Automated Forecasting Dashboards – The Case of CPI

Jan-Hendrik Pretorius[a], Nico Katzke[b]

[a] *Stellenbosch University, Stellenbosch, South Africa*

[b] *SATRIX, Cape Town, South Africa*

Abstract

This paper addresses the challenge of forecasting volatile financial indicators by introducing an automated forecasting dashboard, InCast (Inflation Forecaster), tailored to enhance the precision and reliability of Consumer Price Index (CPI) forecasts in South Africa. The dashboard employs an econometric forecasting methodology to project future values of the Consumer Price Index (CPI) for South Africa. The approach leverages the power of two distinct models: an ARIMA (AutoRegressive Integrated Moving Average) model, adept at capturing underlying trends and patterns, and a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, which excels in modeling volatility and uncertainty. The ARIMA model is utilized to generate point forecasts for CPI values, while the GARCH model is employed to forecast volatility, allowing one to gauge the level of uncertainty associated with the CPI projections. Through this synthesis process, the strengths of both models are combined, producing a comprehensive forecasting framework.

Keywords: GARCH, ARIMA, CPI, Volatility Modelling, R Shiny, Dashboards, Inflation

JEL classification L250, L100

Table of Contents

Corresponding author: Jan-Hendrik Pretorius

Email addresses: janhpret@gmail.com (Jan-Hendrik Pretorius), nfkatzke@gmail.com (Nico Katzke)

Contributions:

The authors would like to thank Codera Analytics for access to their data through EconData. Thank you sincerely.

1. Introduction

In recent years, South Africa has witnessed significant volatility in its Consumer Price Index (CPI). The CPI is not only a fundamental measure for assessing inflation and economic health but also plays a pivotal role in finance, influencing monetary policy, investment strategies, and financial planning. In the realm of finance, the CPI directly impacts interest rate decisions, affects the value of currency, and serves as a critical gauge for adjusting the real returns on investments. As such, volatility in the CPI can introduce substantial uncertainty into financial markets, complicating the assessment of investment risks and opportunities.

This volatility poses a challenge for accurate forecasting and necessitates innovative approaches capable of navigating the complexity and unpredictability of financial indicators. Traditional forecasting models often struggle to separate the underlying signal from the noise inherent in such volatile economic data, leading to forecasts that may be less reliable for economic planning, policy-making, and financial analysis. Recognizing this challenge, this paper introduces an Automated Forecasting Dashboard – InCast (**In**flation Fore**cast**er) – designed specifically to enhance the precision and reliability of CPI forecasts in South Africa. By employing advanced statistical models, including Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, the dashboard aims to refine our understanding of CPI trends and their implications for the financial sector.

This paper details the development and implementation of the forecasting model and its integration into an automated dashboard, emphasizing its potential to significantly impact financial decision-making in South Africa.

2. Literature

In the literature review exploring the optimal modeling procedure for capturing volatility in CPI rates, two significant methodologies are highlighted: ARIMA/ARFIMA and GARCH modeling. The ARIMA model excels at capturing the linear aspects of time series data, making it particularly effective for modeling and forecasting CPI rates over short to medium terms. On the other hand, the GARCH model is adept at capturing and modeling the volatility clustering often observed in financial time series, including CPI rates. This ability allows GARCH models to provide valuable insights into the variability and risk associated with inflation rates, offering a complementary perspective to the ARIMA/ARFIMA models' focus on the level and trend of the series. Together, these methodologies provide a comprehensive toolkit for analyzing CPI volatility, with each addressing different facets of the data's behavior.

A working paper by Nyoni (2019) utilized ARIMA models to forecast CPI in Myanmar using annual data from 1960 to 2017. The ARIMA (2, 2, 1) model was identified as optimal for predicting future CPI values (Nyoni, 2019: 4), indicating an upward trajectory in Myanmar's inflation over the next decade. This research emphasizes the model's stability and acceptability for CPI modeling, advocating for policymakers to adopt stringent monetary and fiscal policies to manage inflation effectively.

