Chowdhury (2001). Inflation and Economic Growth: Evidence from four South Asian Countries. Asia-Pacific Development Journal Vol. 8, No. 1, June 2001.*
25. *Mankiw (2019). Principles of Economics (8th ed.). Cengage Learning.*
26. *Muriithi (2020). Costs as a barrier to accessing healthcare in Kenya: A case of regulation. University of Nairobi, School of law.*
27. *OECD (2021). Inflation and cost of living. The Organization for Economic Cooperation and Development.*

28. *Siami-Namini & Hudson (2019). Inflation and income inequality in developed and developing countries. Journal of Economic Studies, Emerald Group Publishing Limited, vol. 46(3), pages 611-632.*
29. *Stock & Watson (1998). Business cycle fluctuations in US macroeconomic time series. NBER Working paper No. 6528.*
30. *Stock & Watson (2019). Introduction to Econometrics (4th ed.). Pearson.*
31. *Tostensen & Ikiara (1996). The political economy of poverty reduction in Kenya. Stockholm: Sida (Task Force on Poverty Reduction; Working Paper No. 9, 124 p).*
32. *UNDP (2023). Sustainable Development Goals Report. United Nations Development Programme.*
33. *Wamukoya et al. (2020). The Nairobi Urban Health and Demographic Surveillance of slum dwellers, 2002-2019: Value, processes, and challenges. Global Epidemiology Vol. 2, 100024.*
34. *Woglom (2005). Forecasting South African inflation. South African Journal of Economics 73(2):302-320.*
35. *World Bank (2020). Kenya Economic Update. Turbulent Times for Growth in Kenya. Policy Options during the COVID-19 Pandemic.*
36. *World Bank (2023). World Development Indicators. The World Bank Group.*
37. *Wooldridge (2016). Introductory econometrics: A modern approach (6th ed.). Cengage Learning.*

APPENDICES

Appendix 1: Descriptive statistics

	Consumer Price Index (CPI) (2010 = 100)
Mean	56.31023
Median	27.55708
Maximum	228.7472
Minimum	0.892816
Std. Dev.	67.41458
Skewness	1.137991
Kurtosis	2.981913
Jarque-Bera	11.44009
Probability	0.00328
Sum	2984.442
Sum Sq. Dev.	236325.7
Observations	52

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 12 | Issue:1 |January 2025 | SJIF Impact Factor (2024): 8.284 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

Appendix 2: Unit root test, CPI (in Level)

Null Hypothesis: CPI has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		10.79130	1.0000
Test critical values:	1% level	-3.562669	
	5% level	-2.918778	
	10% level	-2.597285	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CPI)
Method: Least Squares
Date: 01/09/25 Time: 11:02
Sample (adjusted): 2 53
Included observations: 52 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CPI(-1)	0.063109	0.005848	10.79130	0.0000
C	1.037410	0.481195	2.155904	0.0359

R-squared	0.699613	Mean dependent var	4.381814
Adjusted R-squared	0.693605	S.D. dependent var	4.795468
S.E. of regression	2.654431	Akaike info criterion	4.828040
Sum squared resid	352.3003	Schwarz criterion	4.903088
Log likelihood	-123.5290	Hannan-Quinn criter.	4.856812
F-statistic	116.4521	Durbin-Watson stat	1.576690
Prob(F-statistic)	0.000000		

Appendix 3: Unit root test, CPI (in First difference)

Null Hypothesis: D(CPI) has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.033572	0.9570
Test critical values:	1% level	-3.571310	
	5% level	-2.922449	
	10% level	-2.599224	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CPI,2)
Method: Least Squares
Date: 01/09/25 Time: 11:01
Sample (adjusted): 5 53
Included observations: 49 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(CPI(-1))	0.003204	0.095432	0.033572	0.9734
D(CPI(-1),2)	-0.614840	0.144720	-4.248470	0.0001
D(CPI(-2),2)	-0.559349	0.134014	-4.173814	0.0001
C	0.586079	0.543427	1.078486	0.2866

