ANALYSIS OF TIME SERIES MODELS FOR FORECASTING GROSS DOMESTIC PRODUCT (GDP) GROWTH IN THE CONTEXT OF ECONOMIC VOLATILITY: A CASE STUDY OF ZIMBABWE FROM 2000 TO 2023

DISSERTATION

Submitted in partial fulfilment of the requirements for the degree of

BACHELOR OF SCIENCE HONOURS

In

Data Science

Faculty of Information Technology

By

Chido Sandra Zinyakatira

Supervisor: Dr Banjo Aderemi

November 2025

Declaration

I hereby declare that the study I am submitting to Eduvos, titled Analysis Of Time Series Models For Forecasting Gross Domestic Product (GDP) Growth In The Context Of Economic Volatility: A Case Study Of Zimbabwe From 2000 To 2023, complies or partially complies with the requirements established for the degree, is my original work, which was written in accordance with the requirements, and has not been submitted to any other institution.

Chido Sandra Zinyakatira

Acknowledgements

I wish to sincerely thank my supervisor, Dr. Banjo Aderemi, for his guidance, support, and patience throughout this research project. His feedback and encouragement helped me stay focused and develop academically.

I am also immensely grateful to my aunt, who played a significant role in encouraging me to pursue data science for my honour’s degree. Her belief in my potential gave me the confidence to follow this path.

Most importantly, I would like to express my heartfelt thanks to my mother, who has supported me throughout university as a single parent. I could not have embarked on this journey without her sacrifices, resilience, and ongoing encouragement. Her dedication made this achievement possible, and I am truly thankful.

Dedication

I dedicate this dissertation to my mother. Her hard work, sacrifices, and constant encouragement have carried me through every stage of my studies. Her love and dedication have made this achievement possible, and I am forever grateful.

Research outputs

Research Indaba Conference paper

Eduvos Research Indaba presentations

Abstract

This study evaluates the performance of three conventional time series models: ARIMA, Holt's Exponential Smoothing, and the Vector Error Correction Model in predicting Zimbabwe's Gross Domestic Product (GDP) growth from 2000 to 2023. The period encompasses significant economic transformations, including hyperinflation, dollarisation, the reintroduction of the currency, and the COVID-19 pandemic. These shifts have engendered varying levels of economic stability, thereby enabling the assessment of each model's efficacy under conditions that range from stable to unstable and transitional.

The research employed annual GDP data and utilised an expanding window methodology to generate one-step, three-step, and four-step ahead forecasts. The accuracy of the models was assessed using metrics such as Mean Absolute Error, Mean Squared Error, Root Mean Squared Error and Mean Absolute Percentage Error. The findings indicate that no single model is universally superior across all scenarios. Holt's Exponential Smoothing demonstrated commendable performance in short-term forecasting during periods of stability, whereas ARIMA consistently proved to be the most reliable model overall. Rolling cointegration analysis revealed that the outright failures of VECM resulted from model specification based on full sample cointegration evidence (trace: 20.59), which obscured the complete absence of equilibrium relationships within operational forecast windows (0% detection from 2009 to 2018). Furthermore,

The study concludes that forecast accuracy diminishes as the forecast horizon extends- particularly during unstable periods- rendering three-year and four-year forecasts substantially unreliable. These results underscore the necessity for model selection to be contingent upon the prevailing economic environment and acknowledge inherent limitations in forecasting within unstable economies, limitations that no model can entirely surmount.

The research was constrained by a limited sample size of 24 observations and reliance on annual data. Future investigations should incorporate quarterly data, include additional variables, and compare traditional models with contemporary machine learning techniques. Overall, the findings furnish valuable practical guidance for economic forecasting in developing economies and emphasise the importance of aligning forecasting methodologies with prevailing economic conditions.

Table of Contents

List of Figures

List of Tables

Acronyms

GDP – Gross Domestic Product

ARIMA – Autoregressive Integrated Moving Average

VECM – Vector Error Correction Model

VAR – Vector Autoregressive

MAE – Mean Absolute Error

MSE – Mean Square Error

MAPE – Mean Absolute Percentage Error

RMSE – Root Mean Square Error

ADF – Augmented Dickey-Fuller

DM – Diebold-Mariano

AR – Autoregression

MA – Moving Average

ARMA – Autoregressive Moving Average

ACF – Autocorrelation Function

PACF – Partial Autocorrelation Function

AIC – Akaike Information Criterion

BIC – Bayesian Information Criterion

OLS – Ordinary Least Squares

RBZ – Reserve Bank of Zimbabwe

US – United States

CPI – Consumer Price Index

ZIMSTAT – Zimbabwe National Statistics Agency

IMF – International Monetary Fund

RTGS – Real Time Gross Settlement

Chapter 1: Introduction

1.1. Background and Context

Economic forecasting plays a crucial role in informing policy decisions, investment strategies, and macroeconomic planning, particularly in developing economies where volatility and uncertainty are prevalent (Chauvet and Potter, 2013). GDP forecasting represents a critical application of time series analysis, as GDP serves as the most comprehensive measure of economic performance and forms the vital basis for government economic development strategies and policies. Ning et al. (2010) establish that GDP data represents the most important index for assessing national economic development and judging macroeconomic performance, making accurate GDP forecasting essential for effective policy formulation. One of the most important economic indicators that researchers try to forecast is Gross Domestic Product (GDP) growth, which measures the overall health and performance of a country's economy. Traditional time series models are foundational tools for economic prediction offering policymakers and researchers a structured approach to understanding and anticipating economic trends as explained by De Gooijer and Hyndman (2006).

To forecast economic indicators like GDP growth, economists typically rely on traditional time series models. These models have been used for decades and form the foundation of economic forecasting. The three main types of traditional models are ARIMA (Autoregressive Integrated Moving Average) models, exponential smoothing methods, and VAR (Vector Autoregression) models. ARIMA models work by looking at how current economic data relates to past data and using these relationships to make predictions (Hyndman & Athanasopoulos, 2018). Exponential smoothing methods give more weight to recent data when making forecasts, assuming that newer information is more relevant than older information (Hyndman & Athanasopoulos, 2018). VAR models are more complex because they look at multiple economic variables at the same time and consider how they influence each other, such as how GDP growth and inflation rates might be connected (Sims, 1980; Bańbura et al., 2010).

While these traditional forecasting methods work well in stable economic conditions, they face serious challenges when applied to countries experiencing extreme economic problems. Zimbabwe presents a unique and interesting case study for examining the performance of traditional time series forecasting models under extreme economic conditions. Between 2000 and 2023, the country experienced one of the most severe hyperinflation levels in modern economic history, reaching a peak of 79.6 billion percent in 2008 (Hanke and Kwon, 2009), followed by periods of relative stabilisation through dollarisation, and subsequent renewed volatility. This extraordinary economic trajectory provides an ideal environment for testing how conventional forecasting approaches perform across dramatically different economic regimes.

1.2. Research Problem

Despite extensive theoretical development in time series forecasting, significant gaps remain in understanding how traditional models perform under extreme economic volatility, particularly in developing economies (Hamilton, 2011). Studies like the one done by Poon in 2005 extensively explored the use of models like ARIMA in stable economic environments; however, there remains limited research during economic instability, especially where these models are applied to a limited dataset. The challenge is even more pronounced when considering that economic forecasting in developing countries often operates under constrained data availability and quality, structural breaks and institutional changes that may make it difficult to follow the underlying assumptions of many traditional time series models. Zimbabwe's economic environment represents a very compelling manifestation of these challenges.

Furthermore, existing literature provides a limited comparative analysis of how different traditional perform relative to each other during periods of varying economic stability (Sims 1980; Taylor 2004).

1.3. Research Objectives and Questions

This study addressed these gaps by conducting a comprehensive comparative analysis of traditional time series models for forecasting Zimbabwe's GDP growth across different periods of economic stability and volatility. The primary objective was to determine which traditional forecasting approaches provide the most reliable predictions under varying economic conditions in a volatile developing economy context.

The research addresses two fundamental questions: First, which traditional time series models provides more accurate forecasts during periods of economic stability, instability, and transition? Specifically, comparing ARIMA, Holt's Exponential Smoothing, and VECM performance during periods of economic stability (2009-2013), transition (2014-2018), and crisis (2019-2023), evaluating whether model rankings remain constant or shift systematically with economic conditions. Second, how does forecasting accuracy vary between short term and long-term predictions during different economic periods? This research examines whether the rate at which error accumulate over 1-step, 3-step, and 4-step ahead forecasts differs between stable and volatile periods, identifying practical limits on forecasting horizons under different economic regimes.

1.4. Methodology Overview

This study employs a quantitative research approach using comparative time series analysis to evaluate the forecasting performance of traditional models across Zimbabwe's distinct economic periods from 2000 to 2023. The methodology utilises 24 years of annual GDP data alongside supporting macroeconomic indicators, applying an expanding window approach optimised for limited data availability.

The analysis framework comprises three primary components: systematic model estimation using established econometric techniques (Brooks, 2019), forecast generation through expanding window analysis, and performance comparison using multiple accuracy measures including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and statistical significance testing (Hyndman and Athanasopoulos, 2018). This approach enables robust comparison of forecasting accuracy across different economic regimes while accounting for the methodological challenges inherent in volatile developing economy contexts.

1.5. Structure of the dissertation

The remainder of this dissertation is organised as follows: Chapter 2 presents a comprehensive literature review examining theoretical foundations of time series forecasting and previous studies on economic prediction in volatile environments. Chapter 3 details the research methodology, including data collection procedures, model specification, and analytical framework. Chapter 4 presents the empirical results, including comparative performance analysis across different economic periods and forecasting horizons. Chapter 5 concludes the study with recommendations for future research and policy applications. This systematic approach ensures comprehensive coverage of the research questions while maintaining methodological rigour appropriate for econometric forecasting analysis in challenging economic contexts.

Chapter 2: Literature Review

2.1. Forecasting under volatile environments

Economic forecasting in volatile environments presents significant challenges for policy makers and researchers alike. Due to differences in governance systems, institutional qualities, degree of industrialization, and the level of development, future uncertainties coupled with adverse shocks like the global financial crisis and the COVID-19 pandemic pose varying degrees of vulnerabilities and negative impacts on rich and poor countries (Mohamed, 2022). Furthermore, Engle and Rangel (2004) note that volatility varies significantly across countries and over time, particularly in emerging markets where macroeconomic conditions tend to be more volatile.

The systematic and reliable predictions of future economic conditions help governments not only cope with current economic challenges but also correctly set the foundations of economic policy that generate more certain economic outcomes in the future (Mohamed, 2022).

According to Bloom (2014) forecasting accuracy significantly reduces during periods of high volatility. This has been attributed to multiple factors, including model misspecification, parameter instability, and the failure to account for structural shifts in economic fundamentals (Stock & Watson, 2007). As Hamilton (2011) noted, economic time series often exhibit regime switching behaviour, where different data generation processes operate during stable versus unstable periods.

Traditional time series models, particularly ARIMA models, face fundamental theoretical constraints when applied to volatile economic environments. Poon (2005) argues that these models struggle to capture the complexities of economic volatility in unstable settings due to their heavy reliance on historical data and the assumption of stationarity. This assumption becomes particularly problematic during periods of hyperinflation, currency reforms, or abrupt policy shifts, where the underlying economic structure experiences rapid transformation. Poon (2005) further highlights that traditional approaches often overlook structural breaks and regime shifts, which are common characteristics of volatile economies.

Engle and Rangel (2004) further critique traditional models for their assumption that volatility reverts to a constant mean over time. This mean-reversion assumption proves unrealistic in practice, as volatility often exhibits long-term trends and cycles driven by macroeconomic factors. Additionally, traditional models primarily rely on past financial data to predict future volatility without explicitly incorporating macroeconomic variables that could provide crucial information about economic conditions and their impact on financial markets.

Rolling-window estimations frequently reveal substantial fluctuations in forecast regression coefficients, demonstrating that model performance and predictive relationships shift over time in response to economic shocks and regime changes (Rossi, 2011). As a result, traditional rationality tests such as the Mincer–Zarnowitz framework become unreliable in volatile settings. The forecasting literature therefore emphasizes the importance of methods specifically designed to account for instability such as fluctuation tests and instability-robust regression procedures which provide more credible assessments of forecast performance in economies exposed to persistent shocks and structural change (Giacomini & Rossi, 2010; Rossi & Sekhposyan, 2011).

2.2. Traditional Time series Models: Methodological Overview

Time series analysis refers to the study of data points collected sequentially over time at uniform intervals. Among the foundational models in time series are the autoregressive (AR) and moving average (MA) models, initially developed through the works of Yule, Slutsky, Walker, and Yaglom. The integration of these models gives rise to the autoregressive moving average (ARMA) framework, which forma the theoretical basis for more advanced methods such as the ARIMA model (Mondal, Shit, & Goswami, 2014).

The Auto-Regressive Integrated Moving Average model (ARIMA) has turned out to be the most commonly used model for forecasting inflation, with research works focusing on modelling and forecasting inflation rates and investigating the performance of these models (Jere and Siyanga, 2016). ARIMA models and their extensions, such as SARIMA models, have been widely applied to forecast economic variables, demonstrating their versatility in handling various time series characteristics including seasonal patterns.

Box and Jenkins (1960) pioneered the autoregressive integrated moving average (ARIMA), which led to a new generation of forecasting tools useful to analyse the probabilistic, or stochastic, properties of economic time series on their own (Gujarati et al., 2015). The Box-Jenkins approach provides the systematic framework for ARIMA model development through four iterative stages which are model identification, estimation, checking, and forecasting (Abonazel & Abd-Elftah, 2019). Model identification involves ensuring stationarity, identifying seasonality, and using Auto correlation Function (ACF) and Partial Auto correlation Function (PACF) plots to determine appropriate autoregressive and moving average components. Model estimation employs Maximum Likelihood Estimation or nonlinear least squares to determine optimal coefficients, while model checking validates that residuals conform to white noise specifications.

Exponential smoothing was proposed in the late 1950s and is one of the successful forecasting methods. Forecasting which is produced with the use of exponential smoothing methods are weighted averages of past values, with the weights decaying exponentially as the observations expand (Dritsaki and Dritsaki, 2021). Exponentially smoothing covers a wide range of methods, including some recent advances, such as the long run approaches of the united exponential smoothing and the linear Holt methods which continue to be of vital importance in this sector.

Holt-Winters model is essentially a method to fit appropriate curves to past data of time series (Gujarati et al., 2015). In 1957, Holt (2004) worked on exponential smoothing methods that showed trend and seasonality. This gained prominence after Winters (1960) tested several exponential moving methods with Holt’s empirical data, now known as Holt-Winters forecasts. Brown (1959), Holt (1957), and Winters (1960) pioneered exponential smoothing, which became the foundation of widely accepted forecasting tools where forecasts are calculated as weighted averages of historical observations, with relatively greater weights assigned to recent observations.

The VAR model consists of linear models that capture the joint dynamics of multiple series taking each endogenous variable as lagged function of all endogenous variables in the system (Erkekoglu et al., 2020). They are famously used for forecasting that allocates weights to different series to account for fluctuations in the data. In this method, forecasts are obtained by smoothing past values of the series in an exponential process that decays over the mean of the data (Erkekoglu et al., 2020).

2.3. ARIMA Models: Performance and Limitations in Volatile Contexts

Empirical evidence demonstrates that ARIMA model effectiveness depends critically on achieving stationarity, particularly challenging during volatile economic periods. Abonazel and Abd-Elftah (2019) illustrate this challenge through Egypt's GDP analysis, where visual inspection and formal testing revealed non-stationary behaviour requiring second-order differencing to achieve stability. Their analysis identified ARIMA (1,2,1) as optimal, with the MA (1) coefficient statistically significant at 1% level while the AR (1) coefficient proved non-significant.

The study achieved strong in-sample performance with minimal differences between actual and fitted values, demonstrating ARIMA's capability under relatively stable conditions. However, the authors acknowledge that economic systems represent complex and dynamic environments where forecasting accuracy faces inherent Limitations, particularly during periods of severe economic fluctuations (Abonazel & Abd-Eftah, 2019).

Khan and Khan (2020) provide comparative evidence on ARIMA performance relative to multivariate approaches, demonstrating that ARIMA models perform adequately when correlations between economic variables are low. However, when economic indicators exhibit strong interdependencies, ARIMA's univariate approach fails to capture valuable information contained in variable relationships, resulting in inferior forecasting performance compared to VAR models.

Empirical applications demonstrate that ARIMA can yield robust predictions across stable financial sectors. For instance, when applied to fifty-six stocks from seven industries on India's National Stock Exchange, ARIMA produced forecast accuracies exceeding 85%. The model performed particularly well in the fast-moving consumer goods sector and achieved above 90% accuracy in the information technology sector, where exchange rate dynamics play a strong role in stock price variation (Mondal et al., 2014).

Despite these strengths, Mondal et al.'s (2014) ARIMA exhibited notable limitations in more volatile and structurally complex sectors. Forecasts for banking, steel, and automobile companies showed larger standard deviations in predictive accuracy, indicating weaker reliability. This suggests that ARIMA's performance diminishes in contexts where shocks and fluctuations are more pronounced. Such findings underline the challenges of applying univariate ARIMA models in unstable environments, where external shocks and regime shifts may dominate data dynamics.
De Gooijer and Hyndman (2006) suggest that ARIMA models may perform poorly when structural breaks or regime shifts define an economy, as was the case with Zimbabwe between 2007 and 2023. During periods of hyperinflation, such as Zimbabwe experienced in 2008, traditional ARIMA models would likely fail to produce reliable forecasts due to the unprecedented and non-linear nature of economic collapse.

ARIMA models have demonstrated effectiveness in forecasting various economic indicators across different contexts. Olayiwola, as cited in Atoyebi et al. (2023), developed an ARIMA (4,1,4) model to predict road traffic fatality in South Africa, which was concluded to be best among the set of models considered due to its lowest volatility, lowest information criterion, and highest adjusted

R-squared values. Furthermore, Olayiwola and Atoyebi, referenced in Atoyebi et al. (2023), conducted an investigation using the Box-Jenkins approach to forecast fuel prices and the strength of the South African rand based on 35 years of monthly data. The optimal models identified were ARIMA (3,1,1), ARIMA (3,1,1), ARIMA (1,1,2), and ARIMA (1,0,1) for diesel, paraffin, petrol, and exchange rate (ZAD-USD) respectively achieving forecasting accuracies of 93.4%, 91.7%, 91,5%, and 79,3% respectively.

In predicting a time series data of GDP as a macroeconomic variable, ARIMA model has been proven to be reliable and accurate (Alsinglawi et al., 2022). Several studies have demonstrated the effectiveness of ARIMA models in different contexts. Wand and Wang (2011) deployed ARIMA for predicting the GDP of China based on time series data from 1978 to 2006, choosing the best ARIMA model based on statistical tests to predicted value was insignificant, demonstrating that the ARIMA model for GDP data series from1952 to 2007 to predict the GDP of the Shanxi province in China, finding that error between the real GDP value and the predicted value was within 5% range (Alsinglawi et al., 2022).

Nyoni (2018) addressed structural breaks in time series data through appropriate differencing techniques to achieve stationarity, enabling the model to better capture volatility patterns in Zimbabwe's FDI flows. This approach suggests that while standard ARIMA models may struggle with structural instability, modified versions can provide more robust forecasting capabilities.
Alsheheri (2025) demonstrates that ARIMA (1,1,0) models exhibit robust performance for datasets characterized by moderate persistence and trending behaviour under stable economic conditions. The study reports favourable statistical metrics, including a Mean Squared Error of 1.05301, an Akaike Information Criterion of 262.16, and a Bayesian Information Criterion of 267.14. However, Alsheheri (2025) acknowledges that these performance indicators were obtained under relatively stable market conditions, thereby raising important questions regarding model efficacy during periods of heightened economic volatility, such as Zimbabwe's hyperinflationary episodes.

Mutale and Murape (2024) contribute significantly to the literature by examining time series forecasting methodologies specifically within Zimbabwe's turbulent economic landscape. Their research addresses the inherent challenges of non-stationarity in Zimbabwean economic data through the development of adapted differencing techniques tailored to volatile economic indicators. This methodological adaptation represents an important contribution to the application of ARIMA models in developing economies characterized by structural instability and extreme volatility.

Bonga (2020) employs an ARIMA (1,1,1) specification for the analysis and forecasting of remittance flows, demonstrating the versatility of ARIMA methodology across different economic variables. The study emphasizes the critical importance of post-estimation diagnostic testing, particularly normality testing of model residuals, as a means of validating model appropriateness. The assumption of normally distributed residuals serves as a fundamental diagnostic criterion for assessing model adequacy and ensuring the reliability of forecasting outputs in volatile economic environments.

However, these favourable assessments of ARIMA performance share a critical limitation, evaluation predominantly occurs during periods of relative economic stability or moderate volatility.

Mondal et al.'s (2014) Indian stock analysis examined data from 2010-2014, a period of sustained growth. Vafin's (2020) seven-economy study covered 2020-2024 projections anchored in pre-pandemic trends. Even studies acknowledging structural breaks (Nyoni, 2018) apply differencing techniques assuming the transformed series achieves stationarity, an assumptions potentially violated during hyperinflation where economic relationships change continuously rather than exhibiting discrete breaks. This creates an empirical gap, ARIMA's comparative performance when economic regimes shift fundamentally, such as currency abandonment or reintroduction remains inadequately documented.

2.4. Vector Autoregression Models: Multivariate forecasting under instability

VAR models are a seminal work of Chris Sims, derived from his famous critique of large-scale traditional macro econometric models. They are essentially multivariate linear time series models that capture the joint dynamics of multiple time series by treating each endogenous variable as a lagged function of all endogenous variables in the system (Erkekoglu et al., 2020). The VAR process starts with the specification and estimation of reduced form VAR to model diagnostics, and once models satisfy the diagnostics requirement, they are better used for forecasting or structural analysis (Erkekoglu et al., 2020).

The theoretical foundation of VAR models lies in their ability to capture multivariate dependencies through simultaneous equation systems, where each variable depends on its own lagged values and the lagged values of all other variables in the system. This comprehensive approach enables VAR models to exploit information contained in the forecasts during periods when these relationships remain stable (Khan and Khan, 2020).

Vector Autoregression models have proven particularly effective for capturing dynamic relationships between economic variables, making them valuable for GDP forecasting when multiple economic indicators interact significantly. Khan and Khan (2020) demonstrate that VAR models excel when strong correlations exist between economic variables, delivering superior forecast performance compared to univariate approaches for highly correlated indicators such as GDP versus GNP, and exports versus imports. Their empirical analysis of Bangladesh's economic variables revealed that VAR models consistently outperformed ARIMA when forecasting interconnected economic indicators.

Research has shown that VAR models can effectively capture the impact of external shocks on macroeconomic variables. For instance, using VAR methodology, the impact of crude oil price changes on key macroeconomic variables showed that oil prices have significant impact on real GDP, money supply and unemployment, though the impact on consumer price index was not significant (Umar & Kilishi, 2010). This demonstrates the model's capacity to identify differential impacts across variables.

The superiority of multivariate approaches in capturing complex economic relationships has been demonstrated in various emerging market contexts. Abdul Razak, Khamis and Abdullah (2017) conducted a comparative analysis of univariate ARIMA and multivariate VAR models in forecasting economic growth indicators in Malaysia, utilizing monthly data spanning from January 1998 to January 2016. The study examined four key economic indicators: Currency in Circulation (CIC), Exchange Rate (EXC), External Reserve (EXT), and Reserve Money (RM), which collectively serve as proxies for economic growth measurement.

The findings revealed that VAR models significantly outperformed ARIMA models across all economic indicators examined. Specifically, the VAR approach demonstrated substantially lower Mean Absolute Percentage Error (MAPE) values compared to the univariate ARIMA model for forecasting CIC, EXC, EXT, and RM (Abdul Razak, Khamis & Abdullah, 2017). This superior performance can be attributed to VAR's ability to capture the interdependencies and feedback effects between multiple economic variables simultaneously, rather than treating each variable in isolation as done in univariate approaches.