Boateng, Gil-Alana, 'Maseka, Siweya & Belete (2016) explore the long memory dynamics of CPI inflation rates in Ghana using ARFIMA (AutoRegressive Fractionally Integrated Moving Average) modeling. This methodology effectively captures both short- and long-term dependencies within the inflation data, indicating persistence and mean-reversion in inflation rates (Boateng *et al.*, 2016: 297). Their findings underscore the significance of employing fractional differencing to accurately model inflationary trends, offering vital insights for policymakers. This approach not only reveals the presence of long memory in Ghana's inflation but also aids in better understanding and forecasting inflationary pressures (Boateng *et al.*, 2016: 299), proving ARFIMA's effectiveness in economic time series analysis.

In his seminal work, Bollerslev (2023) introduced the GARCH model, extending the ARCH model developed by Engle (1982). Bollerslev's GARCH model allows for both autoregressive and moving average components in the variance equation, enabling a more comprehensive analysis of time-series volatility (see p.25). This model has become fundamental in financial econometrics for analyzing and forecasting time-varying volatility, reflecting its capacity to capture the clustering of volatility phenomena observed in financial market returns.

Through the integration of ARIMA and GARCH methodologies, researchers have adeptly navigated the complexities of CPI rate volatility, achieving precise volatility modeling and signal extraction. This blend harnesses ARIMA's strength in understanding data trends and GARCH's prowess in volatility dynamics, offering a robust framework for accurate economic forecasting.

Baillie, Chung & Tieslau (1996) investigates inflation through the ARFIMA-GARCH model, highlighting its effectiveness in capturing long-memory processes and conditional heteroscedasticity in inflation data. Their methodology, combining ARFIMA to model fractional integration and GARCH(1, 1) for volatility, provides a nuanced understanding of inflation's persistence and variability. This approach is particularly adept at analyzing the complex dynamics of inflation rates, offering significant insights into their mean-reverting behavior and the interaction between mean and volatility, in line with Friedman's hypothesis that current inflation significantly influences volatility in future inflation (Baillie *et al.*, 1996: 38).

In an extension of the ARFIMA-GARCH model, Belkhouja & Mootamri (2016) account for time-varying baseline mean and volatility in analyzing G7 inflation dynamics from 1955 to 2014. Their innovative approach, incorporating logistic functions for structural changes (see p.451), addresses the overestimation of long-run and GARCH persistence when such changes are ignored. This model provides a nuanced understanding of inflation's behavior, aligning identified shifts with significant economic and political events, thus offering a more accurate tool for policymakers.

The ARIMA model is adept at capturing and forecasting the underlying patterns in CPI data, while the GARCH model addresses the volatility of residuals from the ARIMA model, a critical aspect given the data's inherent noise and volatility. This sophisticated modeling approach, combined with the dashboard's automation features—such as automated data updating—ensures that forecasts are based on the most current data available, a vital requirement for making informed financial decisions. The paper now turns to the methodology followed to build an automated forecasting dashboard.

3. Data & Methodology

For the purposes of this essay, the focus will fall on using an ARIMA-GARCH model[1] to extract the signal from Consumer Price Index (CPI) data in South Africa. In this section, I will provide an overview of the data sources, the key variables involved, and the methodology employed for the analysis. This approach aims to uncover valuable insights into the behavior and volatility of CPI, a critical economic indicator, in the South African context.

3.1. Automated Data Processes

To conduct this analysis, I leverage an extensive dataset, produced by Statistics South Africa, comprising the Classification of Individual Consumption by Purpose (COICOP) 5-digit CPI values for South Africa. This dataset is sourced directly from Codera Analytics's EconData platform (Statistics South Africa, 2024). It spans the time frame from January 2008 to December 2023.

One distinct advantage of utilizing EconData is its capacity for functional coding, which allows us to access the data seamlessly. This functional coding approach renders the data entirely self-contained, eliminating the necessity for separate data storage. Furthermore, it offers the benefit of automated data updates on a monthly basis. Thus, each time the dashboard (which I will elaborate on later) is accessed, it provides the most up-to-date CPI information available. This real-time data accessibility is instrumental in ensuring the accuracy and timeliness of our analysis.