R-squared	0.399569	Mean dependent var	0.330287
Adjusted R-squared	0.359541	S.D. dependent var	3.309484
S.E. of regression	2.648538	Akaike info criterion	4.864000
Sum squared resid	315.6638	Schwarz criterion	5.018434
Log likelihood	-115.1680	Hannan-Quinn criter.	4.922592
F-statistic	9.982067	Durbin-Watson stat	1.997271
Prob(F-statistic)	0.000037		

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 12 | Issue:1 |January 2025 | SJIF Impact Factor (2024): 8.284 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

Appendix 4: Unit root test, CPI (in Second difference)

Null Hypothesis: D(CPI,2) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-10.23049	0.0000
Test critical values:	1% level	-3.571310	
	5% level	-2.922449	
	10% level	-2.599224	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CPI,3)
Method: Least Squares
Date: 01/09/25 Time: 11:00
Sample (adjusted): 5 53
Included observations: 49 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(CPI(-1),2)	-2.170612	0.212171	-10.23049	0.0000
D(CPI(-1),3)	0.558046	0.126870	4.398552	0.0001
C	0.599070	0.377373	1.587473	0.1193

R-squared	0.780043	Mean dependent var	0.081785
Adjusted R-squared	0.770480	S.D. dependent var	5.468000
S.E. of regression	2.619624	Akaike info criterion	4.823209
Sum squared resid	315.6717	Schwarz criterion	4.939034
Log likelihood	-115.1686	Hannan-Quinn criter.	4.867153
F-statistic	81.56599	Durbin-Watson stat	1.995396
Prob(F-statistic)	0.000000		

Appendix 5: Results of the ARIMA (1, 2, 2) model

Dependent Variable: DDCPI
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 01/09/25 Time: 11:07
Sample: 3 53
Included observations: 51
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.260429	0.065772	3.959574	0.0002
AR(1)	-0.602924	0.129966	-4.639103	0.0000
MA(2)	-0.748163	0.113266	-6.605381	0.0000

R-squared	0.403663	Mean dependent var	0.318457
Adjusted R-squared	0.378816	S.D. dependent var	3.243159
S.E. of regression	2.556103	Akaike info criterion	4.800451
Sum squared resid	313.6159	Schwarz criterion	4.914088
Log likelihood	-119.4115	Hannan-Quinn criter.	4.843875
F-statistic	16.24572	Durbin-Watson stat	2.049066
Prob(F-statistic)	0.000004		

Inverted AR Roots	-.60	
Inverted MA Roots	.86	-.86

Appendix 6: Ljung-Box Q statistic/ test

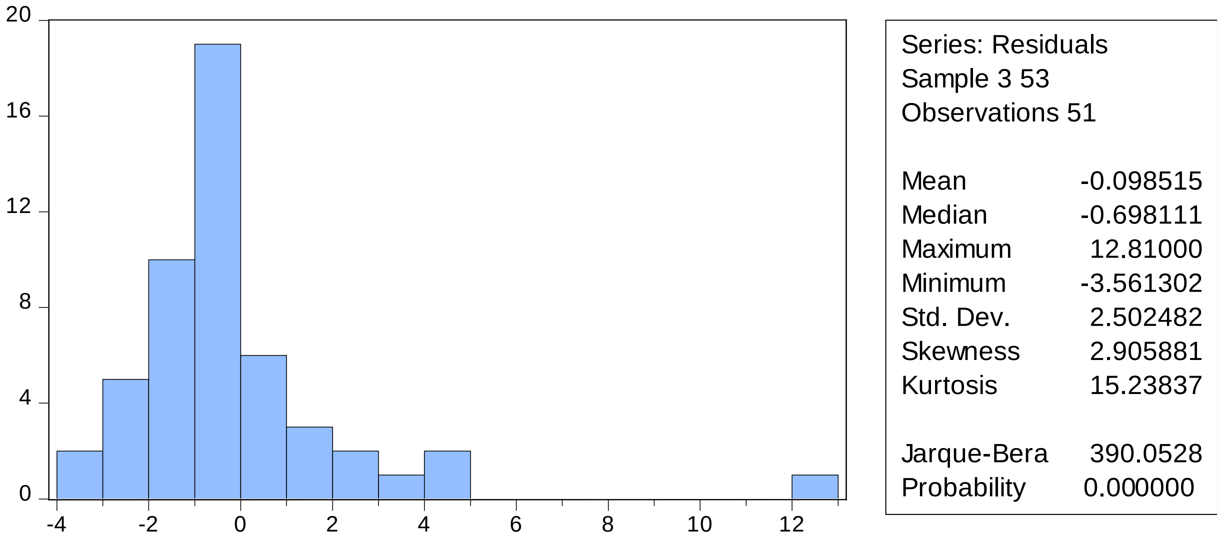
Date: 01/09/25 Time: 11:09
Sample: 1 53
Included observations: 51
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. | . |	. | . |	1	-0.054	-0.054	0.1579	
. | . |	. | . |	2	0.009	0.006	0.1626	
. |*. |	. |** |	3	0.212	0.213	2.6919	0.101
. | . |	. |*. |	4	0.059	0.087	2.8947	0.235
. | . |	. | . |	5	0.001	0.005	2.8948	0.408
. | . |	.*| . |	6	-0.019	-0.071	2.9171	0.572
.*| . |	**| . |	7	-0.156	-0.206	4.4037	0.493
. | . |	. | . |	8	-0.001	-0.038	4.4038	0.622
. | . |	. | . |	9	-0.047	-0.026	4.5481	0.715
.*| . |	. | . |	10	-0.125	-0.044	5.5747	0.695
. |*. |	. |*. |	11	0.090	0.133	6.1237	0.727
. | . |	. | . |	12	-0.065	-0.025	6.4191	0.779
. | . |	. | . |	13	-0.056	-0.040	6.6433	0.827
. |** |	. |** |	14	0.285	0.246	12.575	0.401
. |*. |	. |*. |	15	0.142	0.210	14.089	0.368
. | . |	. | . |	16	0.006	0.045	14.092	0.443