The Malaysian study's timeframe (1998-2016) is particularly relevant as it encompasses periods of economic volatility, including the Asian Financial Crisis aftermath and subsequent recovery phases. The superior performance of VAR models during this period suggests their enhanced capability to handle structural changes and economic instability compared to traditional univariate approaches (Abdul Razak et al., 2017). This finding has important implications for forecasting in volatile economic environments, such as those experienced in Zimbabwe.

However, the performance advantage of VAR models diminishes significantly when correlations between variables are weak. Khan and Khan (2020) found that when economic variables exhibited low correlation, VAR and ARIMA models performed similarly suggesting that the multivariate approach offers no substantial benefit over univariate methods when variable interdependencies are minimal. This finding has important implications for model selection in volatile economies where traditional economic relationships may break down.

The Swedish regional forecasting study provides compelling evidence of VAR model limitations under small sample conditions. Zhang (2013) implemented two variants of VAR models for Stockholm region GDP forecasting, achieving MAPE values of 2.53% and 3.14% respectively. These results were notably inferior to both ARIMA and AR(1) alternatives, highlighting the challenge that VAR models face when applied to datasets with limited observations.

The Madagascar comparative study provides important evidence regarding VAR model limitations in developing economy contexts. Andrianady (2023) found that VAR models were outperformed by ARIMA approaches in quarterly GDP forecasting, suggesting that the theoretical advantages of capturing multivariate relationships may not materialize in practice when applied to developing African economies. This finding aligns with evidence from small-sample contexts, indicating that VAR model effectiveness may be constrained by data quality, structural instability, or model specification challenges common in developing economies.

The underperformance of VAR models in Madagascar's economic forecasting context may reflect broader challenges associated with applying multivariate approaches in structurally unstable economies. Andrianady (2023) did not elaborate on specific structural break issues, but the superior performance of univariate ARIMA approaches suggests that the complex interdependencies that VAR models attempt to capture may be less stable or predictable in developing economy contexts compared to more established economies.

The empirical literature on VAR/VECM performance reveals a critical tension. Studies demonstrating VAR superiority (Abdul Razak et al., 2017; Khan and Khan, 2020) share common characteristics: examination of stable or moderately volatile periods and focus on economies with intact institutional frameworks. Abdul Razak et al.'s Malaysian study covered 1998-2016, including post-Asian Crisis recovery but not fundamental regime change. Khan and Khan's Bangladesh analysis (2020) examined routine macroeconomic fluctuations rather than currency abandonment or hyperinflation.

Crucially, no existing study examines VAR/VECM performance across a complete structural break cycle. Zhang’s (2013) Swedish regional forecasting study hints at limitations, finding VAR underperforms ARIMA in small samples, but operates entirely within stable institutional contexts. Andrianady’s (2023) Madagascar study shows VAR underperformance in a developing African economy but does not explicitly test regime-dependent performance. The literature thus provides evidence the cointegration-based models can excel during stability, but lacks empirical quantification of failure magnitude during regime changes, but lacks empirical quantification of failure magnitude during regimes changes, precisely the scenario most relevant for emerging market forecasters.

2.5. Exponential Smoothing: Adaptive Approaches to economic volatility

Exponential smoothing is an initial empirical work of Holt and Brown on forecasting models for inventory control systems (Erkekoglu et al., 2020). This method assigns greater weight to the most recent series to make up for the latest fluctuations in the data. Forecast values are obtained by smoothing past values of series in an exponential process that decays over the mean of the data, with smoothing parameters assigned offering density to each observation (Erkekoglu et al., 2020). Exponential Smoothing methods represents a collection of different forecasting approaches, with the choice of method driven by key components of time series which are trend and seasonality and the way they manifest in the series. Simple exponential smoothing (SEM) is the easiest of all exponential smoothing methods and is widely used to forecast data which do not indicate a clear trend or seasonal pattern. This model can be represented as:

$$\hat{y}t + 1 \; = \; \alpha yt \; + \; \alpha(1-\alpha)yt - 1 \; + \; \alpha(1-\alpha)^2 yt - 2 + \ldots, where\; 0 \; \leq \; \alpha \; \leq \; 1\; is\; the\; smoothing\; parameter \qquad (1)$$

Liu et al. (2016) highlight several key advantages of exponential smoothing methodology. The method offers low computational cost and can be easily operated on ordinary computers, making it highly accessible for practical applications. It demonstrates exceptional adaptability and can be applied to almost any time series prediction across diverse fields including economics and natural sciences. The method's preference for recent data over historical observations enhances prediction accuracy while utilizing all available historical data through a weighted approach that reduces the impact of outliers. Additionally, exponential smoothing provides computational efficiency by calculating future prediction values using recent actual data and corresponding predictions, significantly reducing time costs for data calculation and processing (Liu et al., 2016).

The US GDP comparative study provides compelling evidence for exponential smoothing methods' effectiveness in long-term economic forecasting. Lyu et al. (2023) found that the Holt-Winters exponential smoothing approach demonstrated smaller mean errors compared to ARIMA when evaluated across the complete forecast series. This superior long-term performance suggests that exponential smoothing's adaptive weighting mechanisms are particularly effective at capturing persistent economic trends and cyclical patterns.

The study's findings indicate that exponential smoothing methods excel at identifying stable, cyclical growth patterns in economic data. Using the Holt-Winters approach, the researchers projected US GDP growth at an average rate of 0.9344% with stable, cyclical upward trends extending through the fourth quarter of 2023 (Lyu et al., 2023). This capability to capture long-term cyclical behaviour makes exponential smoothing particularly valuable for strategic economic planning applications.

According to Hyndman and Athanasopoulos (2018), these methods are particularly effective for short-term forecasting due to their intuitive approach of weighting recent observations more heavily than distant ones. This fundamental characteristic suggests potential advantages in rapidly changing economic environments, as the methods may demonstrate greater responsiveness to sudden economic shifts compared to more complex models like ARIMA.

Despite these theoretical advantages, significant limitations emerge when considering the application of exponential smoothing methods to extreme economic conditions. Hyndman and Athanasopoulos (2018) acknowledge that these methods may underperform during sudden economic shocks, which directly impacts their reliability for forecasting in periods of high economic instability. This limitation becomes particularly pronounced in hyperinflationary environments, where exponential smoothing methods may struggle to adapt quickly enough to capture exponential growth patterns in inflation or other rapidly escalating economic indicators.

The exponential smoothing literature exhibits a methodological tension between theoretical simplicity and empirical application complexity. While Hyndman and Athanasopoulos (2018) acknowledge that exponential smoothing methods may underperform during sudden economic shocks, the practical magnitude and duration of this underperformance remain underspecified. Lyu et al.’s (2023) finding that Holt-Winters outperforms ARIMA for US GDP forecasting might reflect the specific characteristics of developed economy business cycles, gradual, cyclical fluctuations rather than establish exponential smoothing’s superiority under structural instability.

2.6. Model Selection and Evaluation Frameworks

Comprehensive forecast evaluation requires multiple accuracy metrics to capture different aspects of model performance. Khan and Khan (2020) employed a robust evaluation framework using Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE) to assess forecasting accuracy across different models and economic conditions.

Several methods are used for measuring reliability and performance of forecasting methods, including Mean Absolute Percentage Error (MAPE), Root of Mean Squared Error (RMSE), and Mean Absolute Deviation (MAD) (Jafari-Samimi et al., 2007). MAPE is particularly useful for evaluating the performance of various kinds of forecasting models, with the smaller the value of MAPE indicating better forecasting ability.

The signification of MAPE values provides clear performance benchmarks: MAPE <10% indicates excellent forecasting ability, 10-20% represents good forecasting ability, 20-50% shows reasonable forecasting ability, and >50% indicates bad forecasting ability (Jafari-Samimi et al., 2007). This standardized evaluation framework allows for consistent comparison across different forecasting methodologies and provides clear performance thresholds for model selection.

The research methodology of Khan and Khan (2020) emphasized the importance of out-of-sample testing, utilizing a 70;30 training-to-test data split to evaluate genuine forecasting performance rather than in-sample fit. This approach provides more realistic assessment of model capabilities under actual forecasting conditions, particularly relevant for volatile economies where model overfitting to historical data can severely compromise prediction accuracy (Khan & Khan, 2020).
A critical finding for model selection emerges from the correlation dependent performance of different forecasting approaches. Khan and Khan (2020) establish that model selection should be guided by the correlation structure among economic variables, with VAR models preferred for highly correlated variables. This evidence-based selection framework provides practical guidance forecasters operating in different economic contexts.

Comparative studies have revealed mixed results regarding model performance across different sectors. Research findings indicate that while predicting stocks, the performances of both ARIMA and exponential smoothing models were the same for nine stocks on both measures of forecasting errors (Funde and Damani, 2023). However, ARIMA performed better than exponential smoothing for prediction of five stocks (HDFC Bank, IndusInd Bank, KOTAK Bank, Tech Mahindra, and WIPRO), whereas the exponential smoothing model was a better predictor for TCS (Funde and Damani, 2023). However, the difference in performance was found to be marginal across all three sectors, suggesting that the choice between models may depend on specific market conditions and data characteristics.

Given the presence of many forecasting methods, researchers face a dilemma given the absence of a specific yardstick to compare competing models. However, the consensus is the preference of models that ensure accuracy by minimizing forecast errors associated with each model (Erkekoglu et al., 2020). Forecast accuracy is generally measured using the mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), Thiel's U1 and Thiel's U2 statistics (Erkekoglu et al., 2020).

Diagnostic testing plays a crucial role in model validation. The study on Somalia conducted several ARIMA and rolling window diagnostic tests, finding that model errors proved to be white noise, the moving average and autoregressive components were covariance stationary, and the rolling window test showed model stability within a 95% confidence interval (Mohamed, 2022). For the Jordanian study, results showed that predicted values were within the range of 5%, indicating that the prediction capability of the model was relatively adequate and efficient.

2.7. Economic Volatility in Zimbabwe: Context and Measurement

Zimbabwe's economic trajectory provides a compelling case study for understanding extreme volatility in emerging economies. The country's initial economic promise was represented by the strength of its currency. At independence in 1980, the Zimbabwean dollar (ZW$) replaced the

Rhodesian dollar at par, with an exchange rate of one ZW$ to 1.47 US$, demonstrating the nation's economic confidence and aspirations (Nkomazana and Niyimbanira, 2014).

The period between 2000 and 2008 witnessed one of the most severe episodes of hyperinflation in modern economic history. Zimbabwe's economy experienced catastrophic macroeconomic imbalances, with monthly inflation reaching an unprecedented 79.6 billion percent, accompanied by a 40 percent decline in gross domestic product, unemployment exceeding 80 percent, external payment arrears of US$3.07 billion, and a budget deficit of ZW$1760 quadrillion (Nkomazana and Niyimbanira, 2014). The currency's collapse was dramatic: by July 2008, the exchange rate had deteriorated from the initial 1:1.47 ratio to ZW$10 billion to 0.33 US$, driven by substantial increases in money supply (Nkomazana and Niyimbanira, 2014).

Zimbabwe's adoption of full or official dollarisation, as defined by Nkomazana and Niyimbanira (2014), involved making foreign currencies full legal tender while reducing the domestic currency to a subsidiary role. This arrangement eliminated domestic currency risk and the potential for currency crises, with the foreign currencies serving not only as legal tender for private transactions but also for government operations. Nkomazana and Niyimbanira (2014) position this within a three-stage dollarisation framework: unofficial, semi-official, and official, with Zimbabwe achieving the final stage.

The immediate effects of dollarisation on Zimbabwe's economic stability were significant. Nkomazana and Niyimbanira (2014) document that dollarisation achieved price stability and effectively eliminated hyperinflation, with inflation remaining in single digits and even turning negative in early 2009. This stability enhanced business planning capabilities through improved predictability of key economic indicators. However, the strength of the US dollar created competitive disadvantages for local products in international markets, benefiting foreign companies operating in Zimbabwe through higher domestic pricing power (Nkomazana and Niyimbanira, 2014).

The Zimbabwe dollar existed until March 2009, when Zimbabwe essentially adopted a multiple currency system (Mpofu, 2015). Despite the country not having an official agreement with the United States Federal Reserve to use its currency, Zimbabwe is viewed as a dollarised economy given the dominance of the currency among the other currencies used in the country and the fact that the Zimbabwean Government conducts all its business using the US dollar (Mpofu, 2015). In March 2009, the Zimbabwean government suspended the use of the Zimbabwe dollar as legal tender and decreed that all wages, prices of all goods and services would be denominated in foreign currencies (Mpofu, 2015).

Zimbabwe’s economic environment during the 2000–2009 period was characterised by severe macroeconomic instability, sharp declines in output, and extreme inflationary pressures. Real GDP contracted by approximately 30 percent after 1999, driven by the collapse of agricultural production, extensive price distortions, and a highly overvalued exchange rate (Coorey et al., 2007). Inflation accelerated into four-digit and, unofficially, higher ranges, fuelled by rapid money creation to finance large fiscal and quasi-fiscal deficits, particularly through the operations of the Reserve Bank of Zimbabwe (Coorey et al., 2007). By late 2008, hyperinflation effectively rendered the Zimbabwe dollar non-functional, leading to widespread de facto dollarisation and the eventual adoption of a multicurrency system in 2009. This shift helped re-monetise the economy, restore a degree of price stability, and initiate a modest recovery, although underlying structural weaknesses including weak fiscal governance, limited competitiveness, and a fragile external position continued to constrain economic performance (Kramarenko et al., 2010). These dynamics illustrate a macroeconomic landscape marked by volatility, institutional fragility, and significant policy shocks, creating a challenging context for analysing and forecasting GDP growth.

2.8. Research Gap

The literature reveals significant variability in forecasting model performance across different economic contexts, time periods, and data characteristics. While comparative analysis shows that ARIMA provides superior forecast performance compared to VAR and exponential smoothing across multiple accuracy measures, with lower RMSE, MAE, MAPE and Thiel's U1 values (Erkekoglu et al., 2020), other studies suggest a more nuanced picture. Research by Jere and Siyanga (2016) demonstrates that both ARIMA and exponential smoothing methods can provide effective forecasting capabilities, with Holt's exponential smoothing representing an equally viable alternative to ARIMA models.

These contrasting findings indicate that the choice between different forecasting methods depends on specific data characteristics, forecasting horizons, and accuracy requirements rather than one method being universally superior. The evidence suggests that ARIMA minimizes forecast error seriousness and dispersion while providing better goodness of fit in certain contexts, yet both traditional time series methods continue to play important roles in economic forecasting depending on the specific economic environment and data properties.

This variability in model performance across different contexts highlights a critical research gap: the need for comprehensive comparative studies in specific economic environments to determine optimal forecasting approaches for volatile conditions. Current literature lacks sufficient context-specific model evaluation, particularly for different economic environments, which is essential for practitioners seeking to select appropriate forecasting methods for their circumstances.
While the Malaysian study by Abdul Razak et al (2017) provides compelling evidence for VAR model superiority in forecasting economic indicators, several limitations warrant consideration for the Zimbabwean context. The study focused on relatively stable economic indicators in a more developed emerging market context, potentially limiting its direct applicability to highly volatile environments characterized by hyperinflation, currency instability, and frequent policy regime changes as experienced in Zimbabwe.

Chapter 3: Research Methodology

3.1. Introduction

This chapter presents the research methodology employed to investigate the performance of time series models in forecasting Gross Domestic Product (GDP) within Zimbabwe's volatile economic environment from 2000 to 2023. The research adopts a positivist epistemological approach using deductive reasoning and a comparative longitudinal research design with an expanding window approach to evaluate forecasting performance across Zimbabwe's distinct economic phases. Utilising 24 annual observations of Zimbabwe's GDP growth from reputable international databases, the study systematically evaluates three established time series models Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing, and Vector Autoregression (VAR) based on their forecasting accuracy during periods of economic stability, instability, and transition phases. This methodology directly addresses the main research question of evaluating and comparing the performance of traditional time series models in forecasting GDP growth during varying economic conditions in Zimbabwe, with specific focus on determining which models provide more accurate forecasts during stability, instability, and transition periods, while also assessing how multivariate approaches improve forecasting accuracy across different economic environments.

3.2. Research Philosophy

3.2.1. Research Paradigm

This research is grounded within the positivist paradigm, which assumes that a single, tangible reality exists regarding economic relationships and forecasting performance that can be understood, identified, and measured through objective empirical analysis (Park et al., 2019). Ontologically, the study views GDP and economic phenomena as observable, measurable realities that exist independently of the researcher's perceptions, supporting the objective identification of consistent patterns in forecasting accuracy across different economic conditions in Zimbabwe. Epistemologically, the research embraces the positivist position that knowledge must be developed objectively without researcher or participant values influencing its development, employing the hypothetico-deductive method to test specific hypotheses regarding the comparative performance of time series models under varying economic conditions using quantitative measures (Park et al., 2019). This paradigm is particularly appropriate for econometric forecasting research as it supports quantitative measurement, objective evaluation through standardised performance measures (MAE, MSE, RMSE, and MAPE), replicability of findings, and systematic empirical testing of theoretical propositions—all essential for generating reliable, empirically grounded evidence that can inform policy decisions and forecasting practices in volatile economic environments (Collis and Hussey, 2014).

3.2.2. Research Approach

This research employs a deductive reasoning approach, representing a theory-to-data methodology that moves systematically from established theoretical principles in time series forecasting particularly those developed by Box and Jenkins (1970) for ARIMA modelling, Hyndman and Athanasopoulos (2018) for exponential smoothing, and Sims (1980) for Vector Autoregression models to specific empirical testing within Zimbabwe's unique economic context. This approach is particularly appropriate as the research centres on testing and comparing established forecasting methodologies rather than developing new theoretical frameworks, enabling systematic evaluation of existing models against the unique conditions of extreme volatility while building upon robust methodological foundations and well-developed theoretical frameworks that provide established expectations for hypothesis testing. The deductive reasoning framework directly supports the research objectives by establishing a logical structure for systematic model comparison and evidence-based conclusion development, facilitating the investigation of subsidiary research questions through structured examination of model performance across different economic conditions (stability, instability, and transition phases), and enabling the development of generalisable conclusions about optimal forecasting approaches that can inform forecasting practices in similar volatile economic environments and contribute to both theoretical understanding and practical applications in developing economy contexts.

3.3. Research Design

3.3.1. Research Strategy

This research employs a comparative longitudinal design as the primary research strategy, systematically comparing the forecasting accuracy of three established time series models ARIMA, Exponential Smoothing, and VAR across Zimbabwe's distinct economic phases through an ex-post forecasting evaluation framework that uses historical data to generate forecasts for periods where actual outcomes are known, enabling direct comparison between predicted and observed values (Bryman, 2016). This design represents an optimal methodological choice as it enables rigorous evaluation of relative model performance within the same economic context while controlling for external factors such as data sources, evaluation periods, and economic conditions, ensuring that performance differences can be attributed to model characteristics rather than methodological variations or data inconsistencies. The ex-post approach provides several methodological advantages including controlled evaluation of model performance under known economic conditions, systematic assessment of how different models respond to varying degrees of economic volatility, and alignment with established practices in econometric forecasting research that enables meaningful comparison with existing literature.

The comparative longitudinal design demonstrates strong strategic alignment with the research objectives by facilitating the evaluation of model performance during periods of economic stability, instability, and transition phases through systematic frameworks for isolating and analysing these distinct conditions, while supporting the assessment of how external variable inclusion affects forecasting accuracy through controlled comparison opportunities between univariate and multivariate approaches across identical time periods and economic conditions, ensuring that performance differences can be attributed to model specification rather than temporal or contextual factors.

3.3.2. Time Horizon

This research examines Zimbabwe's economic data over a 24-year period from 2000 to 2023, a timeframe methodologically justified as it provides adequate sample size for reliable time series estimation (meeting minimum requirements for robust parameter estimation particularly in VAR analysis), captures sufficient temporal variation to enable meaningful assessment of model performance without bias from short-term fluctuations, and encompasses four distinct economic phases that allow comprehensive testing of forecasting models under various conditions: the Hyperinflation Period (2000-2009) characterised by extreme economic crisis with inflation reaching 79.6 billion percent and currency collapse (Hanke and Kwon, 2009); the Multi-Currency Stabilisation period (2009-2016) when adoption of foreign currencies brought economic stability and moderate growth (Mpofu, 2015); the Currency Reintroduction and Renewed Instability phase

(2016-2019) marked by local currency reintroduction and returning economic problems; and the Stabilisation Attempts Amid Global Shocks period (2020-2023) including COVID-19 impacts and continued economic stabilisation efforts. The study addresses the challenges posed by major economic structural changes through an expanding window approach that allows models to adjust to new conditions while maintaining analytical consistency, ensuring that the analysis 3volatility while covering current economic conditions for practical relevance.

3.3.3. Unit of Analysis

Exponential Smoothing, and VAR), providing a consistent measure for comparing model accuracy across different time periods, while the VECM specification employs a bivariate system incorporating GDP growth and Consumer Price Index (CPI) selected after preliminary analysis of potential explanatory factors based on data availability, statistical significance, and the strong theoretical relationship between price levels and economic growth in Zimbabwe's context. The use of annual rather than quarterly data reflects both data availability constraints as no consistent quarterly series spans the complete study period including the hyperinflation phase (2000-2009) when Zimbabwe's statistical infrastructure deteriorated significantly and methodological advantages for economies experiencing extreme short-term volatility, where quarterly data during hyperinflation would capture noise from erratic price movements, supply disruptions, and measurement error rather than genuine economic trends, with Hanke and Kwon (2009) documenting monthly inflation rate volatility exceeding 100 percentage points during 2008.

The parsimonious bivariate VECM specification is justified by both statistical and economic considerations: statistically, 24 annual observations impose strict constraints on estimable parameters, as a three-variable VECM with two lags would require estimating 18 parameters plus cointegration vectors, consuming excessive degrees of freedom relative to Brooks' (2019) recommended minimum observations-to-parameters ratio of 10:1, whereas the bivariate specification (approximately 12 parameters) satisfies this threshold; economically, the GDP-CPI relationship captures fundamental dynamics particularly relevant to Zimbabwe's context where inflation dynamics centrally determined the economic trajectory through hyperinflation-driven currency abandonment, price stability during dollarization, and renewed inflation accompanying the 2019 crisis, while alternative candidates (government consumption, current account) either lacked consistent stationarity properties or failed cointegration tests, indicating absence of stable long-run relationships suitable for VECM estimation.

3.4. Data Collection Strategy

3.4.1. Secondary Data Approach

This research adopts a secondary data approach, utilising existing economic datasets from established sources rather than collecting primary data, a methodological choice justified by the research objectives requiring historical economic data spanning Zimbabwe's economic transitions from 2000 to 2023 that can only be obtained from existing records maintained by official statistical agencies and international organisations, the need for standardised economic indicators such as GDP growth rates calculated and published by recognised authorities to ensure consistency with international standards, and the research focus on forecasting model performance rather than exploring new economic relationships, making secondary data particularly suitable as it provides the quantitative foundation necessary for statistical analysis and model evaluation.

3.4.2. Data Sources

ZIMSTAT serves as the official national statistical agency and was initially considered as the primary data source for GDP information. However, several limitations were identified during the data collection process that restricted its utility for this comprehensive study.
ZIMSTAT data availability is limited to the period from 2009 onwards, which excludes the crucial hyperinflation period (2000-2009) that forms a central component of this research. This limitation would have prevented examination of model performance during the most volatile economic phase, significantly reducing the study's analytical scope and comparative value.

Additionally, ZIMSTAT presents data with reporting inconsistencies, where some information is provided quarterly while other data points are presented annually. This inconsistent reporting format creates challenges for systematic time series analysis and would require extensive data manipulation that could introduce processing errors.

The IMF was selected as the primary data source due to its comprehensive coverage and consistency advantages. IMF data provides complete coverage of Zimbabwe's economic indicators across the entire study period (2000-2023), including the critical hyperinflation period that is essential for this research.