The central focus of this paper revolves around the South African CPI for all items. However, it is worth noting that I also explore additional facets of the CPI, as elaborated in Table 3.1[2]. These supplementary CPI components are considered in the context of forecasting, and their analysis is

[1] While the ARFIMA model is clearly more adept at handling inflation forecasting, I focus on ARIMA due to its simplistic nature. Future analyses might benefit from the ARFIMA methodology.
[2] Here I present the variable descriptions and the identifier linked to each variable on the EconData platform

facilitated through the utilization of the InCast web application, which serves as a complement to this paper.

Table 3.1: CPI Code Lists as Supplied by EconData

"CPI - All items"	00.0.0.0.TC
"CPI - Food and non alcoholic beverages"	01.0.0.0.TC
"CPI - Alcoholic beverages and tobacco"	02.0.0.0.TC
"CPI - Clothing and footwear"	03.0.0.0.TC
"CPI - Housing water electricity gas and other fuels"	04.0.0.0.TC
"CPI - Furnishings household equipment and routine household maintenance"	05.0.0.0.TC
"CPI - Health"	06.0.0.0.TC
"CPI - Transport"	07.0.0.0.TC
"CPI - Communication"	08.0.0.0.TC
"CPI - Recreation and culture"	09.0.0.0.TC
"CPI - Education"	10.0.0.0.TC
"CPI - Restaurants and hotels"	11.0.0.0.TC
"CPI - Miscellaneous goods and services"	12.0.0.0.TC

3.2. ARIMA and GARCH Implementation

The analysis employs ARIMA and GARCH models to study the CPI in South Africa. ARIMA models capture temporal dependencies and trends, while GARCH models are often used for analyzing volatility. This combined methodology, detailed in the Mathematical Appendix, involves fitting an ARIMA model to the CPI series, extracting residuals, and then applying a GARCH model to these residuals for volatility estimates. The integration of these models aims to enhance forecasting accuracy.

3.3. Automated Dashboard Design

The automated dashboard – InCast; a supplement to the paper – was produced using R Shiny and focuses on interactive and dynamic data visualization. It offers real-time insights into CPI in South Africa[3]. The dashboard integrates the analytical models presented in this paper to provide a comprehensive view of CPI behavior and inflation dynamics – allowing one to also see the model's performance on other CPI series. This approach enables users to interact with the data, adjust certain parameters, and visualize forecasts, enhancing decision-making processes with up-to-date, customized information.

[3]While still in its infancy stage, currently only focusing on CPI data, the future of the dashboard includes plans to extend automated modelling and forecasting for other economic series. The end goal is to produce an accessible and automated modelling & forecasting tool.

4. Results

This section details the application of ARIMA and GARCH models to CPI data, focusing on the accuracy of their predictive capabilities. Figure 4.1 illustrates the overall CPI against its various components. The CPI trend is observed to be relatively linear until an increase in mid-2021, signaling a notable rise in CPI volatility – a trend that reflects the recent unpredictable nature of inflation rates.