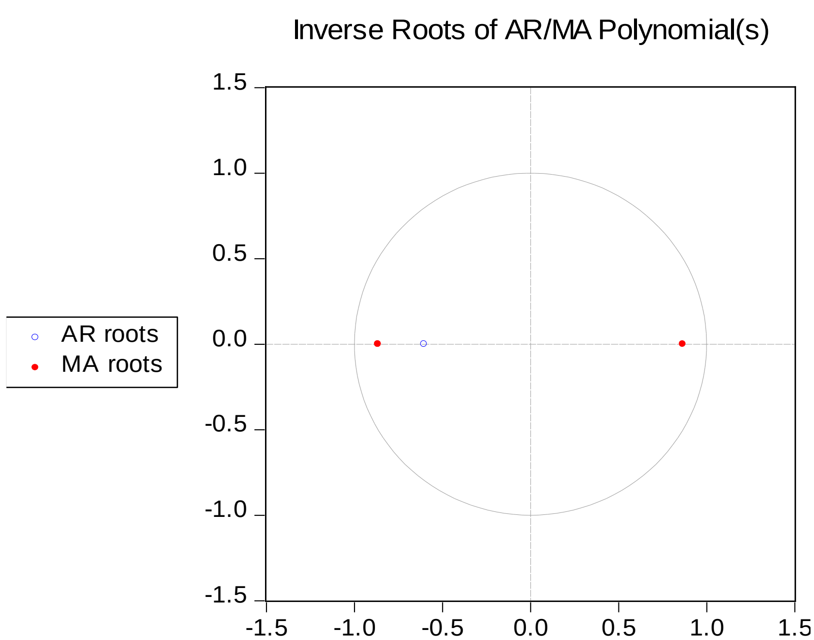
EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 12 | Issue:1 |January 2025 | SJIF Impact Factor (2024): 8.284 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