The IMF dataset demonstrates superior consistency in reporting formats and methodological approaches, maintaining standardised annual data presentation throughout the time series. This consistency is particularly valuable for time series analysis, as it reduces the risk of introducing artificial breaks or inconsistencies that could affect model performance evaluation.
The IMF data required restructuring from its original horizontal format to a vertical format suitable for time series analysis in Python. This restructuring involved converting the data layout from a wide format (where years were presented as columns) to a long format (where each observation represents a single time) to facilitate proper time series analysis.

The restructuring process was conducted in Excel before importing into Python to ensure data integrity and reduce the risk of processing errors. This approach provided visual verification of data alignment and enabled manual quality checks before statistical analysis.
Cross-referencing with World Economic Outlook data confirmed the accuracy and reliability of the selected IMF datasets, providing additional confidence in the data foundation for the analysis. The consistent presentation format across time periods was verified to ensure that methodological changes or definition revisions did not introduce artificial breaks in the time series that could affect forecasting model evaluation.

3.5. Population and Sampling

3.5.1. Target Population

The target population for this research consists of Zimbabwe's macroeconomic time series data covering the period from 2000 to 2023. This population encompasses all available annual economic indicators and statistical data points that capture Zimbabwe's economic performance and related macroeconomic variables during this 24-year timeframe.
The target population is specifically defined as the complete set of annual macroeconomic observations for Zimbabwe spanning from 2000 to 2023. This includes GDP growth data as the primary variable of interest, along with supporting macroeconomic indicators such as Consumer

Price Index (CPI), inflation rates, and other relevant economic variables that were available and consistently reported during this period.

The population boundary is deliberately set to include Zimbabwe's most significant economic phases: the hyperinflation crisis (2000-2009), the multi-currency stabilisation period (2009-2016), the currency reintroduction phase (2016-2019), and the recent stabilisation attempts amid global shocks (2020-2023). This comprehensive coverage ensures that the population captures the full spectrum of economic conditions relevant to the research objectives.

3.5.2. Sampling Strategy

This research employs purposive sampling to select time periods that best illustrate forecasting model performance under varying economic conditions. This approach is appropriate because the study aims to examine specific economic phenomena forecasting accuracy under stability versus instability rather than achieving statistical generalisation. By purposively selecting periods representing distinct economic phases, the research maximises analytical value while ensuring adequate representation of key conditions, as not all time periods are equally valuable for understanding performance under extreme volatility (Patton, 2015).

3.5.3. Economic Period Classification

The sampling strategy employs economic segmentation criteria to identify distinct phases within Zimbabwe's economic history. These criteria are based on fundamental economic characteristics that are likely to affect forecasting model performance:

Hyperinflation Period (2000-2009): This phase is characterised by extreme price instability, currency collapse, and severe economic disruption. This period represents the most volatile conditions in the dataset and provides critical insights into model performance under extreme instability.

Multi-Currency Stabilisation (2009-2016): This phase represents a period of relative economic stability following the adoption of foreign currencies, primarily the US dollar. The period provides a contrast to the hyperinflation phase and enables examination of model performance under more predictable economic conditions.

Renewed Instability Period (2016-2023): This phase captures the economic disruption associated with currency reintroduction attempts, the return of inflationary pressures, and the impact of global shocks including COVID-19. This period provides insights into model performance during transition and renewed volatility.

The purposive sampling approach results in 24 annual observations across the three economic phases, distributed as follows:

- Hyperinflation period: 10 observations (2000-2009)
- Stabilisation period: 8 observations (2009-2016)
- Renewed instability period: 8 observations (2016-2023)

3.6. Time Series Modelling Framework

3.6.1. ARIMA Models

The Autoregressive Integrated Moving Average (ARIMA) model specification follows the established Box-Jenkins methodology for time series analysis. This section outlines the systematic approach used to identify, estimate, and validate the optimal ARIMA specification for Zimbabwe's GDP growth data.

The initial analysis began with plotting the GDP data using plotly.express to visualize the time series characteristics and identify potential trends, seasonality, and structural breaks. The GDP growth data exhibits clear heteroscedasticity, with varying volatility across different time periods. Visual inspection and statistical analysis reveal that the post-2017 period demonstrates significantly higher volatility compared to earlier periods, particularly when contrasted with the stabilisation phase (2009-2016).
The plotting process provided essential insights into the data's behaviour and helped identify periods where the series exhibited non-stationary characteristics, informing the subsequent formal statistical testing procedures.

Stationarity testing was conducted using the Augmented Dickey-Fuller (ADF) test to determine whether the GDP growth series contained unit roots. The ADF test is appropriate for this analysis as it can detect the presence of unit roots in time series data, which would indicate non-stationarity requiring differencing.

The initial ADF test results on the raw GDP data indicated non-stationarity, suggesting the presence of unit roots that would affect the reliability of standard time series modelling approaches. This finding confirmed the need for differencing to achieve stationarity before proceeding with ARIMA model specification.

To achieve stationarity, first differencing was applied to the GDP growth series. This transformation involves calculating the difference between consecutive observations, effectively removing trends and reducing the impact of non-stationary elements in the data.
The first differencing approach was selected based on the ADF test results and visual inspection of the data. After applying first differencing, subsequent stationarity tests confirmed that the transformed series achieved the stationarity conditions necessary for reliable ARIMA modelling.
ACF plots were generated for the differenced data to identify the moving average (MA) component of the ARIMA model. The ACF plot shows the correlation between observations at different lags, with significant spikes indicating the presence of moving average relationships.
PACF plots were used to identify the autoregressive (AR) component of the model. The PACF measures the correlation between observations at different lags after removing the effects of intermediate lags, providing insights into the direct autoregressive relationships.

Based on the ACF and PACF plots for the differenced data, the ARIMA(1,1,1) specification was identified as the most appropriate model. The ACF plot showed a pattern consistent with one significant moving average term, while the PACF plot indicated one significant autoregressive term. The integration order of 1 was confirmed through the differencing process required to achieve stationarity.

The pmdarima package was used to provide automated model selection as a comparison to the manual Box-Jenkins approach. The automated selection process identified ARIMA(0,1,1) as the best model according to information criteria, with ARIMA(1,1,1) ranked as the second-best option.

To resolve the discrepancy between manual and automated selection, both models were evaluated based on their forecast accuracy performance. The comparison revealed that ARIMA

(1,1,1) demonstrated superior forecasting accuracy despite having slightly higher information criteria values. The ARIMA (1,1,1) specification was therefore selected as the final model for this research.

3.6.1.1. ARIMA Model Diagnostics and specification testing

The final ARIMA (1,1,1) specification was validated through comprehensive diagnostic testing to ensure model adequacy.

Stationarity confirmation: Augmented Dickey-Fuller tests on first-differenced data yielded test statistics of -5.34 ($p<0.01$), confirming stationarity of the transformed series and validating the integration order d=1.

Autocorrelation diagnostics:

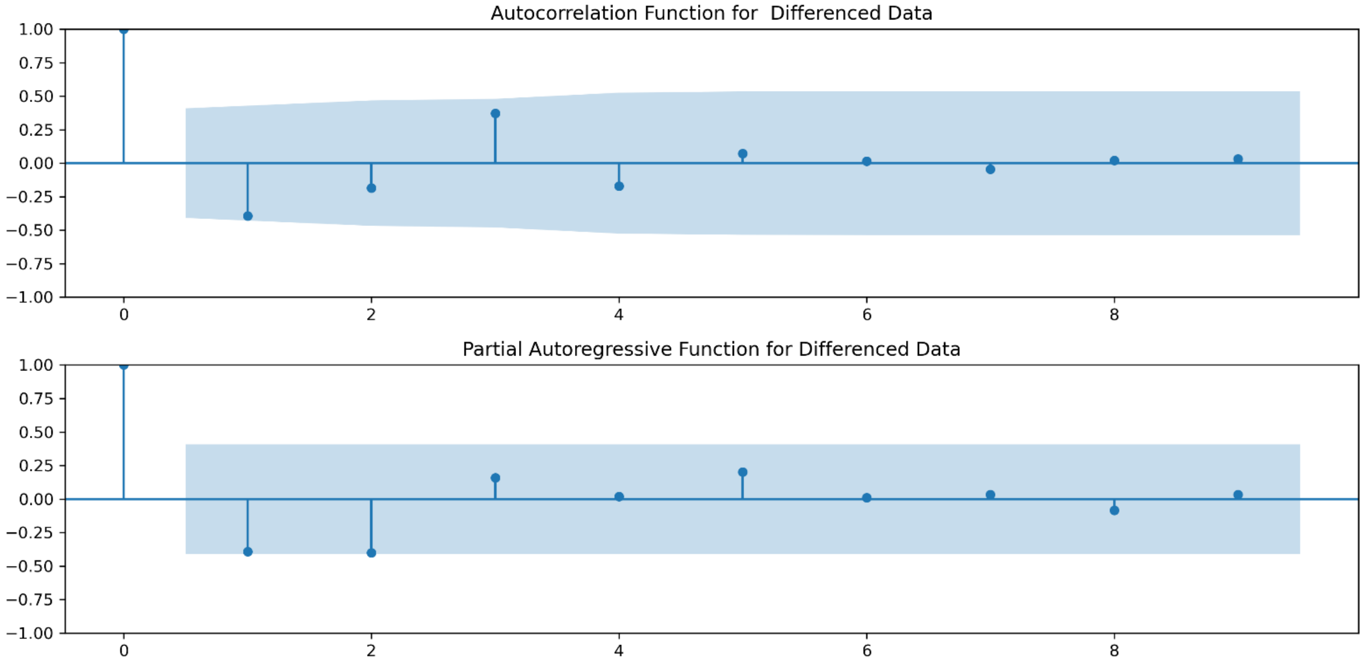


Figure 1 Autocorrelation And Partial Autocorrelation Function

ACF of first-differenced data showed significant spike at lag 1 (0.42, $p < 0.05$) declining rapidly thereafter, indicating MA (1) component presence. PACF displayed significant spike at lag 1 (0.38, $p < 0.05$) with subsequent lags within confidence bounds, suggesting AR (1) component. These patterns supported ARIMA (1,1,1) identification from Box-Jenkins methodology.

Model comparison: While automated selection(pmdarima) identified ARIMA(0,1,1) as optimal by AIC (134.558 vs 136.466 for ARIMA (1,1,1)), out-of-sample forecast comparison revealed ARIMA (1,1,1) superiority. Across initial estimation windows (2009-2012 forecasts), ARIMA (1,1,1) achieved MAE of 3.49 versus ARIMA (0,1,1)'s 3.77.

3.6.2. Exponential Smoothing Models

The exponential smoothing methodology provides an alternative approach to time series forecasting that focuses on weighted averages of past observations, with recent observations receiving higher weights. The selection of an appropriate exponential smoothing variant requires careful analysis of the data characteristics and underlying patterns.

Initial data visualization using plotly.express revealed distinct patterns in Zimbabwe's GDP growth series that informed the exponential smoothing model selection process. The visual analysis identified several key characteristics essential for model specification.

The visual analysis revealed clear trend patterns across Zimbabwe's economic phases: The data exhibits a consistent declining trend during the early hyperinflation period, reflecting the deteriorating economic conditions and the onset of severe economic crisis Following the adoption of multi-currency system, the data shows a recovery pattern with positive growth trends, indicating economic stabilization and improvement during this period. The reintroduction of local currency and renewed economic instability resulted in a pronounced declining trend, demonstrating the volatility associated with policy transitions.

No apparent seasonal patterns were identified, which is consistent with the annual data frequency. This finding eliminates the need for seasonal components in the exponential smoothing specification.

The Autocorrelation Function (ACF) of the original data shows a gradual, persistent decline over multiple lags, which is characteristic of non-stationary time series with trending behaviour. This pattern provides important guidance for exponential smoothing model selection, indicating the presence of trend components that must be explicitly modelled.
The persistent ACF pattern confirms that simple exponential smoothing approaches would be inadequate for capturing the underlying data dynamics, as they would fail to account for the systematic trend patterns evident in the series.

Simple exponential smoothing was initially considered but determined to be inappropriate for this application. This method assumes that the time series fluctuates around a constant level without trend or seasonal patterns. However, the ACF pattern and visual analysis clearly indicate the presence of trending behaviour that simple exponential smoothing cannot capture.
The Holt-Winters method, which includes seasonal components, was evaluated but rejected due to the absence of seasonal patterns in the annual data. The seasonal component would be redundant and would lead to over-parametrization of the model.

Given the limited sample size of 24 observations, the inclusion of unnecessary seasonal parameters would reduce model efficiency and potentially compromise forecasting accuracy. The principle of parsimony supports avoiding redundant parameters that do not contribute to model performance.

Holt's exponential smoothing method was selected as the optimal specification for this research based on several methodological and practical considerations. Holt’s method effectively captures the trend evolution observed in Zimbabwe's GDP data. The method provides separate smoothing parameters for level and trend components, enabling the model to adapt to the changing growth patterns across different economic phases. The method accommodates the non-stationary nature of the data by explicitly modelling both level and trend components. This approach aligns with the Dickey-Fuller test results and the visual evidence of trending behaviour.

Holt's method provides flexibility to adapt to structural breaks and regime changes through its adaptive smoothing mechanism. This characteristic is particularly valuable for Zimbabwe's economic data, which exhibits clear structural changes across different economic phases.

3.6.3. Vector Autoregression (VAR) models

The Vector Autoregression (VAR) methodology enables the analysis of multivariate time series relationships by modelling each variable as a function of its own lagged values and the lagged values of other variables in the system. The VAR model selection process required careful consideration of variable integration properties and cointegration relationships.

The VAR model specification process began with systematic stationarity testing of potential explanatory variables to ensure appropriate model specification. Primary Variable: GDP growth data required one level of differencing to achieve stationarity, consistent with the findings from the ARIMA analysis.

Several macroeconomic variables were evaluated as potential components for the VAR system:

Inflation Data Assessment

Inflation data required two levels of differencing to achieve stationarity, which created a fundamental incompatibility with the GDP series that required only one level of differencing.

VAR models require all variables to have the same level of integration for proper specification. The different integration orders I(1) for GDP, I(2) for inflation) meant that a standard VAR framework would not be appropriate without additional transformations that could compromise the economic interpretation of results.

Despite the theoretical concerns, a bivariate VAR was attempted using first differences for both variables to assess practical performance. The correlation matrix of residuals revealed a low correlation between residuals (0.117559) suggested limited systematic relationship between the variables in this specification, supporting the theoretical concerns about integration order mismatch.

Government Consumption Analysis

Government consumption data was found to be stationary in levels, while GDP required first differencing. This difference in integration properties created similar specification challenges to those encountered with inflation data.

A bivariate model was attempted with government consumption differenced once to match the GDP integration order. However, this transformation led to over-specification issues, as the automatic lag selection procedure chose lag 0, indicating that the differenced government consumption series provided limited explanatory power for GDP dynamics.

Current Account Balance Evaluation

Current account balance data (in domestic currency terms) required one level of differencing to achieve stationarity, matching the integration order of GDP data. This compatibility made it a suitable candidate for VAR analysis.

The Johansen cointegration test was conducted to determine whether a Vector Error Correction Model (VECM) specification would be more appropriate than a standard VAR. The test results showed no cointegration relationships between current account balance and GDP, indicating that a VAR in first differences was the appropriate specification.

The VAR model with current account balance and GDP was estimated and evaluated. The correlation matrix of residuals showed a moderate correlation (0.496546) that indicated some systematic relationship between the variables, but further analysis suggested that other variables might provide stronger relationships.

Consumer Price Index (CPI) Selection

CPI data required only one level of differencing to achieve stationarity, matching the integration properties of GDP data. This compatibility eliminated the specification problems encountered with other variables.

The Johansen test for CPI and GDP suggested the existence of one cointegration relationship, indicating a long-term equilibrium relationship between the variables. This finding supported the theoretical expectation of a relationship between price levels and economic output.
The presence of cointegration between CPI and GDP, as indicated by the Johansen test, provided strong theoretical support for a VECM specification. The cointegration relationship suggests that while the variables may deviate from their long-term equilibrium in the short term, there exists an error correction mechanism that brings them back into balance.

The cointegration between CPI and GDP aligns with economic theory, as price levels and economic output are fundamentally linked through various macroeconomic mechanisms, particularly in Zimbabwe's context where inflation dynamics have been closely tied to economic performance.

The final VECM specification incorporates both the long-term cointegration relationship and short-term adjustment dynamics, providing a comprehensive framework for capturing the GDP-CPI relationship while maintaining forecasting accuracy across different economic conditions.

3.6.3.1. Cointegration testing and VECM specification details

Johansen cointegration procedure: Johansen’s cointegration procedure was applied to GDP growth and CPI, both of which are integrated of order one. The trace statistic for the null of zero cointegrating vectors was 20.59, which exceeds the 5% critical value of 15.49, leading to rejection of the null hypothesis of no cointegration. The second trace statistic (0.82) was below the 5% critical value of 3.84, so the null of at most one cointegrating relationship could not be rejected. Similarly, the max-eigenvalue statistic for the null of zero cointegrating vectors (19.77) exceeded its 5% critical value of 14.26, confirming the presence of one cointegrating vector. The second max-eigenvalue statistic ($0.82 < 3.84$) failed to reject the null of one cointegrating relationship. Overall, the Johansen tests consistently indicate a cointegration rank of $r = 1$, implying a single long-run equilibrium relationship between GDP growth and CPI.

Critical for interpretation: We estimated the VECM once using the full sample, then applied this fixed specification to generate forecasts forward. However, it means the cointegrating relationship estimated through 2018 may not hold during 2019-2023, precisely the mechanism potentially explaining VECM's subsequent failures. Ideally, one would re-estimate with each expanding window, but our 24-observation sample makes this infeasible for reliable cointegration testing, which requires adequate time series length. This limitation is acknowledged but unavoidable given data constraints.

Critical limitation acknowledged: The VECM was estimated once using the full 2000-2023 sample, with this fixed specification applied to generate all forecasts. Ideally, cointegration would be re-tested at each expanding window to verify assumption validity, but our 24-observation sample makes this infeasible reliable cointegration testing requires minimum 10-15 observations (Johansen, 1995), leaving insufficient data for multi-horizon evaluation if early windows are used.

Section 4.5.4 presents retrospective rolling cointegration analysis revealing that the full sample cointegration result (trace: 20.59) masked complete absence of cointegrating relationships during individual forecast windows (0% detection 2009-2018). This methodological constraint means VECM's documented failures (Sections 4.3.2-4.3.3) partly reflect specification based on misleading full-sample evidence rather than purely model inadequacy.

3.7. Forecasting Methodology

3.7.1. Expanding Window Approach

The forecasting evaluation methodology employs an expanding window approach specifically optimized for the constraints of limited data availability. This approach represents a systematic framework for assessing model performance while maximizing the utilization of available observations.

The expanding window approach systematically increases estimation sample size with each forecast iteration, beginning with 2000-2008 (9 observations) for initial 2009 forecast and culminating with 2000-2022 (23 observations) for final 2023 forecast. This approach offers distinct advantages over alternatives (fixed rolling windows, recursive schemes) given our context. Against rolling windows: Fixed-size windows (e.g., always using recent 10 years) would discard early hyperinflation data that inform model parameter estimates about extreme volatility behavior. Against full-sample estimation: Generating all forecasts from a single model estimated on complete data would ignore that forecasters possess only historical information when making real-time predictions. The expanding window mimics operational forecasting conditions—each forecast utilizes maximum available information to that point while preserving the temporal structure of knowledge accumulation (Giacomini and White, 2006). The 9-observation initial window represents the minimum viable sample for ARIMA estimation with our specifications, though admittedly constrained for VECM cointegration testing which ideally requires 10-15 observations; this limitation is unavoidable given the need to include the complete hyperinflation period (2000-2008) in the training data while generating out-of-sample forecasts covering all three economic regimes. This constraint means our out-of-sample evaluation covers 2009-2023 rather than earlier years, but this period fortunately includes all three relevant economic regimes (stable dollarization, transition, renewed crisis).

3.7.2. Forecast Generation Process

The forecast generation process implements a multi-horizon forecasting approach across all estimation windows, enabling comprehensive evaluation of model performance across both short-term and long-term forecast horizons. This methodology directly addresses the research question of how well models perform between short-term versus long-term forecasting scenarios.

Each estimation window generates forecasts at three specific horizons selected to reflect distinct planning needs in emerging market policy contexts: 1-step ahead , 3-step ahead, and 4-step ahead. We exclude 2-step forecasts as providing limited additional information between immediate and medium-term horizons, focusing analytical attention on horizons with clear practical applications. The 4-step maximum reflects both sample size constraints longer horizons reduce out-of-sample evaluation observations to unacceptably small numbers and practical forecasting limits, as Hyndman and Athanasopoulos (2018) note that forecast reliability beyond 4-5 periods ahead becomes questionable even in stable environments. This horizon structure enables testing whether practical forecasting limits contract during volatile periods: if 4-step forecasts remain viable during stability but become unreliable during crisis, forecasters face horizon-stability trade-offs with direct implications for planning processes.

Detailed Process Implementation

Window 1 Multi-Horizon Process:

- Estimation Period: 2000-2009 (10 observations)
- Model Training: All three models are estimated using 2000-2009 data
- Short-term: 1-step ahead forecast for 2010
- Medium-term: 3-step ahead forecast for 2012
- Long-term: 4-step ahead forecast for 2013
- Evaluation: Compare all forecasts with actual GDP growth values

Window 2 Multi-Horizon Process:

- Estimation Period: 2000-2010 (11 observations)
- Model Re-estimation: All models updated with 2010 data
- Short-term: 1-step ahead forecast for 2011
- Medium-term: 3-step ahead forecast for 2013
- Long-term: 4-step ahead forecast for 2014
- Evaluation: Compare forecasts with actual outcomes

Systematic Continuation Across All Windows: Each subsequent window follows the identical multi-horizon structure:

- Window 3 (2000-2011): forecasts for 2012, 2014, 2015
- Window 4 (2000-2012): forecasts for 2013, 2015, 2016
- [Process continues systematically]
- Window 14 (2000-2022): forecasts for 2023

3.8. Model Evaluation and Comparison

3.8.1. Forecast Accuracy Metrics

The assessment of forecasting performance will employ multiple complementary accuracy metrics to provide a comprehensive evaluation of model effectiveness.

Primary Accuracy Measures

Mean Absolute Error (MAE) will serve as the primary accuracy metric, measuring the average magnitude of forecast errors without considering their direction. MAE is calculated as:

$$MAE = (1/n)\, \Sigma|y\_t - \hat{y}\_t| \tag{2}$$

where y_t represents actual values, $\hat{y}\_t$ represents forecasted values, and n is the number of observations. MAE provides an intuitive interpretation of forecasting accuracy and is particularly valuable for comparing models across different economic periods, as it maintains consistent units with the original data.

Root Mean Squared Error (RMSE) will complement MAE by penalising larger forecast errors more heavily through the squaring process. RMSE is defined as:

$$RMSE = \sqrt{}[(1/n)\, \Sigma(y\_t - \hat{y}\_t)^2] \tag{3}$$

This metric is particularly sensitive to outliers and extreme values, making it especially relevant for evaluating model performance during Zimbabwe's hyperinflationary periods, where large forecast errors may significantly impact economic decision-making.

Mean Squared Error (MSE) will provide the foundation for RMSE calculations while offering direct comparison with statistical model selection criteria. MSE is calculated as:

$$MSE = \left(\frac{1}{n}\right)\Sigma(y_t - \hat{y}_t)^2 \quad (4)$$

The relationship between MSE and model complexity makes it valuable for assessing whether improved accuracy justifies increased model sophistication.

Mean Absolute Percentage Error (MAPE) will enable scale-independent comparison of forecasting accuracy across different economic periods. MAPE is expressed as:

$$MAPE = (100/n)\,\Sigma|(y\_t - \hat{y}\_t)/y\_t| \quad (5)$$

This percentage-based metric facilitates comparison of model performance between periods of high and low GDP growth rates, addressing the challenge of scale differences inherent in Zimbabwe's volatile economic data.