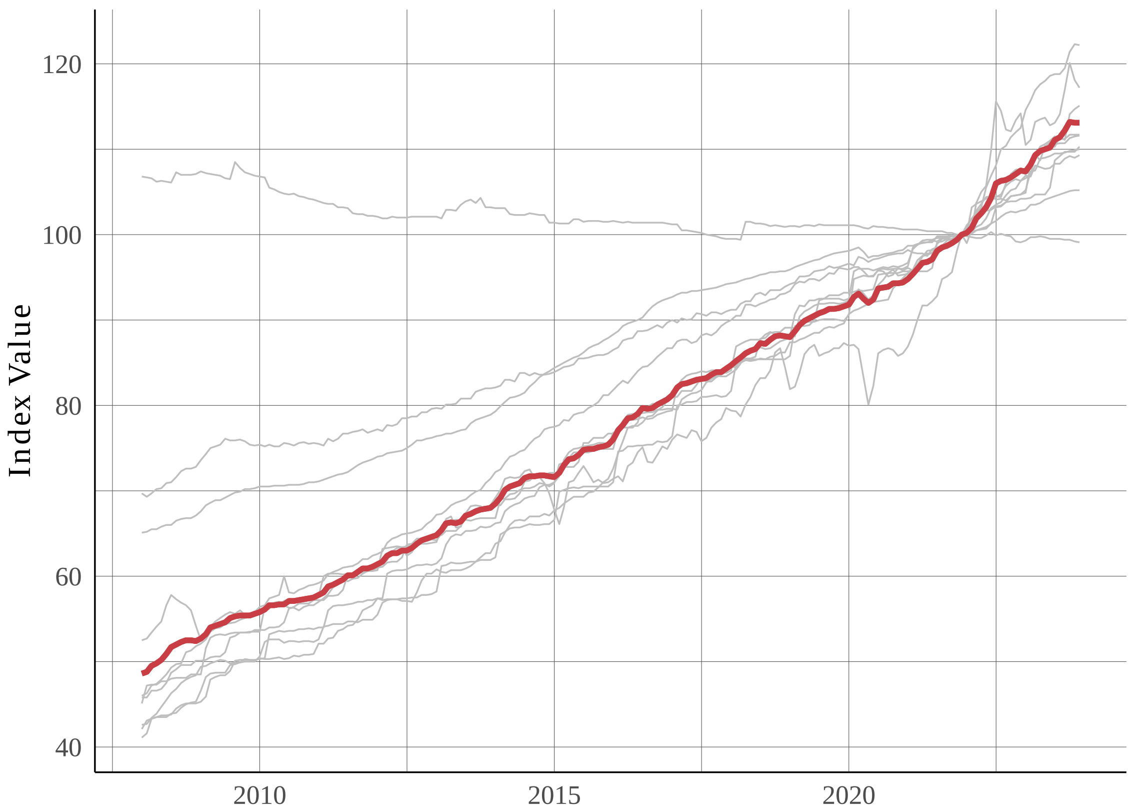

Figure 4.1: CPI for All Items Over Time. CPI (all items) in red. Grey lines represent other CPI components.

Tables 4.1 and 4.2 show the result of the ARIMA model, an ARIMA(3,1,1)(2,0,0) model which indicates a seasonally adjusted series with first-order differencing. With reference to Table 4.1, the autoregressive part consists of three lags with coefficients (AR1, AR2, and AR3) suggesting a moderate positive effect at the first lag and smaller, mixed effects at subsequent lags. The moving average term (MA1) has a strong negative coefficient, implying a quick reversal effect on the series following a shock. Seasonal terms (SAR1 and SAR2) with one year lag indicate positive effects. The relatively low σ^2 suggests minor error variability.

Table 4.1: ARIMA Model Coefficients

Coefficient	Estimate	Std. Error	Statistic
AR1	0.4385	0.0758	5.785
AR2	-0.2619	0.0785	-3.336
AR3	0.0875	0.0767	1.141
MA1	-0.9883	0.0185	-53.423
SAR1	0.3415	0.0715	4.776
SAR2	0.3086	0.0783	3.941
σ^2	0.06534	-	-

The training set error measures (Table 4.2), including ME, RMSE, MAE, and MASE, are relatively low, suggesting a good fit. The near-zero ACF1 indicates little autocorrelation in the residuals.

Table 4.2: Error Measures of the ARIMA Model

Error Measure	Value
ME	0.012364
RMSE	0.250896
MAE	0.198046
MASE	0.842047
ACF1	-0.005765

Tables 4.3, 4.4 show the results from the GARCH model. The model used is a standard GARCH(1,1) with a normal distribution for errors. Table 4.3 suggest that the volatility of the series is persistent, as indicated by the high estimate (close to 1) of the β_1 parameter. The relatively high p-value for the ω and α_1 parameters indicates they are not significantly different from zero at conventional levels, suggesting a limited short-term impact on volatility from previous shocks. The Ljung-Box tests on standardized and squared residuals (Table 4.4) indicate no serial correlation in the residuals or squared residuals, implying a good fit of the model to the data.

Table 4.3: Optimal Parameters of GARCH Model

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
ω	0.000470	0.002491	0.18854	0.85045
α_1	0.043695	0.038903	1.12319	0.26136
β_1	0.955305	0.075796	12.60358	<0.0001

Table 4.4: Weighted Ljung-Box Test on Standardized Squared Residuals

Lag	Statistic	p-value
Lag[1]	2.404	0.12102
Lag[$2(p+q)+(p+q)-1$][5]	7.033	0.05091
Lag[$4(p+q)+(p+q)-1$][9]	9.982	0.05061
Degrees of freedom (d.o.f): 2		

Figure 4.2 illustrates the estimated volatility of a financial series as deduced from a GARCH model, reflecting how risk levels have evolved over time. A noticeable downward trend in volatility from 2008 to 2015 is observed, suggesting a period of relative stability. However, the graph also captures several pronounced spikes, particularly around 2020, indicative of increased market turbulence or significant economic events – such as the COVID-19 pandemic. The conditional volatility captured by the GARCH model only serves to substantiate its need for modelling inflation.