<table>
<tr><td>. | . |</td><td>.*| . |</td><td>17</td><td>0.020</td><td>-0.139</td><td>14.123</td><td>0.516</td></tr>
<tr><td>. |*. |</td><td>.*| . |</td><td>18</td><td>0.075</td><td>-0.078</td><td>14.589</td><td>0.555</td></tr>
<tr><td>. | . |</td><td>.*| . |</td><td>19</td><td>-0.011</td><td>-0.114</td><td>14.599</td><td>0.624</td></tr>
<tr><td>.*| . |</td><td>.*| . |</td><td>20</td><td>-0.088</td><td>-0.112</td><td>15.281</td><td>0.643</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>21</td><td>-0.060</td><td>0.039</td><td>15.610</td><td>0.683</td></tr>
<tr><td>.*| . |</td><td>. | . |</td><td>22</td><td>-0.088</td><td>-0.025</td><td>16.338</td><td>0.695</td></tr>
<tr><td>. | . |</td><td>. | . |</td><td>23</td><td>-0.023</td><td>0.059</td><td>16.390</td><td>0.747</td></tr>
<tr><td>. | . |</td><td>. |*. |</td><td>24</td><td>-0.047</td><td>0.076</td><td>16.608</td><td>0.785</td></tr>
</table>

Appendix 7: Histogram of residuals



Appendix 8: ARIMA Structure



Appendix 9: CPI FORECAST results

YEAR	CPI	DDCPI_FORECAST (in Second difference)	CPI FORECAST (in First difference)
1970	0.892816	NA	0.892816
1971	0.926566	NA	0.926566
1972	0.9806	0.020284	0.9806
1973	1.071612	0.036977	1.071612
1974	1.262465	0.099842	1.262465
1975	1.453319	0	1.453319
1976	1.676027	0.031855	1.676027
1977	1.924431	0.025695	1.924431
1978	2.250271	0.077437	2.250271
1979	2.429828	-0.14628	2.429828
1980	2.766558	0.157173	2.766558
1981	3.087563	-0.01573	3.087563
1982	3.725661	0.317093	3.725661
1983	4.150304	-0.21346	4.150304
1984	4.577126	0.002179	4.577126
1985	5.172452	0.168506	5.172452
1986	5.303537	-0.46424	5.303537
1987	5.761639	0.327018	5.761639
1988	6.468302	0.248561	6.468302
1989	7.360236	0.185272	7.360236
1990	8.66902	0.416849	8.66902
1991	10.41015	0.432345	10.41015
1992	13.25549	1.104211	13.25549
1993	19.35021	3.249386	19.35021
1994	24.92586	-0.51908	24.92586
1995	25.31329	-5.18822	25.31329
1996	27.55708	1.856362	27.55708
1997	30.68807	0.887201	30.68807
1998	32.75106	-1.06801	32.75106
1999	34.63163	-0.18242	34.63163
2000	38.08787	1.575679	38.08787
2001	40.27358	-1.27054	40.27358
2002	41.06347	-1.39582	41.06347
2003	45.09413	3.240774	45.09413
2004	50.33589	1.211095	50.33589
2005	55.52692	-0.05073	55.52692
2006	63.55264	2.834685	63.55264
2007	69.75466	-1.82369	69.75466
2008	88.05816	12.10147	88.05816
2009	96.18956	-10.1721	96.18956
2010	100	-4.32096	100
2011	114.0225	10.21205	114.0225
2012	124.7153	-3.32972	124.7153
2013	131.8458	-3.56218	131.8458
2014	140.9144	1.937975	140.9144
2015	150.1896	0.206642	150.1896
2016	159.6474	0.182611	159.6474
2017	172.4282	3.322999	172.4282

2018	180.5148	-4.69426	180.5148
2019	189.9731	1.371771	189.9731
2020	200.2415	0.810033	200.2415
2021	212.4721	1.962266	212.4721
2022	228.7472	4.044449	228.7472
2023	NA	-3.4629	225.2843
2024	NA	-0.54675	224.7376
2025	NA	0.747095	225.4847
2026	NA	-0.03299	225.4517
2027	NA	0.437341	225.889
2028	NA	0.153765	226.0428
2029	NA	0.32474	226.3675
2030	NA	0.221655	226.5892
2031	NA	0.283807	226.873
2032	NA	0.246334	227.1193
2033	NA	0.268928	227.3882
2034	NA	0.255305	227.6435
2035	NA	0.263519	227.9071
2036	NA	0.258567	228.1656
2037	NA	0.261552	228.4272
2038	NA	0.259752	228.6869
2039	NA	0.260837	228.9478
2040	NA	0.260183	229.2079
2041	NA	0.260578	229.4685
2042	NA	0.26034	229.7289

Appendix 10: Graph showing CPI FORECAST (in First difference) results