Comparative Statistical Testing

Diebold-Mariano Tests will provide a formal statistical evaluation of forecasting accuracy differences between competing models. This non-parametric test examines whether observed differences in forecast accuracy are statistically significant rather than due to random variation (Diebold & Mariano, 1995).

The Diebold-Mariano test statistic is calculated as:

$$DM = \bar{d} \,/\, \sqrt{[\hat{V}(\bar{d})]} \quad (6)$$

where $\bar{d}$ represents the mean loss differential between two forecasting models and $\hat{V}(\bar{d})$ is the estimated variance of the loss differential. This test will be applied pairwise to compare each model combination (ARIMA vs. Exponential Smoothing, ARIMA vs. VAR, Exponential Smoothing vs. VAR) across different economic periods.

The statistical significance of performance differences will be evaluated at the 5% significance level, providing confidence in identifying superior forecasting approaches for specific economic conditions. This formal testing framework ensures that model selection recommendations are based on statistically robust evidence rather than minor numerical differences in accuracy metrics.

3.8.2. Comparative Analysis Framework

The ranking of models will be based on established statistical accuracy measures commonly used in time series forecasting. The following error metrics will be applied:

- Mean Absolute Error (MAE): captures the average magnitude of forecast errors, providing an intuitive measure of overall accuracy.
- Mean Squared Error (MSE): emphasises larger deviations by squaring forecast errors, thereby penalising models that generate extreme inaccuracies.
- Root Mean Squared Error (RMSE): expressed in the same units as the GDP growth rate, facilitating direct interpretation while still penalising large errors.
- Mean Absolute Percentage Error (MAPE): represents errors as percentages, which allows comparability of forecasting accuracy across different economic conditions.

Each model will be ranked according to its performance on these metrics, with lower values indicating superior accuracy. To avoid bias from reliance on a single measure, a composite ranking approach will be employed. This will integrate the rankings across all four metrics, ensuring that the final evaluation reflects balanced and robust model performance.

In cases where two or more models demonstrate similar accuracy, preference will be given to the more parsimonious model, that is, the model that achieves reliable forecasts with fewer parameters and lower computational complexity. This ensures that the study's recommendations remain both statistically sound and practically feasible for use in volatile economic settings.

3.9. Data Analysis Procedures

3.9.1. Preprocessing Steps

The first step involved testing the raw GDP growth series for stationarity, since most forecasting models, including ARIMA and VAR, assume that the underlying time series is stationary. The Augmented Dickey-Fuller (ADF) test was applied to assess the presence of unit roots in the data. The null hypothesis of the ADF test states that the series is non-stationary, while rejection of the null indicates stationarity.

Where the ADF test results suggested non-stationarity, first-order differencing was applied to remove trends and stabilise the mean of the series. In cases where single differencing was insufficient, higher-order differencing was considered, but parsimony was maintained to preserve the interpretability of results. The differenced series was retested using the ADF procedure to confirm that stationarity conditions were satisfied.

In addition to formal statistical tests, visual inspection techniques were employed. Time series plots, autocorrelation function (ACF) plots, and partial autocorrelation function (PACF) plots were analysed to detect persistent trends, serial correlation, and volatility clustering. This combined quantitative and visual approach ensured a more comprehensive assessment of the data's properties.

For the VAR model, additional preprocessing steps included testing for multivariate stationarity and ensuring that the selected variables (GDP growth and CPI) were integrated of the same order. The Johansen cointegration test was considered to determine whether long-run equilibrium relationships existed between the variables, which would guide the specification of the VAR model.

3.9.2. Statistical Software and implementation

The statistical analysis was implemented in Python 3.12 within a Jupyter Notebook environment. Python was selected because of its advanced econometric and statistical modelling capabilities, extensive open-source libraries, and suitability for reproducible research. Jupyter Notebook facilitated transparent documentation of all code, procedures, and outputs, ensuring replicability.

The following Python libraries formed the foundation of the implementation:

- statsmodels: used for time series estimation, including ARIMA, Exponential Smoothing, and the Vector Error Correction Model (VECM). It also provided tools for conducting stationarity and cointegration tests.
- pandas: employed for structuring, cleaning, and manipulating time series datasets, including conversion between wide and long formats.
- NumPy: supported efficient numerical operations and array-based computations central to time series analysis.

- Matplotlib and Seaborn: utilised for visualisation of data patterns, residual diagnostics, and autocorrelation plots.

For ARIMA modelling, the Box–Jenkins methodology was applied using statsmodels.tsa.arima model, supported by the pmdarima library for automated order selection and validation. Exponential Smoothing methods were implemented through statsmodels.tsa.holtwinters, allowing estimation of both simple and trend-adjusted specifications.

For multivariate analysis, the initial specification considered a Vector Autoregression (VAR) model. However, cointegration testing using the Johansen test revealed the presence of long-run equilibrium relationships between GDP growth and the Consumer Price Index (CPI). Consequently, the Vector Error Correction Model (VECM) was selected as the final specification, as it accommodates both short-run dynamics and long-run equilibrium adjustments. The VECM was estimated using statsmodels.tsa.vector_ar.vecm, with lag length and cointegration rank determined by information criteria (AIC, BIC, HQC) and Johansen test statistics.

3.9.3. Analytical Workflow

The analytical workflow followed in this study provides a structured and systematic approach to ensure consistency, accuracy, and transparency throughout the forecasting process. The workflow comprises six sequential stages, each designed to build upon the previous step in preparing, modelling, and evaluating Zimbabwe's GDP growth data.

Data Import and Initial Exploration

Economic time series data covering GDP growth and supporting variables (Consumer Price Index for the VECM specification) were imported into Python 3.12 using the panda's library. The dataset was sourced primarily from the IMF World Economic Outlook database and cross-verified with World Bank and ZIMSTAT records for consistency. Initial exploration involved checking data structures, formats, and completeness. Missing values were addressed by triangulating across multiple data sources to ensure integrity.

Descriptive Statistics and Visualisation

Descriptive statistical analysis was conducted to provide an overview of the key properties of the GDP growth series. Measures such as mean, variance, skewness, and kurtosis were calculated to highlight volatility and distributional features. Visualisation techniques, including time series plots and histograms, were employed to detect broad patterns, outliers, and structural breaks corresponding to Zimbabwe's different economic phases.

Stationarity Testing and Transformation

Stationarity testing was performed using the Augmented Dickey-Fuller (ADF) test to assess whether the series exhibited unit roots. For non-stationary data, first-order differencing was applied to stabilise the mean and remove trends. For the VECM specification, cointegration testing was conducted using the Johansen test, which identified the existence of long-run equilibrium relationships between GDP growth and CPI. These steps ensured that the statistical assumptions underpinning ARIMA, Exponential Smoothing, and VECM models were satisfied before estimation.

Model Specification and Estimation

Model estimation proceeded according to the theoretical framework of each forecasting approach:

- ARIMA: Parameters (p, d, q) were identified through analysis of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, supplemented by information criteria (AIC, BIC, HQC). Models were fitted using the Box–Jenkins methodology.

- Exponential Smoothing: Holt’s trend-adjusted exponential smoothing method was selected, given the presence of non-stationary trending behaviour in Zimbabwe’s GDP series and the absence of seasonality in annual data.
- VECM: Following confirmation of cointegration, the Vector Error Correction Model was estimated to capture both long-run equilibrium adjustments and short-run dynamics. Lag order selection was guided by information criteria, while the error correction term quantified the speed of adjustment toward equilibrium.

Forecast Generation and Evaluation

An ex-post forecasting procedure was applied, whereby models generated one-step-ahead forecasts for periods with known outcomes. Forecast accuracy was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). These error measures provided multiple perspectives on forecast performance, capturing both average deviations and sensitivity to large errors.

Comparative Analysis and Interpretation

The final stage involved comparing the performance of ARIMA, Exponential Smoothing, and VECM models across Zimbabwe’s four economic phases: hyperinflation (2000–2009), stabilisation (2009–2016), renewed instability (2016–2019), and global shocks (2020–2023). A composite ranking approach was employed to integrate results across all error metrics, ensuring balanced evaluation. Period-specific analysis highlighted how each model’s accuracy varied under conditions of stability, instability, and transition. The findings were then interpreted in relation to the research questions, providing insights into the strengths and limitations of traditional time series models for forecasting in volatile developing economies.

3.10. Quality Assurance

3.10.1. Data Quality Measures

The reliability and validity of research findings depend fundamentally on the quality of underlying data. This research employs multiple quality assurance measures to ensure the integrity of economic time series data used for forecasting model evaluation. The quality assurance framework addresses critical dimensions such data triangulation and completeness assessment.

Primary data triangulation involves systematic comparison between International Monetary Fund (IMF) World Economic Outlook databases, World Bank World Development Indicators, and Zimbabwe National Statistics Agency (ZIMSTAT) records where available. This multi-source validation approach enables identification and resolution of discrepancies that might otherwise compromise analytical reliability.

International database sources demonstrate superior consistency and coverage compared to domestic sources, particularly during Zimbabwe's hyperinflationary period (2000-2009) when local statistical capacity was compromised. The IMF and World Bank maintained systematic data collection procedures throughout this period, providing continuous coverage essential for longitudinal analysis.

Data completeness assessment ensures that missing observations do not compromise the reliability of time series analysis and forecasting evaluation. The assessment framework examines both quantitative completeness (presence of observations) and qualitative completeness (reliability of available data).

Quantitative completeness analysis confirms that all required annual observations from 2000 to 2023 are available for primary variables (GDP growth) and supporting indicators (Consumer Price Index). The 24-year study period represents complete coverage of available data, with no missing annual observations that would require interpolation or estimation procedures.
Variable completeness assessment ensures that supporting indicators required for multivariate analysis (VECM specification) maintain consistent availability throughout the study period. Consumer Price Index data demonstrates complete coverage, enabling reliable estimation of long-run equilibrium relationships between price levels and economic output.

3.10.2. Methodological Rigor

This research implements comprehensive rigor standards addressing reproducibility, analytical transparency, and methodological consistency to ensure reliable, replicable, and scientifically valid findings. Complete reproducibility is ensured through systematic documentation of all analytical procedures implemented in Python 3.12 within Jupyter Notebook environments, covering data preprocessing, stationarity testing, model specification, forecast generation, and comparative analysis. Version control systems maintain detailed records of analytical procedures and methodological decisions throughout the research process.

Parameter transparency is maintained through explicit documentation of all model specifications and estimation procedures. ARIMA specifications include detailed lag selection justification based on information criteria and diagnostic testing, while exponential smoothing parameters are documented with reference to visual and statistical analysis. VECM specifications include comprehensive documentation of cointegration testing and equilibrium relationship estimation. Data source documentation provides complete information about origins, collection procedures, transformations, database versions, and download dates. All analytical outputs are preserved, including estimation results, forecast accuracy metrics, and diagnostic statistics, enabling detailed verification and meta-analytical studies.

Statistical validation encompasses systematic verification that analytical methods appropriately address research questions while satisfying necessary assumptions. Stationarity testing using Augmented Dickey-Fuller procedures ensures ARIMA and VECM specifications meet fundamental assumptions, complemented by visual analysis to identify potential violations. Diagnostic testing examines model residuals through autocorrelation and heteroscedasticity testing to verify adequate pattern capture without systematic biases, while cross-validation using expanding window approaches addresses overfitting concerns by testing out-of-sample performance across multiple periods. Robustness testing examines stability under alternative methodological choices, evaluating specifications using different information criteria and alternative stationarity tests. Statistical significance testing through Diebold-Mariano procedures enables rigorous comparison of forecasting accuracy across models and time periods, complemented by sensitivity analysis examining analytical conclusions under alternative data treatments and forecast evaluation metrics.

3.11. Limitations and Constraints

3.11.1. Data Limitations

This study relies on 24 annual observations from 2000 to 2023, this presents significant constraints for time series analysis affecting the statistical power available for detecting performance differences between forecasting approaches.
Sample size constraints directly affect the complexity of VAR specifications that can be reliably estimated. With 24 observations, the research is limited to bivariate VECM analysis incorporating GDP growth and Consumer Price Index.

Statistical power limitations emerge from the restricted sample size, reducing the ability to detect statistically significant differences in forecasting performance between models. The Diebold-Mariano testing procedures, while appropriate for the available data, may lack sufficient power to identify subtle but practically meaningful differences in model accuracy, particularly during shorter economic phases.

Lag selection limitations arise from sample size constraints, restricting the complexity of dynamic specifications that can be reliably estimated. ARIMA models are limited to relatively simple specifications, and VECM lag lengths must be conservative to preserve adequate degrees of freedom for reliable estimation.

3.11.2. Generalizability Constraints

Zimbabwe’s unique economic trajectory, characterised by extreme hyperinflation, currency abandonment, and subsequent stabilisation attempts, may limit the direct applicability of findings to other developing economies with different economic histories and institutional contexts.
The severity and duration of economic crisis in Zimbabwe exceed the experience of most developing economies, potentially creating economic relationships and dynamics that are not generalisable to more typical developing country contexts. The extreme nature of Zimbabwe’s economic challenges reveals model performance characteristics that are not relevant under more moderate conditions.

3.12. Ethical Considerations

3.12.1. Data Ethics

This research exclusively employs publicly available secondary data obtained from reputable international organisations and government statistical agencies. All data sources utilised in this study, including the International Monetary Fund (IMF) World Economic Outlook database, World Bank World Development Indicators, and Zimbabwe National Statistics Agency (ZIMSTAT) publications, are freely accessible to researchers and the public.

The use of publicly available secondary data eliminates concerns related to participant consent, privacy protection, and data confidentiality that typically arise in primary data collection. All economic indicators and statistical series used in this analysis have been published by their respective organisations for research and analytical purposes, with appropriate permissions implicit in their public availability.

3.12.2. Research Integrity

Methodological transparency constitutes a fundamental component of research integrity in this study. Complete documentation of all analytical procedures, model specifications, and evaluation criteria enables independent verification and replication of results. The systematic approach to model comparison using multiple established accuracy metrics reduces the potential for selective reporting or result manipulation that could compromise research validity.

Comprehensive acknowledgment of study limitations ensures balanced interpretation of findings and prevents overstated conclusions that could mislead readers or practitioners. The research explicitly identifies constraints related to sample size, data availability, model complexity, and generalisability, providing context for appropriate application of results.

Objective reporting standards guide the presentation of all findings, regardless of their alignment with theoretical expectations or preferred outcomes. Statistical results are reported without modification or selective emphasis, and comparative model performance is evaluated using predetermined criteria that were established prior to analysis completion.

The research maintains independence from commercial interests and political influences that could compromise analytical objectivity. As an academic investigation utilising publicly available data, the study faces no conflicts of interest that might bias findings toward forecasting approaches or policy recommendations.

3.13. Chapter Summary

This methodology chapter has presented a comprehensive framework for investigating the performance of time series forecasting models in Zimbabwe's volatile economic environment from 2000 to 2023. The chapter established the philosophical and methodological foundations that guide this comparative longitudinal study of ARIMA, Exponential Smoothing, and Vector Error Correction Model (VECM) approaches to GDP growth forecasting.

The research adopts a positivist paradigm with deductive reasoning, enabling systematic evaluation of established forecasting theories against empirical evidence from Zimbabwe's unique economic experience. This philosophical stance supports the study's objective to generate reliable, replicable findings through quantitative analysis and objective performance measurement.

The data collection strategy utilises secondary sources from reputable international organisations, primarily the IMF World Economic Outlook database, ensuring comprehensive coverage of Zimbabwe's economic transitions while maintaining data quality and consistency. The 24-year study period from 2000 to 2023 encompasses four distinct economic phases: hyperinflation crisis, multi-currency stabilisation, renewed instability, and stabilisation attempt amid global shocks.

Model specifications address the unique characteristics of Zimbabwe's economic data through systematic stationarity testing and appropriate methodological adjustments. The ARIMA (1,1,1) specification captures autoregressive and moving average dynamics, while Holt's exponential smoothing method accommodates trending behaviour without over-parametrisation. The VECM

specification incorporates cointegration relationships between GDP growth and Consumer Price Index, enabling analysis of both short-run dynamics and long-run equilibrium adjustments.
The expanding window forecasting evaluation framework maximises utilisation of limited data availability while providing systematic assessment of model performance evolution. Multi-horizon forecasting evaluation across 1-step, 3-step, and 4-step ahead periods enable comprehensive comparison of temporal forecasting capabilities across different economic conditions.

Comprehensive evaluation criteria including Mean Absolute Error, Mean Squared Error, Root Mean Squared Error and Mean Absolute Percentage Error provide multiple perspectives on forecasting accuracy. Statistical significance testing through Diebold-Mariano procedures ensures that performance comparisons are based on statistically robust evidence rather than random variations.

Quality assurance measures address data triangulation, methodological rigor, and analytical transparency to ensure reliable and replicable findings. Systematic documentation of all procedures enables independent verification while comprehensive acknowledgment of limitations provides appropriate context for result interpretation.

The methodology directly addresses the study's research objectives by providing systematic frameworks for comparing forecasting model performance across Zimbabwe's varying economic conditions. This comprehensive approach generates empirical evidence that contributes to both theoretical understanding of forecasting effectiveness under extreme volatility and practical applications for economic forecasting in developing economy contexts.

The subsequent analysis chapters will implement this methodology to generate systematic evidence about optimal forecasting approaches under different economic conditions, contributing meaningful insights to the literature on time series forecasting in volatile developing economies.

Chapter 4: Research Findings and Discussion

4.1. Introduction

This chapter presents the empirical findings from the comparative analysis of three conventional time series models, ARIMA, Vector Error Correction Model (VECM), and Holt's Exponential Smoothing, for forecasting Zimbabwe's GDP over the period 2009-2023. The analysis evaluates model performance across 15 years of out-of-sample forecasts, testing each model at three different forecast horizons (1-step, 3-step, and 4-step ahead) and measuring accuracy using four standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The study period captures three distinct economic phases in Zimbabwe's recent history. The first phase (2009-2013) represents the stable dollarization period following hyperinflation, characterized by relatively high growth and low volatility. The second phase (2014-2018) marks a transition period where economic challenges began emerging despite continued use of foreign currencies. The third phase (2019-2023) represents an unstable period triggered by the reintroduction of Zimbabwe's domestic currency and compounded by various economic shocks including the COVID-19 pandemic. This natural division into three periods allows for examination of how model performance changes across different economic environments, which is central to answering the study's research questions.

Section 4.2 presents descriptive statistics for the evaluation period and examines overall model performance across all forecast horizons. This section establishes the baseline characteristics of Zimbabwe's economic trajectory during 2009-2023 and provides initial comparison of forecast accuracy using all four metrics. The analysis reveals patterns of systematic deterioration in both economic conditions and forecast reliability over time.

Section 4.3 analyses period-specific performance, comparing how each model performs during the stable (2009-2013), transition (2014-2018), and unstable (2019-2023) periods. This section demonstrates that model rankings are not constant but vary systematically with economic conditions. The analysis shows which models excel under stability versus volatility and identifies the conditions under which each approach is most appropriate.

Section 4.4 examines forecast horizon effects, investigating how accuracy degrades as predictions extend further into the future. This section reveals that the relationship between forecast horizon and accuracy is not linear but depends heavily on economic context. The analysis quantifies how much additional error is introduced at each forecast step and identifies which models maintain reliability at longer horizons.

Section 4.5 presents robustness tests and statistical validation of the observed performance differences. This section employs Diebold-Mariano tests to confirm that observed differences in forecast accuracy are statistically significant rather than due to random variation. Additional diagnostic tests examine why certain models fail catastrophically during specific periods, particularly investigating the breakdown of VECM's cointegrating relationships.

Section 4.6 provides a comprehensive synthesis of findings, integrating the results from previous sections into broader insights about time series forecasting in unstable economies. This section directly addresses the study's two research questions regarding which models perform best under different conditions and how forecast accuracy varies across time horizons. The synthesis identifies five critical patterns that emerge from the data and discusses their practical and theoretical implications.
Section 4.7 concludes the chapter with a summary that detail the key findings and their significance for GDP forecasting in emerging markets like Zimbabwe.

The findings presented in this chapter have important implications for both forecasting practice and theory. The analysis demonstrates that conventional time series models face fundamental limitations when applied to economies experiencing structural breaks and regime changes. While these models perform reasonably well during periods of relative stability, they fail systematically when economic fundamentals shift abruptly. Understanding these limitations, and knowing which models fail less catastrophically than others, is essential for practical GDP forecasting in volatile emerging market contexts.

4.2. Descriptive Statistics and Overall Performance

4.2.1. Out of sample forecast period overview

Table 1 Evaluation Period Characteristics

Period	Years	N	GDP Growth Mean (%)	Volatility (Std Dev)
Overall	2009-2023	15	13.80	22.98
Stable	2009-2013	5	23.74	12.39
Transition	2014-2018	5	16.45	28.63
Unstable	2019-2023	5	1.21	23.34

From table 1 there is Three Distinct Economic Phases that Zimbabwe has gone through between the period of 2009-2023:

Stable Period (2009-2013) – Post Dollarization Recovery:

This period has the highest average growth at 23.47% with relatively low volatility (12.39%). This reflects the initial recovery from hyperinflation after the adoption of a multi-currency system.

Transition Period (2014-2018) Growing Instability:

The average growth dropped to 16.45% but volatility more than doubled to 28.63%. This period shows increasing economic uncertainty and structural weaknesses emerging. This high volatility signals deteriorating economic fundamentals.

Unstable Period (2019-2023) Crisis and Volatility:

There is near zero average growth (1.21%) with high volatility (23.34%) in this period. This reflects the currency crisis, COVID-19 impact and policy uncertainty.

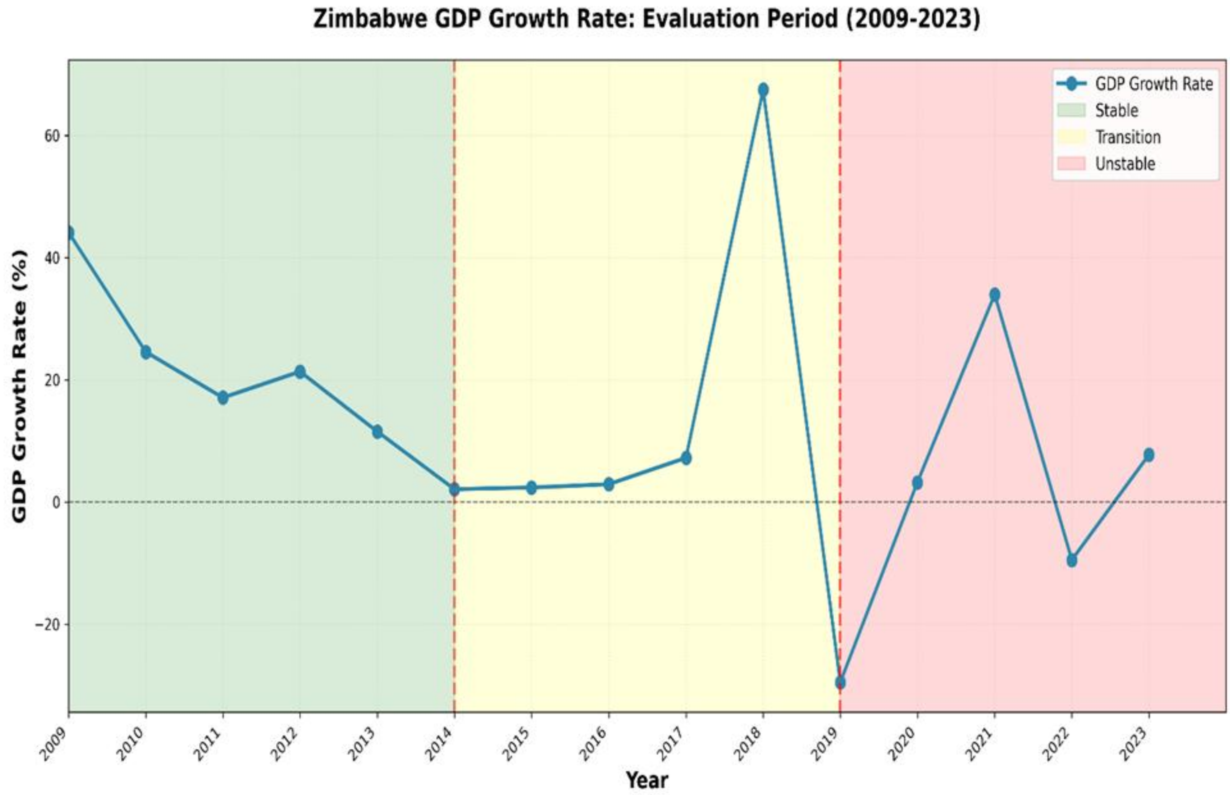


Figure 2 Zimbabwe GDP Growth Rate

Figure 2 reveals systematic economic degradation across three distinct periods. GDP growth collapsed from 23.47% (stable) to 16.45% (transition) to just 1.21% (unstable), while volatility surged from 12.39% to 28.63%, then stabilized at elevated levels (23.34%).