Figure 4.2: GARCH Model Conditional Volatility (Sigma) Over Time

Figure 4.3 illustrates the forecasted change in the Consumer Price Index (CPI) with 95% confidence intervals. The red line represents the median prediction of CPI changes over time, while the shaded area depicts the range within which the true CPI changes are expected to lie with a 95% probability. The forecast extends from the beginning from January 2024 to December 2025, showing fluctuations in the CPI with periods of both increases and potential retractions.

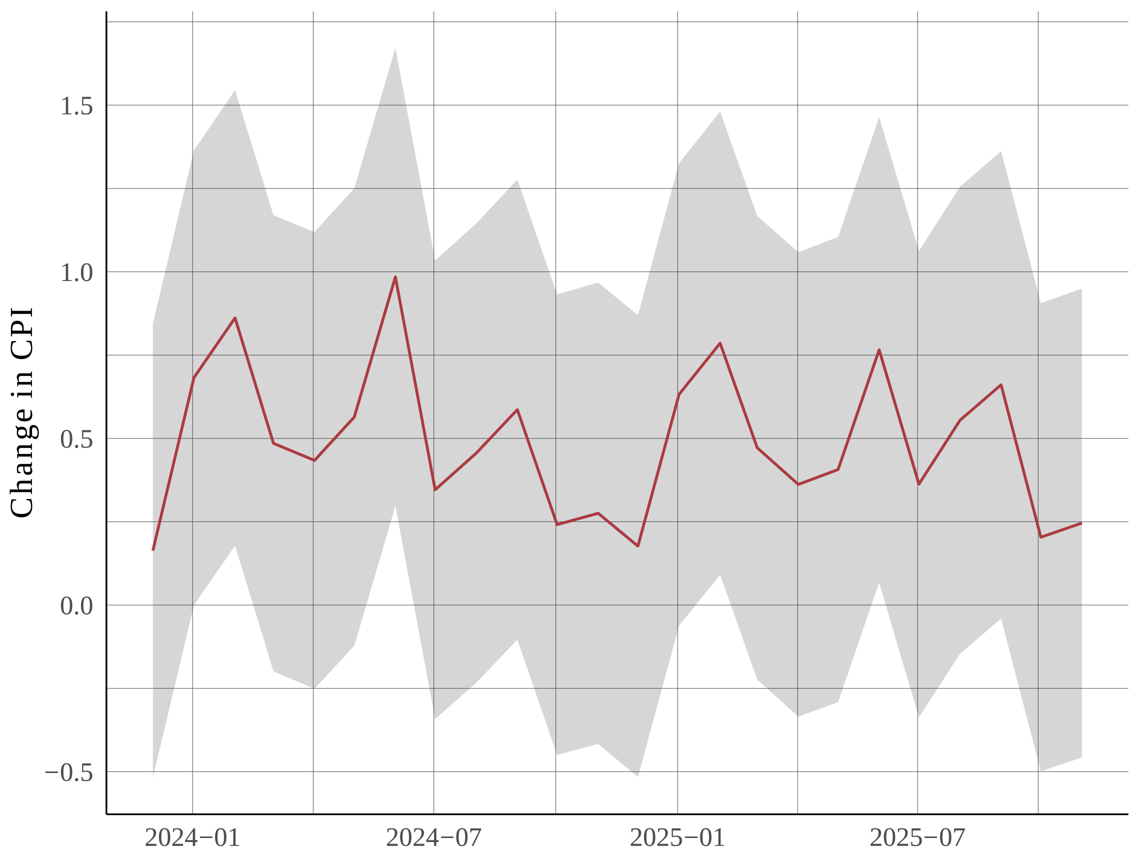

Figure 4.3: Forecasted change in CPI (with 95% Confidence Intervals)

Of course, a model is evaluated on the basis of the accuracy of its forecasts. Figure 4.4 displays a comparative analysis between the actual CPI for all items and the forecasted CPI over a period of 16 months. The forecast for this period was produced based on the observed values of the period before July 2022. The blue line represents the actual CPI values, while the red line denotes the forecasted CPI values. Both lines exhibit an increasing trend over the observed period. The forecasted CPI closely mirrors the actual data, suggesting that the forecasting model is capturing the underlying trend in the CPI effectively. This visual comparison serves as a testament to the model's predictive capability in estimating future inflationary trends.

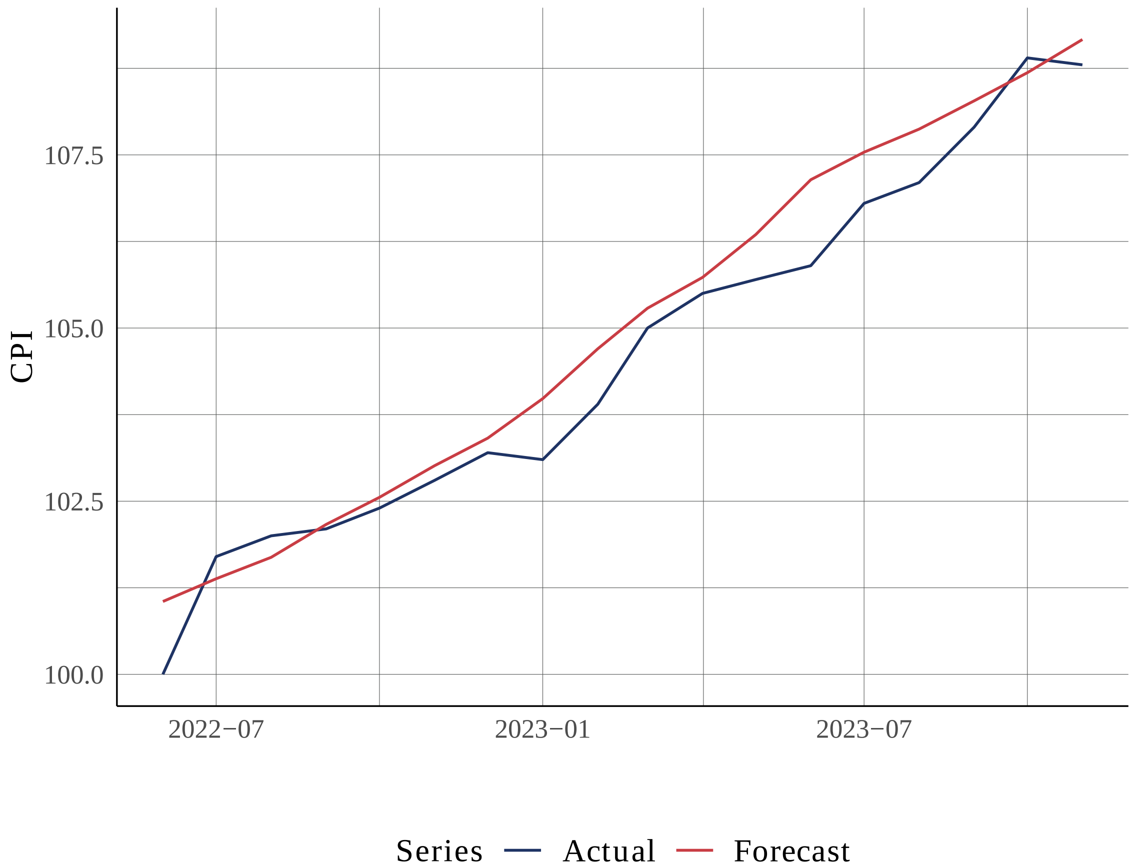

Figure 4.4: Forecasted change in CPI (with 95% Confidence Intervals)

5. Discussion & Conclusion

The study engaged in a sophisticated modeling procedure incorporating ARIMA and GARCH models to analyze and forecast the CPI in South Africa, with an emphasis on capturing the inherent volatility and trend components. The fusion of these models into a cohesive forecasting framework was operationalized through a user-friendly web application built using R Shiny, which facilitated interactive engagement with the data and model outputs.