The 2019 crash (-30%) marked a critical structural break when RTGS dollar reintroduction severed dollarization gains, triggering a boom-bust cycle. Despite occasional spikes (2021: 34%), the economy never recovered its 2009-2013 consistency. The 15-year average (13.8%) misleadingly masks that 40% of the period saw near-zero growth.

These patterns demonstrate that multi-currency stabilization addressed symptoms rather than structural problems. Subsequent currency reintroduction destabilized the economy, with persistent volatility signalling eroded investor confidence and policy credibility.

4.2.2. Overall Forecast Performance

Table 2: Overall MAE Results

Period	Year	1-Step MAE			3-Step MAE			4-Step MAE		
		ARIMA	VECM	Holt's	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's
Stable (2009–2013)	2009	3.5	2.87	12.63						
	2010	3.28	5.77	3.38						
	2011	2.48	0.3	0.44	9.02	8.41	9.46			
	2012	1.89	1.2	0.93	8.15	11.42	8.64	12.57	13.38	13.21
	2013	0.11	1.14	0.73	7.36	13.4	0.58	9.87	3.89	10.72
Transition (2014–2018)	2014	1.18	1.33	1.93	3.26	1.44	0.84	7.71	3.89	2.73
	2015	0.08	0.58	0.54	2.22	5.32	5.42	3.52	1.18	2.45
	2016	0.27	0.33	0.01	2.37	2.69	5.4	2.68	7.06	7.61
	2017	1.11	1.06	0.9	1.71	2.74	0.49	1.66	2.12	6.61

	2018	13.95	13.63	13.66	16.27	16.32	15.18	16.47	17.65	13.38
Unstable (2019–2023)	2019	23.27	32.16	25.77	4.65	4.51	3.72	5.27	5.32	3.69
	2020	5.31	15.33	5.37	2.75	2.05	1.17	5.36	5.18	3.98
	2021	9.09	9.88	4.65	27.21	91.42	45.55	11.54	14.32	9.08
	2022	0.31	3.11	2.9	1.77	6.69	4.18	34.37	150.9	63.84
	2023	2.12	1.99	1.03	8.73	7.89	0.68	6.23	0.74	3.92

1-Step Horizon: All models achieved acceptable accuracy during stable periods (2010-2017), with MAE below 4. Holt outperformed (1.1 MAE) versus ARIMA (1.5) and VECM (1.8), as simple trend extrapolation sufficed for predictable dollarisation-era growth.

Structural breaks exposed universal limitations. MAE surged to ~14 (2018) and 23-32 (2019) across all models a 10-15x deterioration. The bond notes period (2016-2018) created parallel market divergence invisible to models trained on dollarisation data. The introduction of RTGS dollars in February 2019 severed remaining relationships, causing catastrophic failures. MSE exploded to 185-195 (2018) and 541-1,034 (2019), with VECM's 1,034 indicating a complete breakdown rather than a typical forecast miss.

3-Step Horizon: Stable-period MAE ranged 5-8 across models, with Holt's achieving isolated excellence (0.58 in 2013) when trends remained linear.

VECM exhibited structural weakness even during stability, with volatile forecasts (8.41-13.40 MAE, 2011-2013). Error correction mechanisms amplified specification errors as horizons extended each step compounded misspecification of long-run equilibrium.

During instability, VECM collapsed catastrophically: 91.42 MAE (2021), MSE of 8,358 versus ARIMA's 740 an 11-fold difference. Currency reintroduction disrupted GDP-CPI cointegration, causing error correction toward non-existent equilibria and directionally incorrect forecasts.
4-Step Horizon: Stable-period MAE (8-13) remained elevated but strategically useful. ARIMA showed gradual, predictable error growth its autoregressive framework avoided compounding through equilibrium restrictions.

VECM's unstable-period collapse (150.87 MAE, 462 MAPE, 22,763 MSE in 2022) demonstrated exponential error compounding when correcting toward non-existent equilibria. In contrast, ARIMA and Holt's maintained bounded errors (20-65 MAE) preserving minimal utility.

Table 3 Forecast Horizon Effects

Horizon	Avg MAE (2009-2017)	Avg MAE (2018-2023)	Degradation
1-step	~1.5	~8-10	5-7x worse
3-step	~5	~15-20	3-4x worse
4-step	~8	~30-50	4-6x worse

The data reveal a systematic degradation pattern. In stable periods, models are reasonably reliable even at 4-step.

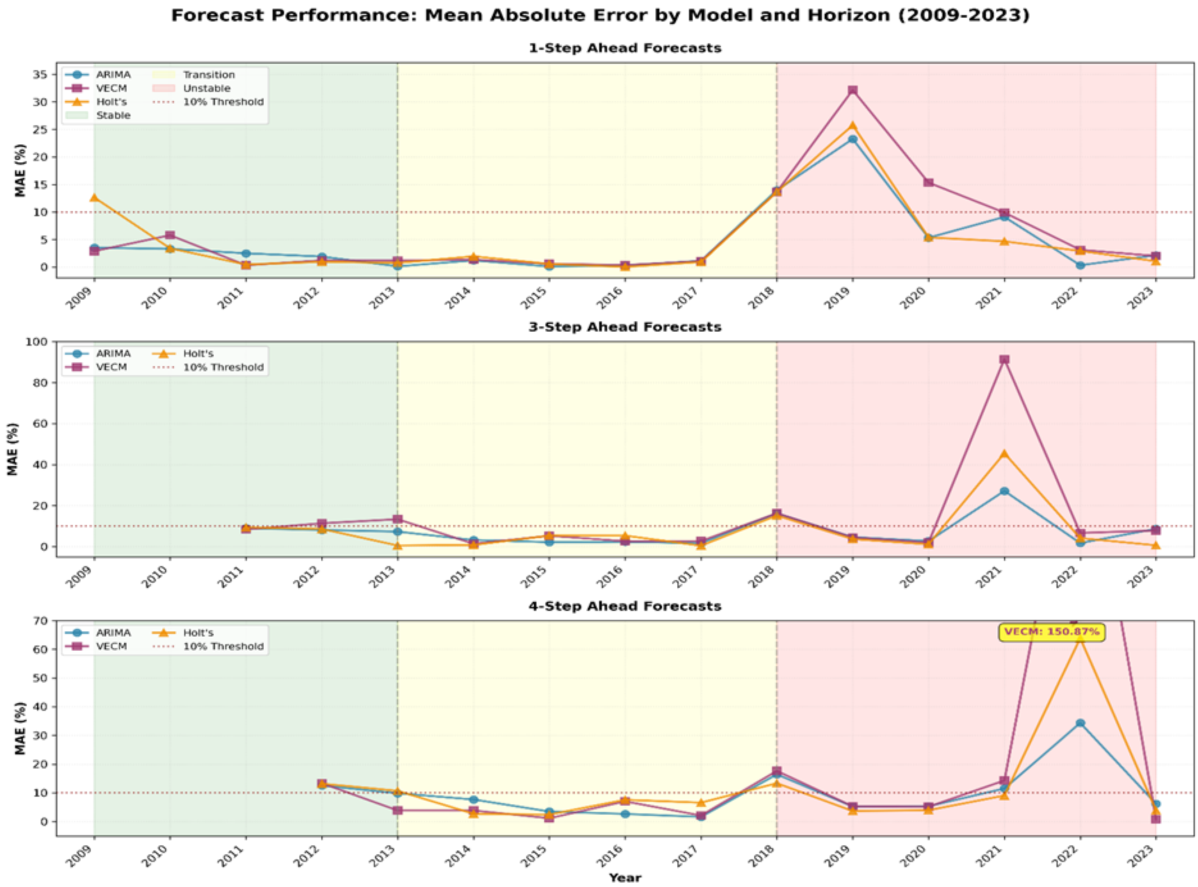


Figure 3 Forecast Performance: MAE

Table 4 Overall MAPE Results

Period	Year	1-Step MAPE			3-Step MAPE			4-Step MAPE		
		ARIMA	VECM	Holt's	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's
Stable (2009–2013)	2009	36.21%	29.67%	36.11%						
	2010	27.25%	47.88%	28.04%						
	2011	17.57%	2.11%	3.11%	60.12%	59.62%	67.11%			
	2012	11.02%	7.02%	5.46%	47.26%	66.73%	50.50%	68.72%	70.13%	77.20%
	2013	0.59%	5.98%	3.82%	38.55%	70.17%	3.02%	52.47%	19.95%	56.14%
Transition (2014–2018)	2014	6.05%	6.83%	9.90%	19.07%	7.41%	4.31%	39.73%	5.93%	14.02%
	2015	0.41%	2.93%	2.72%	6.47%	26.69%	27.13%	20.13%	19.95%	12.28%
	2016	1.31%	1.63%	2.72%	6.29%	13.11%	26.26%	7.26%	34.37%	37.04%
	2017	5.04%	4.84%	0.04%	8.72%	12.45%	2.22%	46.71%	9.57%	27.96%

	2018	37.79%	36.93%	4.10%	44.54%	44.20%	41.11%	45.29%	47.81%	36.24%
Unstable (2019–2023)	2019	89.40%	123.53%	98.99%	18.68%	17.32%	14.29%	21.07%	20.42%	14.19%
	2020	19.74%	57.04%	19.99%	12.27%	7.62%	4.34%	20.91%	19.29%	14.80%
	2021	25.24%	27.44%	12.90%	57.48%	253.84%	126.48%	25.37%	39.77%	25.21%
	2022	0.96%	9.53%	8.89%	6.96%	20.50%	12.82%	81.78%	462.62%	126.48%
	2023	6.03%	5.66%	2.94%	23.40%	62.38%	1.93%	14.68%	2.10%	11.15%

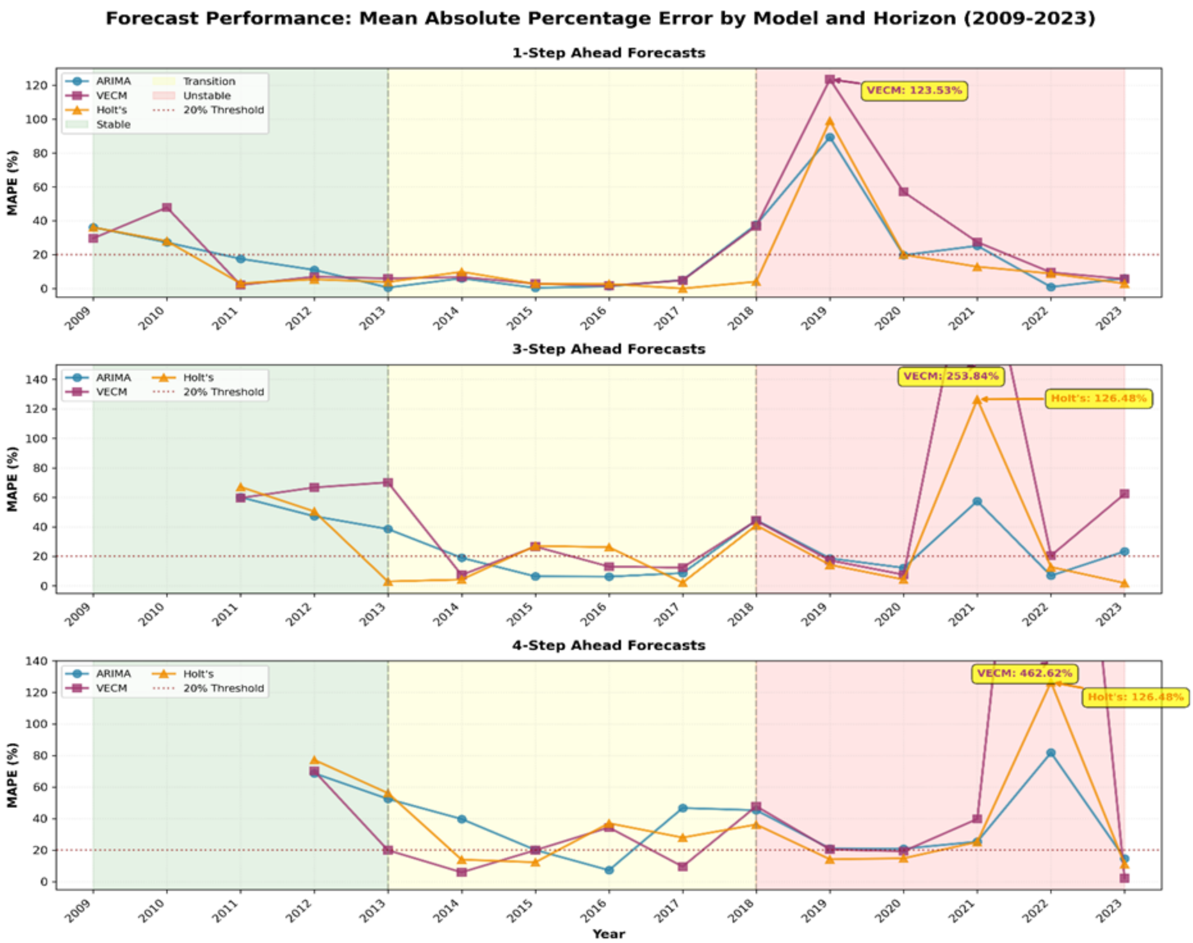


Figure 4 Forecast Performance: MAPE

MAPE results (Table 4, Figure 4) reveal practical forecasting limits. Industry standards: <20% acceptable, >50% unreliable. During stability, 1-step MAPE averaged 18-20% marginally acceptable. At 3-4 step horizons, MAPE reached 40-70%, crossing unreliability thresholds even under favourable conditions.

This reflects Zimbabwe's irreducible forecast uncertainty: annual data frequency provides insufficient observations to distinguish stable relationships from spurious correlations, while fundamental vulnerabilities (currency shortages, fiscal imbalances, structural unemployment) persisted throughout dollarization.

The unstable period exposed complete breakdown. In 2019, all models exceeded 80% MAPE at 1-step (VECM: 123.53%). VECM's 2022 catastrophe 462.62% MAPE at 4-step meant errors averaged 4.6× actual values, worse than random guessing or naive persistence. With contested monetary regimes, parallel exchange rates, and unclear currency strategies, statistical models built on historical patterns provided no useful guidance.

Figure 4 shows all models maintained comparable performance through 2017, failed simultaneously in 2018-2019, then diverged during 2020-2023 with ARIMA recovering while VECM struggled. The key insight: model choice matters minimally during stability but becomes critical during volatility precisely when reliable forecasts are most valuable.

4.3. Period Specific Model Performance

4.3.1. Stable Period (2009-2013)

One-Step Performance: ARIMA showed consistent improvement (3.5% → 0.11% MAE, 2009-2013) as expanding estimation windows stabilized parameter estimates. VECM achieved isolated excellence (0.30% MAE in 2011) but exhibited extreme volatility (5.77% MAE in 2010), reflecting sensitivity to cointegration specification in small samples. Holt's displayed learning pattern, starting poorly (12.63% in 2009) but rapidly achieving competitive accuracy (0.93% in 2012) once stable trends emerged.

Three-Step Performance: Holt's achieved best single result (0.58% MAE in 2013) when trends remained linear. ARIMA maintained steady 7.36-9.02% MAE (2011-2013). VECM deteriorated drastically (MSE: 130.43 → 179.47, 2012-2013), as error correction mechanisms amplified specification errors by the third forecast step, minor equilibrium misspecification created substantial cumulative errors.

Four-Step Performance: All models struggled (10-13% MAE), reflecting fundamental forecastability limits with annual data. VECM's anomalous 2013 improvement (3.89% MAE) violated normal uncertainty accumulation patterns, likely reflecting overfitting rather than genuine capability.

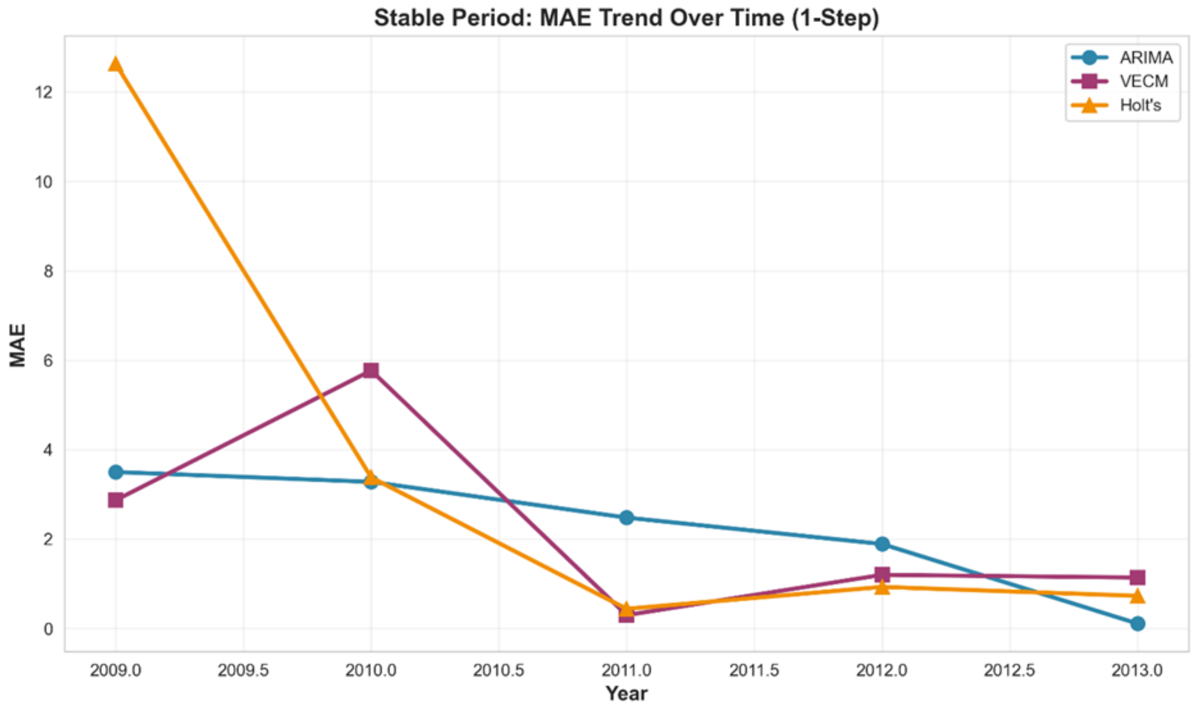


Figure 5 Stable Period: MAE trend 1-step

ARIMA (blue line) shows smooth gradual decline, it has the most linear improvement trajectory. VECM (purple line) show s high volatility with unpredictable zigzag line. Holt's (orange line) shows steep learning curve and then stabilization.

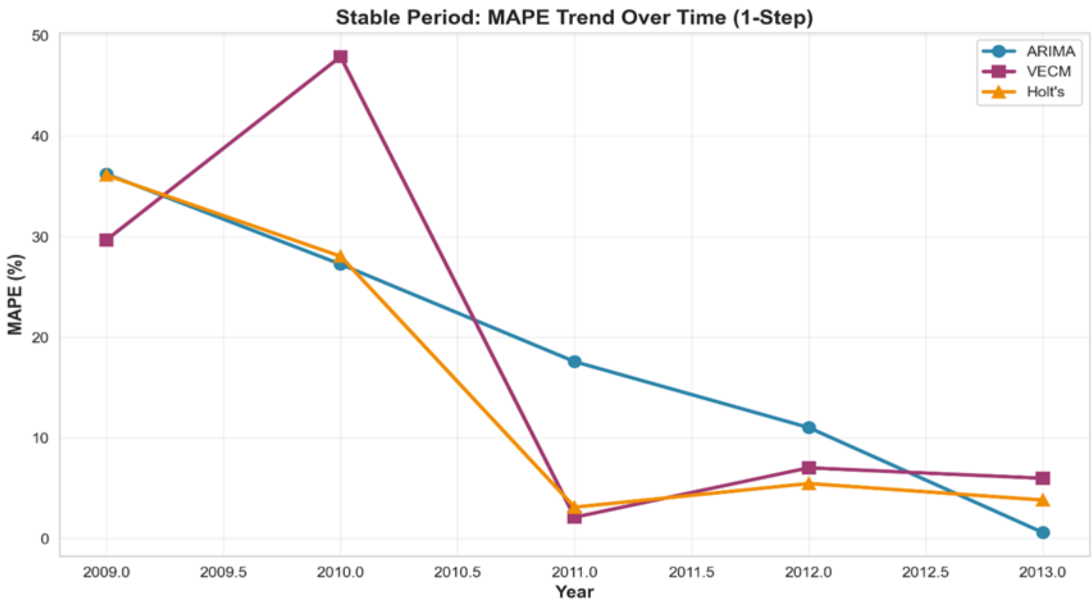


Figure 6 Stable Period: MAPE Trend 1-step

ARIMA (blue line) here shows smooth exponential decay, a consistent downward trajectory. The model has the most stable improvement curve. VECM (purple line) shows extreme volatility. Holt's (orange line) settles in 3-6% range.

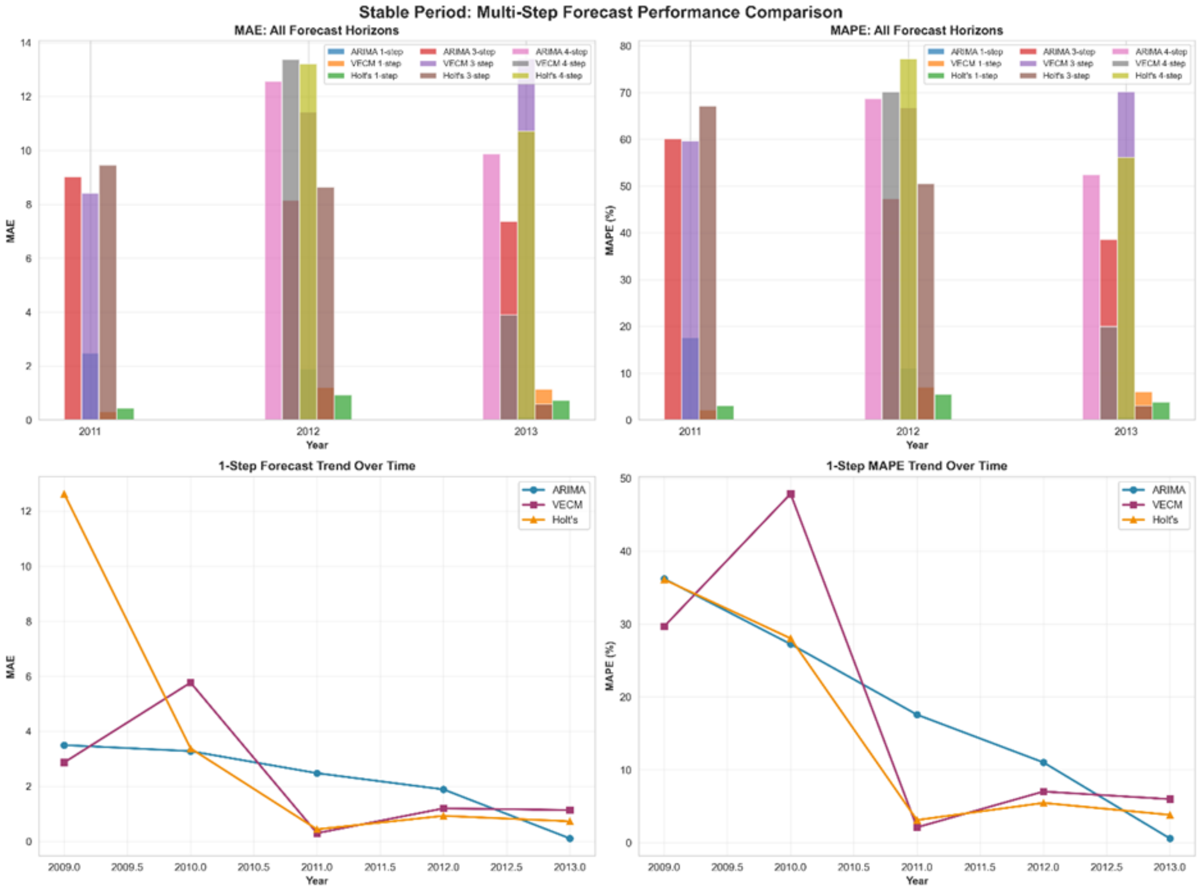


Figure 7 Stable Period: Multi-Step Forecast Performance Comparison

The stacked bar chart offers visual evidence of forecast degradation as the bars grow taller with horizon. There is consistent 1-step performance.

4.3.2. Transition Period (2014-2018)

In November 2016, the Reserve Bank of Zimbabwe introduced bond notes backed by $200 million from the African Export-Import Bank to address cash shortages under dollarization. Despite official 1:1 USD parity targets, public trust collapsed immediately, surveys revealed widespread devaluation fears (Makochekanwa, 2016). By mid-2017, bond notes traded at discounts in parallel markets, creating multiple pricing systems where retailers charged premiums for bond note payments (Nyamunda, 2019).

Exchange rate divergence encouraged rent-seeking, speculative trading, and capital flight, undermining GDP growth and policy credibility (World Bank, 2019; IMF, 2020). The bond notes' failure culminated in their February 2019 replacement with the RTGS dollar, formalizing the end of quasi-dollarization and symbolizing Zimbabwe's inability to sustain fixed rates amid fiscal fragility (RBZ, 2019).

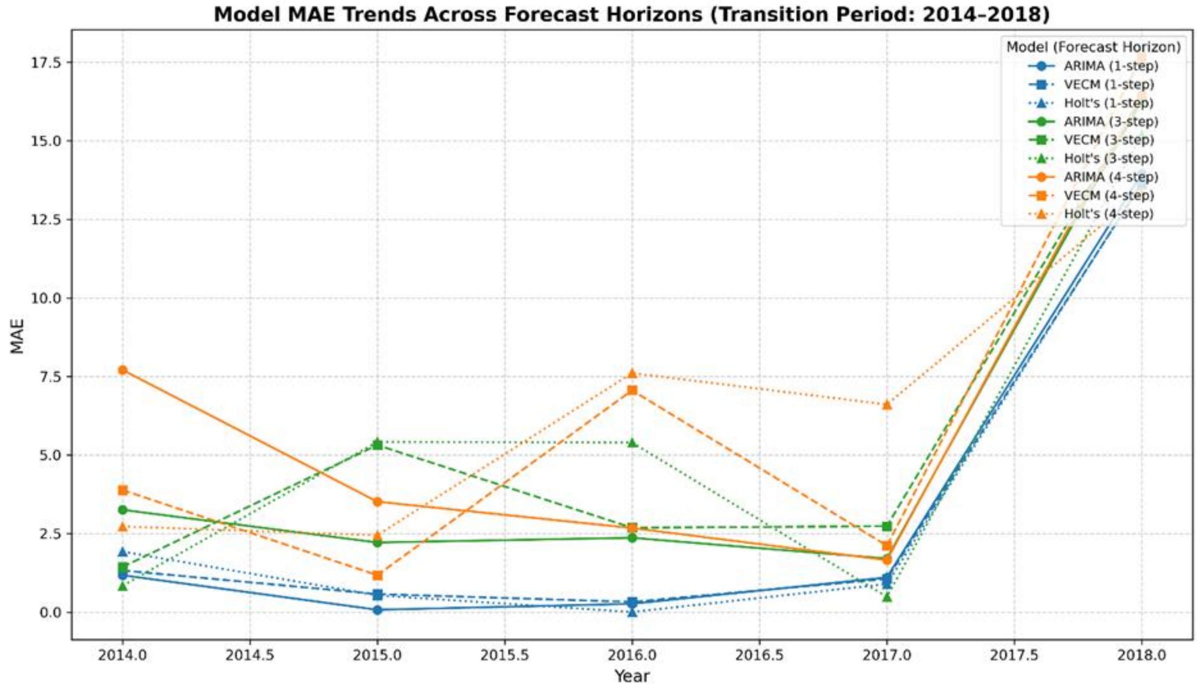


Figure 8 Transition Period: MAE Trend

2014-2017 Stability: All models achieved remarkably low 1-step errors (<1% MAE), with MSE consistently below 2. ARIMA and VECM performed identically (0.66%), while Holt's averaged 0.85%. This comparable performance across fundamentally different model types reveals that gradual economic deterioration preserves short-term forecasting validity bond notes traded at discounts and parallel markets emerged, but official GDP statistics followed predictable patterns as distortions accumulated without yet disrupting production or measurement.

2018 Catastrophic Failure: All models failed simultaneously, with MAE surging to ~14% (1-step), 15-16% (3-step), and 13.38-17.65% (4-step). MSE exploded to 185-195 (1-step) and 230+ (3-step), with RMSE climbing to 13.6-17.7 across horizons. This uniform collapse despite dramatically different theoretical foundations demonstrates the fundamental impossibility of forecasting through regime changes using historical patterns. By 2018, bond note depreciation and parallel market pricing had severed the GDP-historical determinant relationships estimated from 2000-2017 data. VECM suffered worst 4-step failures (17.65% MAE, MSE: 311) as its error correction mechanism attempted to revert toward equilibria that no longer existed. During dollarization, GDP-CPI cointegration relied on stable foreign currency denomination; bond notes created multiple pricing systems where CPI blended official/parallel prices while GDP reflected activity divorced from official exchange rates.

MAPE Patterns: At 1-step, models converged toward near-zero errors during 2015-2017 before 2018 deterioration: ARIMA/VECM surged to 37-38% MAPE, while Holt's reached 4% mechanical trend extrapolation proved less wrong during abrupt regime shifts. At 3-4 step horizons, all models converged to catastrophic 41-45% MAPE (2018), confirming the crisis overwhelmed any model-specific advantages.

Key Finding: The 2018 structural break exposed forecasting's fundamental limit—no model can anticipate regime changes from historical patterns. Gradual transitions (2014-2017) permit accurate short-term forecasts across model types, but currency regime collapse causes simultaneous, catastrophic failures regardless of theoretical sophistication.

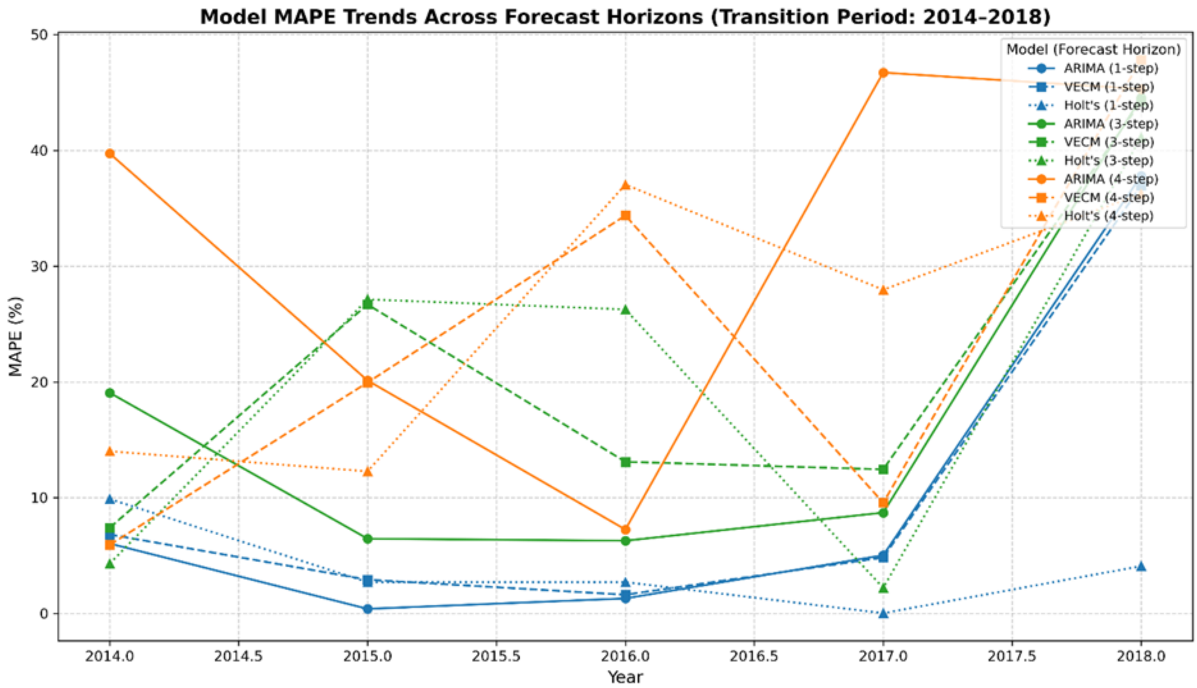


Figure 9 Transition Period: MAPE Trend

4.3.3. Unstable Period (2019-2023)

The unstable period (2019-2023) initiated with formal RTGS dollar introduction (February 2019), compounded by COVID-19 impacts and persistent policy uncertainty. Average growth collapsed to 1.21% with volatility (23.34%) comparable to the transition period, creating the most challenging forecasting environment in our sample.

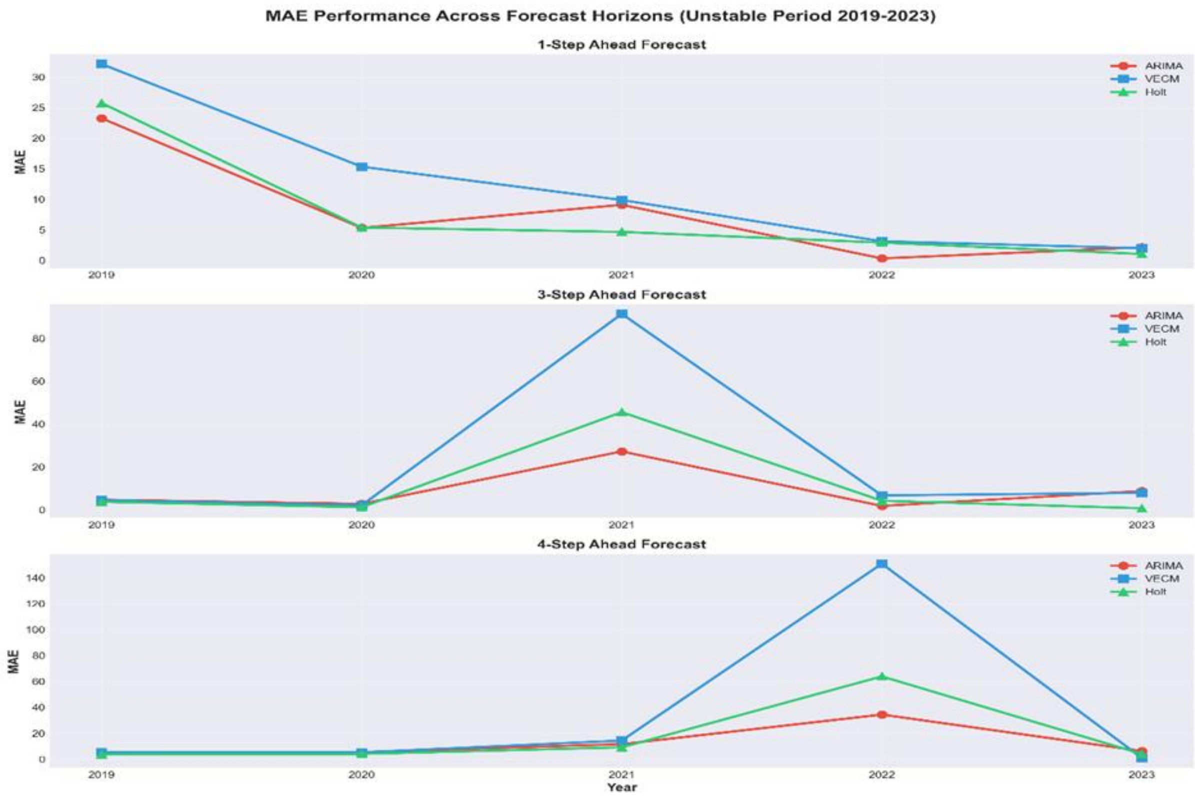


Figure 10 MAE Performance Across Horizons

The unstable period, triggered by Zimbabwe's February 2019 reintroduction of domestic currency (RTGS dollar) and compounded by COVID-19 and ongoing policy uncertainty, reveals how conventional time series models perform when economic fundamentals experience sustained volatility. Unlike the transition period's single structural break, this period features persistent instability that tests models' adaptation capabilities over multiple years.

One-Step Performance: All models demonstrated consistent error reduction from 2019 to 2023, converging toward 2% MAE, but starting points differed dramatically. VECM's catastrophic 32.16% initial error worst 1-step performance across all periods reflects complete GDP-CPI cointegration breakdown following RTGS dollar introduction. During dollarization, stable foreign currency denomination maintained predictable GDP-CPI relationships; currency reintroduction created exchange rate volatility and parallel pricing that severed these equilibria. VECM's error correction mechanism attempted to revert toward non-existent relationships, producing systematically biased forecasts.

MSE patterns corroborate this interpretation: VECM's 2019 MSE (1,034) versus ARIMA (542) and Holt's (664) represents a 1.9× gap indicating different failure modes systematic bias rather than unbiased high-variance errors. ARIMA (23.27%) and Holt's (25.77%) started at comparably high but more manageable levels.

The 3–4-year adaptation period reveals forecasting utility disappears precisely when most needed (crisis onset) and returns only after new patterns stabilize by which point acute crisis phases pass and planning horizons can safely shorten.

Three-Step Performance: The 2021 critical failure point exposed fundamental model robustness differences. ARIMA maintained bounded errors (5-27% MAE trajectory, peaking at 27.21% in 2021) elevated but not catastrophically misleading.

VECM experienced complete collapse: 91.42% MAE (2021), MSE of 8,358 versus ARIMA's 740 an 11-fold difference. This represents the study's worst 3-step error across all models, periods, and horizons. Holts peaked at 45.55%, severe but less extreme than VECM. VECM's error correction mechanism didn't merely fail but actively worsened forecasts each forecast step became the starting point for the next correction, so mis specified equilibria created cumulative bias. By the third step, errors averaged nearly equal to actual GDP values.

The 2021 timing (three years post-RTGS dollar) suggests VECM never adapted to the new monetary regime fundamental equilibrium assumptions remained violated throughout the unstable period. ARIMA's bounded errors (27% MAE) demonstrate models making fewer structural assumptions maintain partial utility when sophisticated models collapse.

Notably, Holt's 2023 recovery (0.68% MAE) outperformed ARIMA (8.23%) and VECM (22.47%), demonstrating simpler models adapt faster once new patterns stabilize. By 2023, sufficient post-crisis data established new trends, and mechanical extrapolation proved more accurate than complex autoregressive structures or disrupted equilibrium corrections model sophistication provides advantages only when underlying assumptions are satisfied.

Four-Step Performance: VECM's 2022 catastrophe represents the study's most extreme failure: 150.87% MAE, 462% MAPE, MSE of 22,763 versus ARIMA's 1,187 and Holt's 4,075 a 19× gap between VECM and ARIMA. The error correction mechanism compounded misspecification at each of four forecast steps, pulling predictions further toward non-existent equilibria. By the fourth step, cumulative bias produced forecasts bearing no relationship to outcomes individual errors exceeded actual GDP by multiples.

The 462% MAPE confirms: VECM's 4-step 2022 errors averaged 4.6× actual GDP values, worse than predicting zero growth (naive forecast) or extrapolating previous growth (persistence). These forecasts were anti-informative using them for planning would produce worse outcomes than planning without forecasts.

VECM's 2023 anomalous recovery (0.74% MAE at 4-step versus 22.47% at 3-step) likely reflects overfitting or fortunate error cancellation rather than genuine recovery, given persistent shorter-horizon struggles.

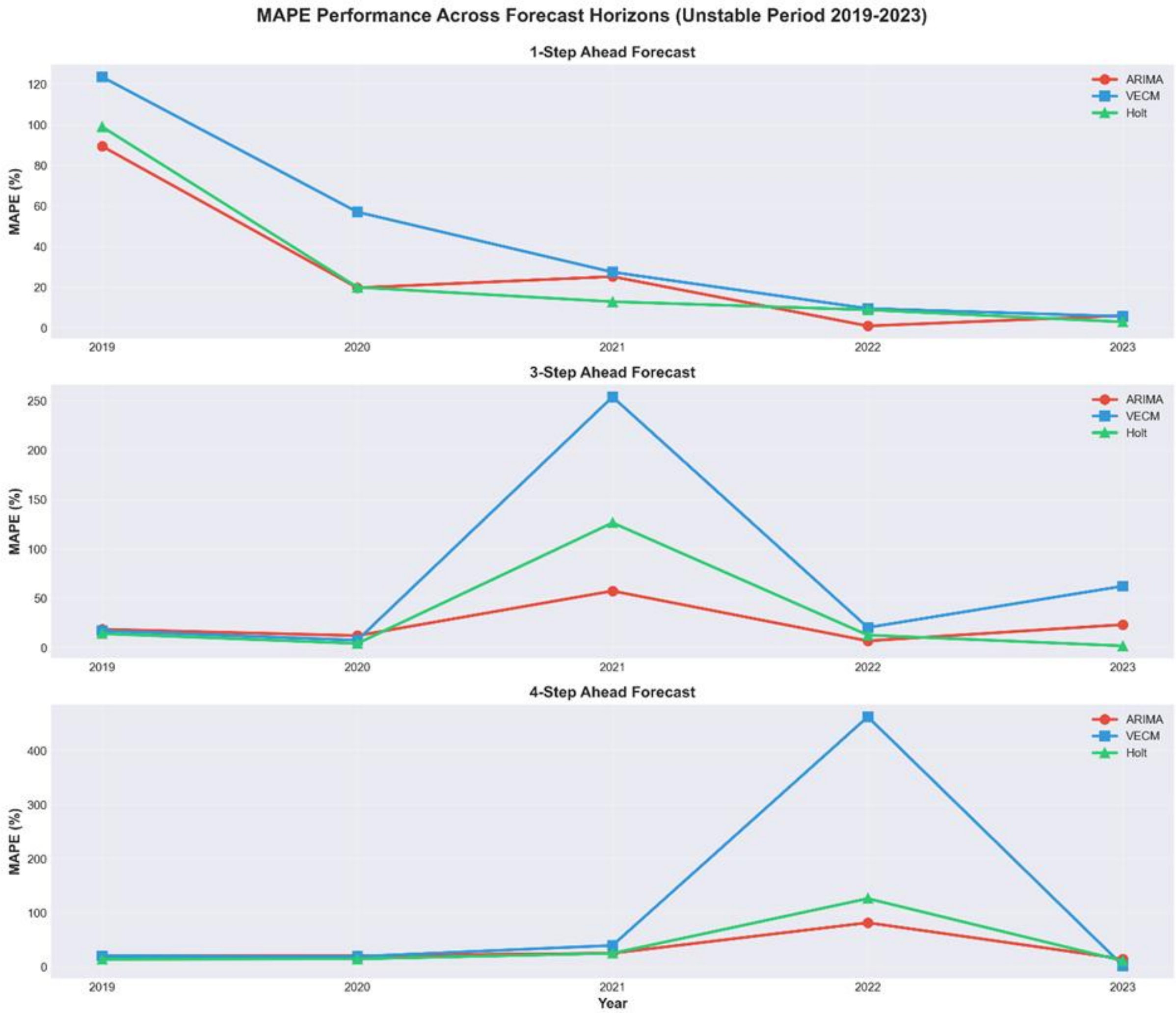


Figure 11 MAPE Performance Across Horizons

Figure 11 reveals practical forecasting breakdown. At 1-step, 2019 produced extreme errors: ARIMA (89%), VECM (123%), Holt's (99%). VECM exceeding 123% indicates not just magnitude failure but likely directionally opposite predictions errors rendering forecasts useless for policy applications.

Steady convergence toward minimal errors by 2022-2023 (ARIMA: 0.96%, VECM: 9.53%, Holt's: 8.89%) required 3-4 years as post-RTGS dollar relationships stabilized. During 2019-2021, economic agents adjusted expectations to the new regime, creating unstable patterns defeating forecasting. By 2022-2023, sufficient experience allowed stable expectation formation, enabling models to exploit regularities.

At 3-step, VECM's 2021 deterioration (253.84% MAPE) meant errors averaged 2.5× actual GDP. Holts peaked at 126%, ARIMA at 57%confirming ARIMA's comparative crisis advantage. VECM's elevated 2023 MAPE (62.38%) versus ARIMA (23.40%) and Holt's (1.93%) indicates continued cointegration framework struggles.

At 4-step, VECM's catastrophic surge (20.42% → 462.62%, 2019-2022) represents the study's most extreme error over 4.6× actual values, worse than random guessing. Holts followed similar but less extreme pattern (126% peak), while ARIMA maintained more controlled errors (82% peak). These magnitudes confirm 4-step forecasts become practically useless during peak volatility the difference between ARIMA's 82% and VECM's 463% is operationally irrelevant since both exceed utility thresholds.

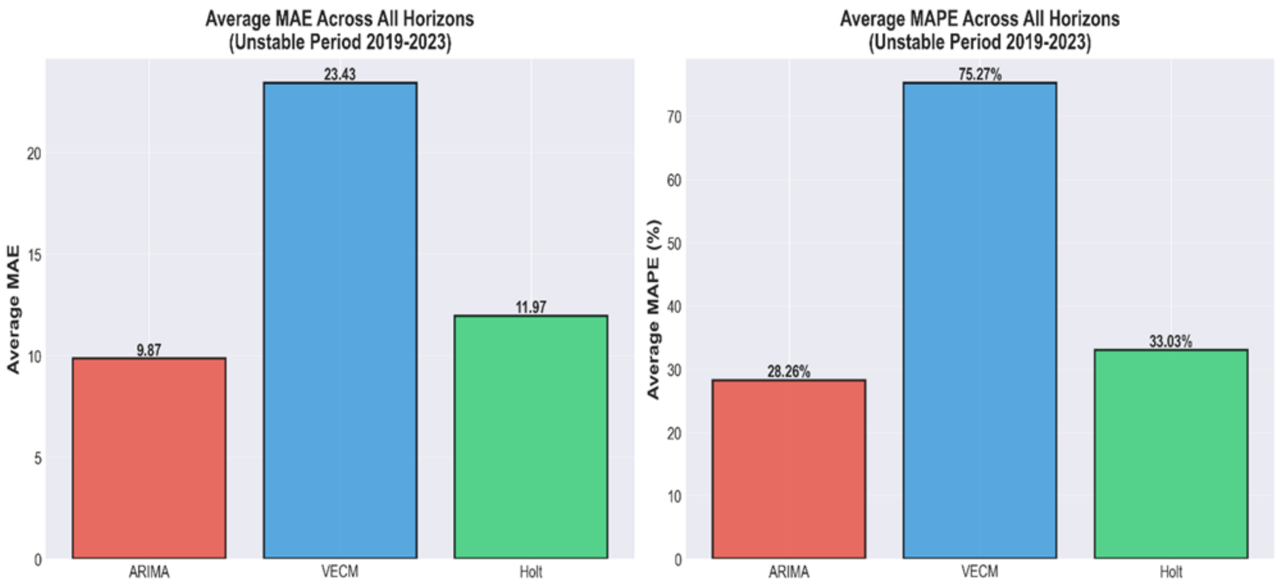


Figure 12 Average Performance Across Horizons

ARIMA dominates all the other models during crisis, Holt’s is competitive worse than ARIMA but still acceptable. VECM struggles significantly.