The ARIMA model adeptly captured the temporal dependencies within the CPI series, while the GARCH model effectively encapsulated the volatility patterns, a critical aspect given the recent unpredictability in inflation rates. The performance of these models was benchmarked against actual CPI data, revealing a commendable forecasting prowess that promises to be an invaluable asset for policymakers and economists alike.

In conclusion, the integration of these statistical models within an automated dashboard presents a powerful tool for real-time economic analysis, offering both precision in forecasting and ease of use through the web application. This study not only contributes to the existing financial econometrics literature but also paves the way for future research to refine and expand upon the methodologies and technologies employed.

References

Baillie, R.T., Chung, C.-F. & Tieslau, M.A. 1996. Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of applied econometrics (Chichester, England)*. 11(1):23–40.

Belkhouja, M. & Mootamri, I. 2016. Long memory and structural change in the G7 inflation dynamics. *Economic modelling*. 54:450–462.

Boateng, A., Gil-Alana, L.A., 'Maseka, L., Siweya, H. & Belete, A. 2016. LONG MEMORY AND ARFIMA MODELLING: THE CASE OF CPI INFLATION RATE IN GHANA. *The Journal of developing areas*. 50(3):287–304.

Bollerslev, T. 2023. Reprint of: Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*. 234:25–37.

Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*. 50(4):987–1007.

Nyoni, T. 2019. Modeling and forecasting CPI in myanmar: An application of ARIMA models. *IDEAS Working Paper Series from RePEc*.

Statistics South Africa. 2024.

Appendix

Mathematical Appendix

The modelling approach involves a three-step process:

1. **ARIMA Modeling**: In the first step, I apply ARIMA modeling to the CPI data to extract the underlying trends and seasonality patterns.

Given the time series CPI data sequence (Y_t), the ARIMA(p,d,q) model, can be expressed as:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)(1 - L)^d Y_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\epsilon_t$$

Where:

- Y_t represents the observed value at time t.

- L is the lag operator, which shifts the time index back by one step.
- p is the order of the autoregressive (AR) component.
- d is the degree of differencing applied to make the series stationary.
- q is the order of the moving average (MA) component.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.
- ϵ_t represents white noise, assumed to be independent and identically distributed with mean zero and constant variance.

This expression represents the general form of an ARIMA model, and the specific coefficients (ϕ and θ) are then estimated.

2. **GARCH Modeling**: Following the ARIMA analysis, I implement a GARCH model to examine the volatility and conditional heteroskedasticity in the CPI series. That is, I use the residuals from the ARIMA process as an input into a GARCH model. This step allows one to assess the degree of uncertainty and risk associated with inflation in South Africa. Given a time series of squared residuals (ϵ_t^2), a GARCH(1,1) model can be expressed as:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Where:

- σ_t^2 is the conditional variance at time t.
- ω is the constant term representing the long-term average of the conditional variance.
- α_1 is the autoregressive parameter for the conditional variance.
- ϵ_{t-1}^2 is the squared residual error from the previous time step.
- β_1 is the moving average parameter for the conditional variance.
- σ_{t-1}^2 is the conditional variance from the previous time step.

3. **Synthesis**: The synthesis of the ARIMA and GARCH models involves combining the point forecasts from the ARIMA model with the volatility forecasts from the GARCH model. Let Y_t represent the observed values of the Consumer Price Index (CPI) time series. The ARIMA model provides point forecasts for Y_t as $\hat{Y}_t$. The GARCH model provides forecasts of conditional variance or volatility, denoted as $\hat{\sigma}_t$, for each time step in the forecast horizon. Then the 95% confidence interval for $\hat{Y}_t$ can be calculated as:

$$\hat{Y}_{t,\text{Upper}} = \hat{Y}_t + z_{\alpha/2} \cdot \hat{\sigma}_t$$
$$\hat{Y}_{t,\text{Lower}} = \hat{Y}_t - z_{\alpha/2} \cdot \hat{\sigma}_t$$

Where:

- $z_{\alpha/2}$ is the critical value of the standard normal distribution corresponding to the desired confidence level (e.g., $z_{0.025}$ for a 95% confidence interval).