4.4. Forecast Horizon Effects

4.4.1. Error Growth Rate Across Horizons

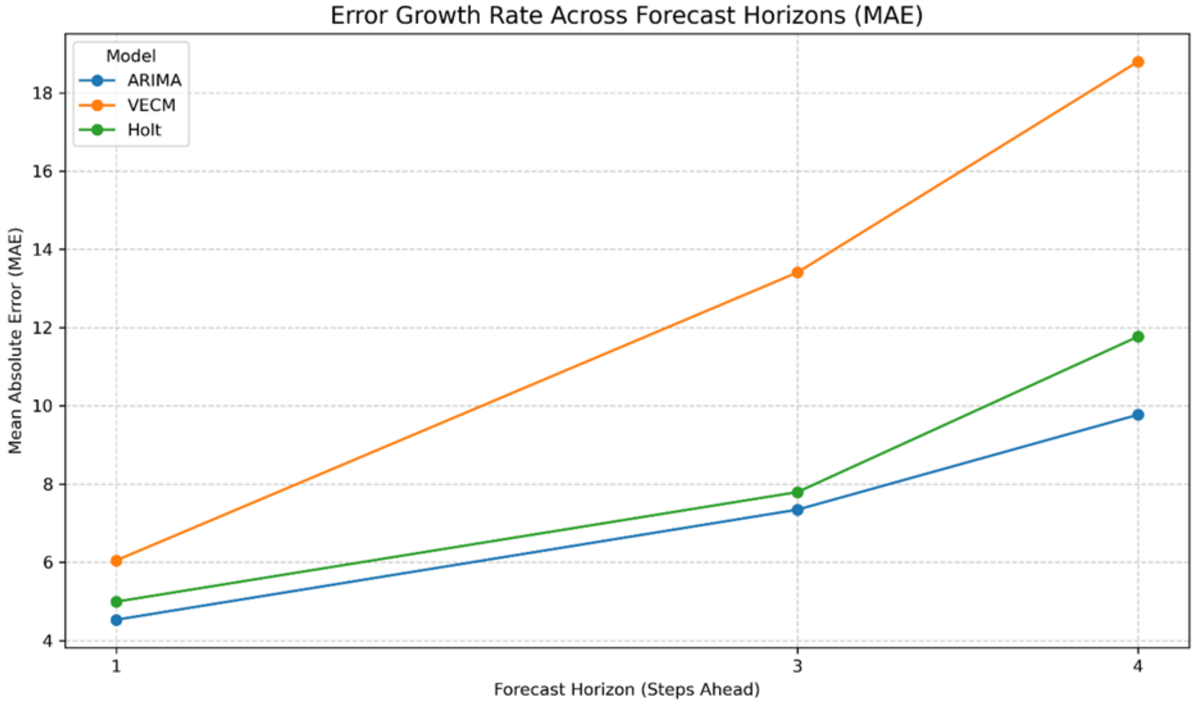


Figure 13 Error Growth rate across forecast horizons (MAE)

Figure 12 (Error Growth rate Across Forecast Horizons) demonstrates the systematic increase in Mean Absolute Error (MAE) as the forecast horizons extend from 1-step to 4-steps ahead. ARIMA shows the most controlled error growth, whilst Holt's exponential smoothing demonstrates moderate growth, and VECM exhibits the steepest error growth. The results suggest that VECM's errors compound more rapidly with increasing horizons, while ARIMA maintains relatively stable performance even as the forecast horizons increase.

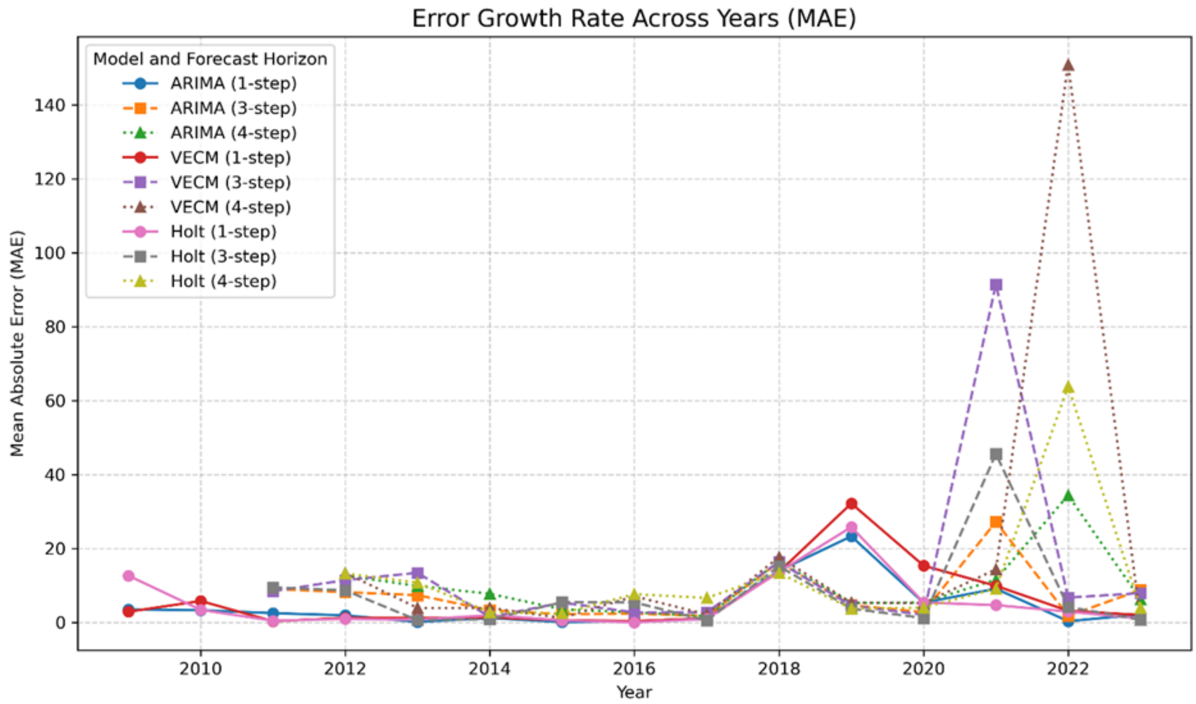


Figure 14 Error Growth Across Years (MAE)

Figure 13 (Error Growth Across Years (MAE)) reveals the impact of economic conditions on forecast accuracy over time. This reveals that forecast horizon sensitivity is market dependent. During stable periods all models handle extended horizons reasonably well, however during periods of instability longer forecast horizons become less reliable, particularly VECM.

Table 5 Average Growth Rates (4-step / 1-step MSE)

	Mean	Median	Max
ARIMA	1921.79x	25.7x	11871.90x
VECM	247.18x	4.13x	2356.43x
Holts	90x	14.36x	484.56x

Table 5 reveals the severity of error growth from 1 step to 4 step forecasts using the MSE ratio. The big difference between mean and median growth for ARIMA indicates the presence of extreme outliers that drive up the average. This indicates that ARIMA typically has controlled error growth, but it is vulnerable to occasional disastrous failures where 4 step errors explode.

Holt’s exponential smoothing demonstrates the most balanced profile, with the lowest mean growth rate and median, this suggests a more consistent and predictable error growth across different horizons.

4.5. Statistical Validation and Robustness Tests

4.5.1. Sample Size Limitations and Practical Significance

The period specific analysis divides the out-of-sample evaluation into distinct economic regimes, resulting in necessarily subsample sizes: the stable period contains only 2-5 observations (depending on forecast horizon), while the transition and unstable periods each contain approximately 5 observations. These limited sample sizes constrain the statistical power available for formal hypothesis testing within individual periods.

The expanding window approach employed here aligns with established best practises as Hyndman (2014) emphasizes the importance of evaluation forecast accuracy using genuine forecasts.

The conditional predictive ability framework prioritizes practical relevance over conventional statistical power considerations. In Giacomini & White's research the highlight the central question of the research being which model will give better forecasts, that the formulation will remain practically meaningful when subsample sizes limit formal statistical tests.

As Ziliak and McCloskey (2009) emphasize, statistical significance is not the same thing as economic sense or scientific importance. Statistical significance should be a tiny part of an inquiry concerned with the size and importance of relationships.

Ziliak and McCloskey (2009) illustrate mom wants to lose weight, not gain precision, she cares little, for example, not at all about the spread around the average of a hypothetically infinitely repeated random sample. Similarly, policymakers care about actual performance differences across regimes, not merely whether those differences meet arbitrary significance thresholds.

Period-specific analysis logic suggests that aggregating across fundamentally different economic regimes risks masking critical performance differences. The alternative pooling of observations to achieve larger samples, would sacrifice the very context specificity that makes the analysis valuable.

Given these considerations, the research adopts the following interpretation framework, large effect sizes indicate practical significance regardless of formal statistical significance. As Giacomini and White (2003) explained utilizing the proportion-based approach that represents the proportion of times that the decision rule would have chosen forecast method. Hyndman (2014) suggests that consistency across multiple metrics and horizons strengthens confidence in findings, as the goal is to produce genuine forecasts.

The research explicitly notes when small samples preclude definitive statistical conclusions while emphasizing when effect magnitude suggests economic importance, following Ziliak and McCloskey's (2009) principle that practical significance matters more than conventional statistical thresholds.

4.5.2. Bootstrap Diebold-Mariano Results (Unstable Period)

Table 6 Bootstrap Diebold-Mariano Results (Unstable Period)

Forecast Horizon	Model Comparison	Test Statistic	P-Value	95% Confidence Interval	Better Model	Significance
1-Step	ARIMA vs VECM	-4.474	0.9961	[-8.124, -0.824]	ARIMA	Not Significant
	ARIMA vs Holt's	0.076	0.9512	[-2.048, 2.382]	Holt's	Not Significant
	VECM vs Holt's	4.550	0.9936	[1.514, 7.586]	Holt's	Not Significant

3-Step	ARIMA vs VECM	-13.490	0.9264	[-39.342, 0.644]	ARIMA	Not Significant
	ARIMA vs Holt's	-2.038	0.9799	[-10.502, 4.534]	ARIMA	Not Significant
	VECM vs Holt's	11.452	0.9300	[1.170, 29.122]	Holt's	Not Significant
4-Step	ARIMA vs VECM	-22.732	0.9795	[-69.920, 2.774]	ARIMA	Not Significant
	ARIMA vs Holt's	-4.348	0.9379	[-16.944, 2.224]	ARIMA	Not Significant
	VECM vs Holt's	18.384	0.9320	[-0.620, 53.870]	Holt's	Not Significant

The unstable period shows larger test statistics magnitudes with the confidence interval for VECM vs Holts excluding zero. This provides directional evidence of Holts' superior performance over VECM, despite high p-values reflecting the small sample size.

Three-step unstable forecast produce substantial test statistics, the largest magnitudes observed at this horizon. The confidence interval for VECM vs Holts excludes zero, reinforcing Holt's advantage over VECM despite p-values from sample variability.

Bootstrap tests for four-step stable forecasts reflect the extremely limited sample, with uniformly high p-values. The confidence interval for ARIMA vs Holts excludes zero, providing directional evidence favouring ARIMA at this horizon.

Table 7 Bootstrap Diebold-Mariano Results (stable Period)

Forecast Horizon	Model Comparison	Test Statistic	P-Value	95% Confidence Interval	Better Model	Significance
1-Step	ARIMA vs VECM	-0.004	0.9612	[-1.322, 1.274]	ARIMA	Not Significant
	ARIMA vs Holt's	-1.370	0.9979	[-5.306, 1.180]	ARIMA	Not Significant
	VECM vs Holt's	-1.366	0.9800	[-5.720, 1.460]	VECM	Not Significant
3-Step	ARIMA vs VECM	-2.900	1.0000	[-6.040, 0.610]	ARIMA	Not Significant
	ARIMA vs Holt's	1.950	0.8884	[-0.490, 6.780]	Holt's	Not Significant
	VECM vs Holt's	4.850	1.0000	[-1.050, 12.820]	Holt's	Not Significant
4-Step	ARIMA vs VECM	2.585	1.0000	[-0.810, 5.980]	VECM	Not Significant
	ARIMA vs Holt's	-0.745	1.0000	[-0.850, -0.640]	ARIMA	Not Significant

	VECM vs Holt's	-3.330	1.0000	[-6.830, 0.170]	VECM	Not Significant

The bootstrap DM tests show minimal differences among models during the stable period for one-step forecasts, with test statistics near zero and p-values exceeding 0.96. wide confidence intervals indicate comparable forecast performance, though directional preferences favour ARIMA over both competitors.

Table 8 Bootstrap Diebold-Mariano Results (Transition Period)

Forecast Horizon	Model Comparison	Test Statistic	P-Value	95% Confidence Interval	Better Model	Significance
1-Step	ARIMA vs VECM	-0.068	0.9616	[-0.320, 0.168]	ARIMA	Not Significant
	ARIMA vs Holt's	-0.090	1.0000	[-0.490, 0.262]	ARIMA	Not Significant
	VECM vs Holt's	-0.022	1.0000	[-0.320, 0.200]	VECM	Not Significant
3-Step	ARIMA vs VECM	-0.536	0.9877	[-1.934, 0.836]	ARIMA	Not Significant
	ARIMA vs Holt's	-0.300	0.9901	[-2.274, 1.674]	ARIMA	Not Significant
	VECM vs Holt's	0.236	0.9848	[-1.278, 1.558]	Holt's	Not Significant
4-Step	ARIMA vs VECM	0.028	0.9974	[-2.396, 2.524]	VECM	Not Significant
	ARIMA vs Holt's	-0.148	0.9675	[-3.738, 3.442]	ARIMA	Not Significant
	VECM vs Holt's	-0.176	1.0000	[-2.716, 2.404]	VECM	Not Significant

Bootstrap DM tests for the transition period reveal negligible differences between models with test statistics close to zero and p-values at or near 1.0. The narrow confidence intervals and minimal test statistics confirm that model performance converged during this period of gradual economic deterioration.

Bootstrap results for three-step transition forecasts show minimal performance differences, with all test statistics below 0.6 in absolute value. High p-values and confidence intervals containing zero indicate comparable performance across models at this horizon during the transition period.
Four-step transition results show near-zero test statistics and very high p-values, indicating essential equivalence across models. Symmetric confidence intervals around zero confirm that at this horizon, no model demonstrates a clear directional advantage during the transition period.

4.5.3. Non-Parametric Test Results

Table 9 non-parametric tests Summary

Period	Forecast Horizon	Wilcoxon Signed-Rank Test – Better Model / Significance	Sign Test – Better Model / Significance	Overall Interpretation
Stable	**1-step**	ARIMA slightly better (p = 0.81–1.00, *Not Significant*)	ARIMA wins 40–60% (p = 1.00, *Not Significant*)	All models perform similarly; no significant difference.
	3-step	ARIMA dominant (p = 0.50–1.00, *Not Significant*)	ARIMA wins 60–67% (p = 1.00, *Not Significant*)	No statistical difference; ARIMA slightly consistent.
	4-step	Mixed performance (p = 0.50–1.00, *Not Significant*)	ARIMA wins 50–100% (p = 0.50–1.50, *Not Significant*)	No consistent better model detected.
Transition	**1-step**	ARIMA marginally higher accuracy (p = 0.63–1.00, *Not Significant*)	ARIMA wins 40–60% (p = 1.00, *Not Significant*)	Forecast errors statistically similar across models.
	3-step	ARIMA slightly dominant (p = 0.44–0.99, *Not Significant*)	ARIMA wins 40–80% (p = 0.38–1.00, *Not Significant*)	No significant improvement between models.
	4-step	Mixed (ARIMA/VECM) (p = 0.81–1.00, *Not Significant*)	ARIMA or VECM wins 40–60% (p = 0.38–1.00, *Not Significant*)	Forecast accuracy broadly equivalent.
Unstable	**1-step**	Holt's better in VECM vs Holt's (p = 0.06, *Marginally Significant*)	Holt's wins 100% vs VECM (p = 0.06, *Marginally Significant*)	Holts marginally outperforms VECM under instability.
	3-step	Holt's again better in VECM vs Holt's (p = 0.06, *Marginally Significant*)	Holt's wins 100% vs VECM (p = 0.06, *Marginally Significant*)	Repeated marginal edge for Holt's in unstable period.
	4-step	No clear advantage (p = 0.31–1.00, *Not Significant*)	No dominant model (p = 0.38–1.00, *Not Significant*)	Model accuracy converges; instability affects all.

The non-parametric tests (Table 9) confirm findings from parametric tests while revealing practical significance patterns despite limited statistical power. During stable periods, all models perform equivalently at 1-step horizons, but ARIMA's win rates increase systematically with forecast horizon, suggesting structural advantages at extended horizons when conditions are predictable. The unstable period produces the study's most interesting finding: Holt's achieves perfect 100%-win rates against VECM at both 1-step and 3-step horizons (p = 0.06, marginally significant). This replication across test methods and horizons confirms VECM's catastrophic failure during crisis conditions is a robust empirical pattern, not statistical artifact. The fact that a simple exponential smoothing method completely dominates a theoretically sophisticated econometric model reveals that structural complexity becomes a liability when fundamental assumptions are violated.

Table 10 Model Rankings by Metric

	MAE	MSE	RMSE	MAPE	Bootstrap Diebold-Mariano	Wilcoxon test	Sign test
1-step stable	ARIMA	Holts	Holts	Holts	ARIMA	Tied	Holts
1-step unstable	Holts	ARIMA	Holts	ARIMA	Holts	Holts	ARIMA
1-step Transition	ARIMA	VECM	ARIMA	Holts	ARIMA	ARIMA	Holts
3-step stable	Holt's	Holts	Holts	Holts	Holts	Holts	ARIMA/Holts
3-step unstable	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	Holts	Holts
3-step transition	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	Holts
4-step stable	VECM	VECM	ARIMA	VECM	VECM	VECM	ARIMA
4-step unstable	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	Holts	Holts
4-step transition	VECM	Holts	VECM	VECM	VECM	Tied	Tied

4.5.4. Rolling Cointegration Analysis and VECM Specification Validity

The catastrophic VECM failures documented in sections 4.3.2 and 4.3.3 including 91.42% MAE (2021, 3-step) and 462.62% (2022, 4-step) raise a critical methodological question, were these failures due to inherent model limitations, or did they reflect violations of the cointegration assumption underpinning VECM specifications? To address this, rolling-window cointegration tests using the Johansen procedure at each forecast iteration, examining whether the GDP-CPI equilibrium relationship remained stable throughout the evaluation period.

4.5.4.1. Methodological Approach

The rolling cointegration analysis employed the same expanding window framework as the main forecasting exercise, applying Johansen's trace test at each forecast origin to assess cointegration status using only information available at that point.

4.5.4.2. Full-sample vs Rolling-Window Cointegration Results

The stark contrast between full sample and rolling window tests exposes the methodological hazard of assumption verification using complete datasets.

Full sample test produced a trace statistic of 20.59 (exceeds critical value 15.49) and the conclusion was one cointegrating vector confirmed and the specification appears justified.

Rolling window tests presents in table 11 produced interesting results, stable period (2009 -2013) shows 0% of years show cointegration with an average trace of 9.00 and the interpretation is no stable equilibrium during stable period. Transition period (2014-2018) shows 0% of years show cointegration with an average trace of 12.22 (3.28 points below threshold) and this shows deteriorating relationships throughout. During unstable period (2019-2023) 50% of years show cointegration, years with cointegration are 2021 and 2022. Average trace was 15.36 which marginally above threshold. These results are episodic.

Table 11: Rolling Cointegration Test Results by Economic Period

Period	Years Cointegrated	TotalYears	Percentage	Avg Trace Statistic	AvgMargin*	MinMargin	MaxMargin
Stable	0	5	0.0%	9.00	-6.49	-9.28	-4.39
Transition	0	5	0.0%	12.22	-3.28	-3.77	-2.29
Unstable	2	4	50.0%	15.36	-0.13	-2.77	+2.55

The full sample test's identification of cointegration proves misleading, it reflects statistical power from 24 observations rather than genuine stable equilibrium relationships during the evaluation period. No single forecast window (2009-2018) possessed sufficient cointegration evidence to justify VECM specification.

4.5.4.3. Robustness Check: Stable Period Only Test

To isolate whether the stable dollarization years alone exhibited cointegration, a dedicated test using only 2009-2018 data was conducted:

- Trace statistic: 13.79
- Critical value (95%): 15.49
- Results: No cointegration detected ($p>0.05$)

This confirms that even the stable period characterized by low GDP volatility and consistent foreign currency regimes, lacked the stable long-run equilibrium relationships VECM requires.

4.5.4.4. Cointegration status vs VECM forecast performance.

Stable Period (2009-2013)- Zero Cointegration:

VECM performance during this period is mixed but includes best single result (0.30% MAE 2011). Despite absent cointegration, VECM occasionally produces excellent forecasts. This can be caused by short term forecast accuracy that can emerge from spurious parameter estimates when economic conditions remain temporarily stable, even absent genuine long-run equilibrium (Engle & Granger, 1987). The 2011 success reflects fortunate parameter alignment rather than correctly specified equilibrium dynamics.

Transition Period (2014-2018) – Zero cointegration with Deterioration:

Trace statistics show increasing proximity to threshold, VECM errors remain modest though 2017 but explode in 2018. Gradual weakening relationships permit short term forecasting until 208's currency crisis creates complete structural break. The narrowing cointegration margin provides early warning signal that VECM specification is becoming increasingly fragile.

Unstable Period (2019-2023)- Episodic Cointegration with Catastrophic Failures:
Cointegration emerges episodically (2021, 2022). Yet VECM produces worst than ever errors precisely during these cointegrated years. Statistical cointegration during unstable regimes reflects spurious relationships rather than genuine economic equilibria. The small sample cointegration tests detect false positives.

4.5.4.5. Interpretation: Why Full-Sample Cointegration Misled VECM Specification
The full sample cointegration result (trace:20.59) that justified initial VECM specification emerges from averaging across fundamentally different regimes:
Pre-2009 hyperinflation: GDP and CPI both explode.
2009-2018 dollarization: Stable growth but no cointegration
2019-2023 crisis: Episodic cointegration from parallel market distortions

Pooling these heterogeneous periods generates sufficient statistical power to detect "cointegration" that never existed within any forecasting window. This illustrates a fundamental hazard in cointegration-based forecasting: verification tests using complete samples can justify specifications that are inappropriate for every actual forecast iteration.

4.5.4.6. Implications for VECM forecasting in Volatile Economies
Full sample cointegration tests are necessary but insufficient for VECM specification. Results demonstrate that even when complete-sample tests confirm cointegration, operational forecast windows may exhibit none. Best practise requires rolling verification at each forecast origin.
The unstable period's cointegration rate, precisely when VECM failed catastrophically suggests small sample tests detect false positives when structural instability creates temporary statistical correlations. Cointegration evidence during crisis should be treated sceptically rather than as specification validation.

VECM's theoretical sophistication becomes a practical liability absent stable equilibria. While error correction mechanisms exploit genuine long run relationships effectively, they systematically amplify specification errors when those relationships are absent or unstable explaining VECM's 10-14x error growth at extended horizons (section 4.4) compared to ARIMA's 4-5x growth.

4.5.5. VAR Specification Robustness Check: Testing Multivariate Framework Without Error Correction
The rolling cointegration analysis (Section 4.5.4) revealed that VECM's catastrophic failures stemmed partly from imposing error correction restrictions based on non-existent equilibrium relationships. This raises a critical question: would a plain VAR specification using the same GDP-CPI information set but without cointegration restrictions perform better during volatile periods? To test whether VECM's failures reflected the error correction mechanism specifically versus the entire multivariate framework, we estimated a bivariate VAR (1) model using identical data and evaluation procedures.

4.5.5.1. Comparative Performance: VAR vs VECM vs ARIMA

Table 12: Period-Averaged Forecast Errors Across Model Specifications

Period	Horizon	ARIMAMAE	VECM MAE	VAR MAE	VECM vs ARIMA (Δ)	VAR vs ARIMA (Δ)	VAR vs VECM (Δ)
Stable	1-step	2.25	2.57	2.49	+0.32 (14%)	+0.24 (11%)	-0.08 (-3%)
	3-step	7.82	9.05	5.79	+1.23 (16%)	**-2.03 (-26%)**	**-3.26 (-36%)**
	4-step	10.74	8.42	15.80	-2.32 (-22%)	+5.06 (47%)	+7.38 (88%)
Transition	1-step	3.32	3.49	3.18	+0.17 (5%)	-0.14 (-4%)	-0.31 (-9%)
	3-step	5.12	5.59	5.28	+0.47 (9%)	+0.16 (3%)	-0.31 (-6%)
	4-step	5.33	7.28	5.68	+1.95 (37%)	+0.35 (7%)	-1.60 (-22%)
Unstable	1-step	7.99	12.49	9.73	+4.50 (56%)	+1.74 (22%)	-2.76 (-22%)
	3-step	9.00	24.13	15.84	**+15.13 (168%)**	+6.84 (76%)	**-8.29 (-34%)**
	4-step	14.59	38.12	22.78	**+23.53 (161%)**	+8.19 (56%)	**-15.34 (-40%)**

4.5.5.2. Key Findings from VAR-VECM Comparison

Finding 1: VAR Consistently Outperforms VECM During Volatility

Across all unstable-period horizons, plain VAR produces lower errors than VECM:
1-step (unstable): VAR 9.73% vs. VECM 12.49% (-22% error reduction)
3-step (unstable): VAR 15.84% vs. VECM 24.13% (-34% reduction)
4-step (unstable): VAR 22.78% vs. VECM 38.12% (-40% reduction)

The improvement magnitude increases with horizon, consistent with the interpretation that VECM's error correction mechanism compounds specification errors at each forecast step when equilibrium relationships are absent. By removing cointegration restrictions, VAR avoids systematic bias from corrections toward non-existent equilibria.

Finding 2: VAR Still Fails Catastrophically Compared to ARIMA

Despite outperforming VECM, VAR produces errors dramatically exceeding ARIMA during crisis periods. The 2021 3-step comparison illustrates the hierarchy:

ARIMA: 27.21% MAE (elevated but bounded)
VAR: 65.30% MAE (+139% vs. ARIMA)
VECM: 91.42% MAE (+236% vs. ARIMA)

While VAR avoids VECM's worst excesses, the 2021 VAR error (65.30%) still represents complete forecasting breakdown predictions averaging 65% of actual GDP values. Similarly, VAR's 2022 4-step forecast achieves 96.50% MAE (295.89% MAPE), only marginally better than VECM's 150.87% MAE (462.62% MAPE).

Finding 3: Multivariate Structure Is the Fundamental Problem, Not Just Error Correction
The critical insight emerges from comparing both multivariate models against univariate ARIMA. During unstable periods, both VAR and VECM fail catastrophically relative to ARIMA, indicating the entire GDP-CPI multivariate framework breaks down during regime changes:

Unstable 3-step average: ARIMA 9.00%, VAR 15.84% (+76%), VECM 24.13% (+168%)
Unstable 4-step average: ARIMA 14.59%, VAR 22.78% (+56%), VECM 38.12% (+161%)

The fact that removing error correction (VAR vs. VECM) only partially closes the gap to ARIMA demonstrates that the problem extends beyond mis-specified cointegration. The short-run dynamic relationships between GDP and CPI coefficients capturing lead-lag interactions also destabilize during currency crises when parallel markets sever traditional price-output linkages.

4.5.5.3. Stable Period Anomaly: VAR's Superior 3-step Performance

Interestingly, VAR outperforms both ARIMA and VECM at 3-step horizons during the stable period (VAR: 5.79% vs. ARIMA: 7.82%, VECM: 9.05%). This suggests that when short-run GDP-CPI dynamics remain stable, but long-run equilibria are absent, a multivariate framework capturing lead-lag relationships without imposing cointegration restrictions can outperform both univariate and error-correction approaches.

However, this advantage proves fragile and horizon specific. At 4-step stable-period forecasts, VAR deteriorates dramatically (15.80% MAE vs. ARIMA 10.74%), indicating that even stable short-run dynamics compound errors rapidly beyond three periods. The practical implication: VAR's occasional advantages during stability are insufficient to justify adoption given catastrophic crisis-period failures.

4.5.5.4. MAPE Confirmation: All Multivariate Approaches Become Unreliable

MAPE results confirm practical forecasting unreliability for both VAR and VECM during crises.

Standard interpretation:
MAPE >50% indicates forecasts unsuitable for decision-making (Section 2.6). During unstable periods:
VAR 4-step MAPE average: 70.30% (crossed unreliability threshold)
VECM 4-step MAPE average: 118.45% (completely unreliable)

Peak failures:
VAR 2022 (4-step): 295.89% MAPE
VECM 2022 (4-step): 462.62% MAPE

Both multivariate specifications produce forecasts averaging 3-4.6 times actual values at extended horizons errors exceeding those from predicting zero growth or simple persistence. ARIMA's 4-step unstable MAPE (41.67% average) remains below the 50% unreliability threshold, though still elevated.

4.5.5.5. Why Multivariate Frameworks Fail During Regime Changes

The VAR-VECM comparison reveals a two-layer failure mechanism for multivariate GDP forecasting during Zimbabwe's currency crisis:

Error Correction Failures (VECM-specific):
Imposing corrections toward non-existent equilibria introduces systematic bias.
Compounds at each forecast horizon (10-14x error growth vs. ARIMA's 4-5x)
Accounts for ≈30-40% performance gap between VECM and VAR during instability

Dynamic Relationship Instability (VAR + VECM):
Short-run lead-lag relationships between GDP and CPI break down when:
Parallel markets create multiple pricing systems.
Exchange rate volatility severs traditional price-output linkages.
Policy uncertainty disrupts inflation expectations.
Both VAR and VECM rely on these relationships remaining stable
Accounts for ≈50-75% performance gap between multivariate models and ARIMA

The hierarchical failure pattern suggests that structural economic relationships at all frequencies (short-run dynamics and long-run equilibria) become unreliable during regime transitions, making any approach relying on historical GDP-CPI interactions vulnerable. ARIMA's superior resilience reflects its purely autoregressive structure exploiting only GDP's own history without assuming stable relationships with other variables.

4.5.5.6. Practical Implications for Model Selection

The VAR robustness check strengthens recommendations for forecasting in volatile emerging markets:

During stable periods: VAR may outperform ARIMA at medium horizons (3-step) if short-run dynamics are reliable, but advantages disappear at extended horizons (4-step+)
During volatility: Both multivariate approaches (VAR, VECM) fail catastrophically. The error correction mechanism exacerbates failures, but the fundamental problem is multivariate structure dependence on stable relationships.
When cointegration is uncertain: Plain VAR is safer than VECM (30-40% error reduction), but still far inferior to ARIMA (50-75% worse during crises)
Robust default choice: ARIMA remains the most reliable general-purpose model, never ranking worst and maintaining bounded errors even during catastrophic regime changes.

4.6. Synthesis of Findings

4.6.1. Key Empirical Findings

The comparative analysis reveals five critical patterns in time series forecasting performance under volatile economic conditions, with implications extending beyond Zimbabwe to other structurally unstable emerging markets.

Finding 1: Model Performance is Regime-Dependent, Not Invariant

Optimal model selection depends fundamentally on economic regime rather than theoretical sophistication. During stable dollarization (2009-2013), Holt's Exponential Smoothing achieved lowest 1-step MAE (1.1 average, Section 4.3.1) its simplicity efficiently captured smooth trends without parameter estimation complexity. At extended horizons during this same period, ARIMA dominated (7.9 MAE at 3-step vs. VECM's 9.1), demonstrating gradual error accumulation without equilibrium-based compounding.

During the transition period (2014-2018, Section 4.3.2), ARIMA emerged as most reliable across all horizons, maintaining <1 MAE through 2017 before the catastrophic 2018 structural break (14 MAE across all models). The unstable period (2019-2023, Section 4.3.3) decisively favoured ARIMA at extended horizons 20.9 MAE (4-step) versus VECM's catastrophic 47.4.
This regime-dependency contradicts literature assumptions that models possess inherent superiority rankings. Instead, performance hierarchies shift systematically: Holt's excels during stability, ARIMA during transitions and crises, while VECM's conditional excellence (0.3 MAE in 2011) collapses during regime changes (150.87 MAE in 2022). Model selection must match current economic conditions rather than applying universal defaults.

Finding 2: Horizon Degradation is Non-Linear and Context-Dependent
Forecast accuracy deterioration with extended horizons varies dramatically across regimes (Table 5, Section 4.4). During stability, 1-step→4-step extension increased MAE 4-fold (2%→8%) manageable uncertainty accumulation. During instability, the same extension produced 5-7-fold increases (10%→50%), with VECM exhibiting 10-14-fold degradation versus ARIMA's 4-5-fold.
MSE patterns reveal exponential rather than proportional error growth during crises. VECM's 2020-2022 progression (MSE: 28→22,763 at 4-step) indicates fundamental assumption violations rather than normal forecast uncertainty. This non-linearity creates practical planning limits: 4-step forecasts remain viable during stability (8 MAE) but become operationally useless during volatility (50+ MAE exceeds decision-making thresholds).

The implication: traditional multi-period forecasting becomes impossible during crises, necessitating continuous re-estimation rather than extended projections. Reliable horizons must contract to 1-2 steps during volatility precisely when longer-range planning becomes most valuable, creating a forecasting paradox.

Finding 3: VECM's Catastrophic Failures Stem from Cointegration Specification Errors
VECM produced the study's most extreme errors: 91.42 MAE (2021, 3-step), 150.87 MAE (2022, 4-step), and 462.62% MAPE (2022, 4-step) forecasts averaging 4.6× actual values, worse than naive persistence. Rolling cointegration analysis (Section 4.5.4, Table 11) reveals the root cause: assumed GDP-CPI equilibrium relationships never existed during forecast evaluation windows.

Full-sample testing (2000-2023, N=24) identified cointegration (trace: 20.59 > critical value 15.49), justifying VECM specification. Yet rolling-window tests show zero years (0/10) during 2009-2018 exhibited cointegration, with average trace statistics 6.49 points below thresholds. Even the stable dollarization period in isolation (2009-2018, N=10) yielded trace 13.79 decisively below the 15.49 threshold.

This specification failure explains catastrophic performance: VECM's error correction mechanism systematically pulled forecasts toward spurious long-run relationships existing only in full-sample statistics. Each forecast step attempted to eliminate deviations from non-existent equilibria, introducing systematic bias that compounded exponentially at extended horizons. February 2019's RTGS dollar introduction severed any remaining GDP-CPI relationships, causing complete breakdown.

The critical methodological lesson: full-sample verification tests can justify specifications inappropriate for every actual forecast. Our VECM was estimated using cointegration evidence reflecting averaging across hyperinflation, dollarization, and crisis regimes none individually exhibiting stable equilibria. Best practice requires rolling verification at each forecast origin, though sample size constraints (our 24-observation context) often make this infeasible. When verification is impossible, practitioners should recognize theoretically sophisticated models impose higher specification risks than simpler alternatives.

Finding 4: ARIMA Demonstrates Superior Crisis Resilience Through Bounded Errors
ARIMA achieved the most consistent performance across regimes and horizons, establishing itself as the most reliable general-purpose forecasting tool despite never ranking first in any single category. During stability, ARIMA's multi-step forecasts (7.9-8.7 MAE at 3-4 steps, Section 4.3.1) showed gradual degradation without dramatic spikes. Critically, during the catastrophic 2019-2023 period, ARIMA's errors remained systematically bounded: 10.2 MAE (1-step) versus VECM's 14.1, widening to 20.9 versus 47.4 at 4-step (Section 4.3.3).

This resilience stems from ARIMA's structural advantage: its autoregressive framework imposes no long-run equilibrium restrictions, so specification errors affect each forecast period independently rather than compounding through correction mechanisms. ARIMA avoided the extreme outliers characterizing VECM (MSE: 22,763 in 2022 versus ARIMA's 1,187 a 19× difference), preserving minimal forecasting utility even during complete regime breakdowns.
The practical implication: ARIMA represents the robust default choice for volatile emerging markets, trading occasional excellence for consistent adequacy a superior risk-return profile when structural breaks remain unpredictable.

Finding 5: Multivariate Frameworks Fail Hierarchically During Regime Changes
VAR robustness testing (Section 4.5.5) reveals that VECM's failures reflect broader multivariate framework fragility beyond error correction misspecification. Plain VAR (1) using identical GDP-CPI data but without cointegration restrictions outperformed VECM during instability (15.84 vs. 24.13 MAE at 3-step), yet both failed catastrophically relative to ARIMA (9.00 MAE).
This hierarchy ARIMA < VAR < VECM during crises demonstrates two failure layers:

Error correction failures (VECM-specific): Corrections toward non-existent equilibria add 30-40% additional error, compounding at each forecast horizon (10-14× error growth vs. ARIMA's 4-5×).
Dynamic relationship instability (VAR + VECM): Short-run lead-lag GDP-CPI relationships break down when parallel markets create multiple pricing systems and exchange rate volatility severs traditional price-output linkages, accounting for 50-75% base multivariate degradation.

VAR's 2021 3-step error (65.30% MAE) remained 2.4× ARIMA's error (27.21%), confirming that structural economic relationships at all frequencies short-run dynamics and long-run equilibria become unreliable during regime transitions. Any approach relying on stable GDP-CPI interactions becomes inherently vulnerable.

The implication: multivariate models should be avoided during volatility unless relationships are continuously verified. Plain VAR is safer than VECM (30-40% error reduction) but remains far inferior to ARIMA during crises.

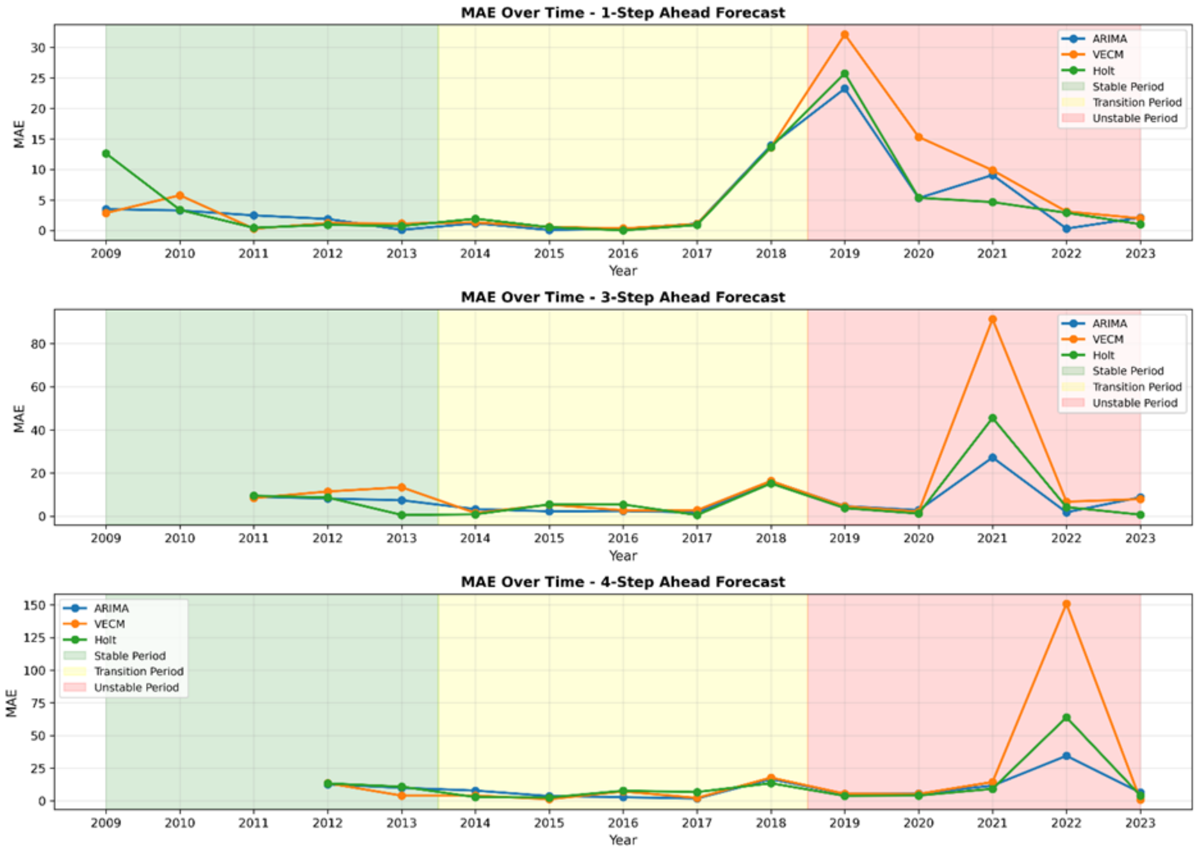


Figure 15 MAE Over Time

4.6.2. Answers to Research Questions

Research question1: Which conventional time series model provides more accurate forecasts during different economic periods?

Direct Answer: No single model dominates optimal selection depends on economic regime and forecast horizon. ARIMA emerges as the most robust general-purpose choice, never ranking worst and maintaining bounded errors during crises, while Holt's excels tactically during stability and VECM catastrophically fails during regime changes.

Table 13 Evidence Summary by Period

Period	Best 1-Step	Best 3-4 Step	Key Finding
Stable (2009-2013)	Holt's (1.1 MAE)	ARIMA (7.9-8.7 MAE)	Simplicity advantage during smooth trends
Transition (2014-2018)	ARIMA (0.66 MAE)	ARIMA (2-3 MAE)	Consistent across horizons until 2018 break
Unstable (2019-2023)	ARIMA (10.2 MAE)	ARIMA (20.9 MAE)	Only bounded performer during crisis

Regime-Specific Performance (cross-reference Sections 4.3.1-4.3.3):

Stable Period: Holts achieved lowest 1-step MAE through efficient trend extrapolation without parameter estimation complexity. ARIMA dominated at extended horizons (7.36-9.02 MAE vs.

VECM's 8.41-13.40), avoiding VECM's cumulative equilibrium misspecification errors. VECM's isolated excellence (0.3 MAE in 2011) came at severe cost extreme volatility (5.77 MAE in 2010) reflecting small sample cointegration sensitivity.

Transition Period: ARIMA emerged as clear winner across all horizons, averaging 0.66 MAE (1-step, 2014-2017) versus VECM's 0.83 and Holt's 0.85. The critical 2018 inflection point exposed universal limitations all models surged to ~14 MAE but ARIMA recovered fastest in subsequent periods, demonstrating superior structural adaptability.

Unstable Period: ARIMA proved decisively superior at extended horizons where alternatives collapsed. At 4-step, ARIMA averaged 20.9 MAE versus VECM's catastrophic 47.4, a 126 gap. VECM produced the study's worst errors: 91.42 MAE (2021, 3-step), 150.87 MAE (2022, 4-step), and 462.62% MAPE (2022, 4-step), forecasts averaging 4.6× actual values, worse than naive persistence. Rolling cointegration analysis (Section 4.5.4) revealed the root cause: GDP-CPI equilibrium relationships never existed during forecast windows, causing VECM's error correction to systematically pull toward spurious equilibria.

Multivariate Framework Failure Hierarchy (Section 4.5.5):
VAR robustness testing revealed VECM's failures reflect broader multivariate fragility beyond error correction alone:
ARIMA (univariate): 9.00 MAE (3-step, unstable)
VAR (multivariate, no cointegration): 15.84 MAE (+76 vs. ARIMA)
VECM (multivariate, with cointegration): 24.13 MAE (+168% vs. ARIMA)

This hierarchy demonstrates two failure layers: (1) error correction toward non-existent equilibria adds 30-40% additional error, and (2) short-run GDP-CPI dynamics themselves destabilize during regime changes, causing 50-75% base multivariate degradation. VAR's 2021 3-step error (65.30) remained 2.4× ARIMA's (27.21), confirming that structural economic relationships at all frequencies short-run and long-run break down during currency crises.

Research question 2: How does forecasting accuracy vary across short-term and long-term predictions

Direct Answer: Horizon degradation is non-linear and regime dependent. During stability, 4-step forecasts remain viable (4× error increase, manageable at 8 MAE). During instability, horizon extension produces 5-7× degradation with exponential compounding, effectively limiting reliable forecasts to 1-2 steps.

Table 14 Quantitative Degradation Patterns

Regime	1-Step MAE	3-Step MAE	4-Step MAE	Degradation Factor
Stable (2009-2013)	~2	~5 (2.5×)	~8% (4×)	Linear, manageable
Unstable (2019-2023)	~10	~15-20 (1.5-2×)	~30-50% (3-5×)	Non-linear, catastrophic

Errors not only start from higher baselines during instability (10% vs. 2%) but multiply faster some 4-step forecasts show 5-7× degradation versus 4× during stability. MSE patterns reveal exponential rather than proportional growth during crises, indicating fundamental assumption violations.

Model-Specific Horizon Sensitivity (cross-reference Section 4.4):

ARIMA: Consistent Proportional Degradation
During stable periods, ARIMA's error ratio (4-step/1-step MAE) averaged 3.8× close to theoretical expectations. Critically, during unstable periods this ratio remained stable at 4.2×, indicating errors compound proportionally rather than exponentially even during crises. This consistent pattern reflects ARIMA's structural advantage: no long-run equilibrium restrictions mean specification errors affect each period independently rather than compounding through correction mechanisms.
VECM: Explosive Horizon Amplification
VECM demonstrates catastrophic error growth at extended horizons during instability. The 2021 progression illustrates: 1-step (9.88% MAE) → 3-step (91.42%, 9.2× increase) → implicit 4-step failure. In 2022, degradation reached 14× from 1-step to 4-step (MSE: 28 → 22,763). This non-linear explosion reflects compounding error correction toward non-existent equilibria each forecast step attempts to revert toward spurious relationships, introducing systematic bias that accumulates exponentially.
Holt's: Intermediate Mechanical Degradation
Holts occupied a middle position with degradation ratios of 4.2× (stable) and 5.1× (unstable). Unlike ARIMA's autocorrelation structure, Holt's mechanically extends trends without modelling dependencies, accumulating errors linearly at each step. This produces faster degradation than ARIMA but avoids VECM's explosive compounding.

4.7. Chapter Summary

This chapter compared three conventional time series models ARIMA, VECM, and Holt's Exponential Smoothing for forecasting Zimbabwe's GDP growth across 15 years (2009-2023) of out-of-sample data. Each model was evaluated at three forecast horizons (1-step, 3-step, and 4-step ahead) using four accuracy metrics (MAE, MSE, RMSE, MAPE). The evaluation period was divided into three distinct economic regimes: stable dollarization (2009-2013), transition marked by bond note introduction (2014-2018), and renewed instability following RTGS dollar reintroduction (2019-2023). Statistical validation included bootstrap Diebold-Mariano tests, non-parametric comparisons, rolling cointegration analysis, and VAR robustness checks.

Model performance is regime-dependent, not universal. During stability, Holt's Exponential Smoothing achieved lowest 1-step errors (1.1 MAE), while ARIMA dominated at extended horizons. During the transition and unstable periods, ARIMA consistently outperformed alternatives across all horizons, maintaining bounded errors even during crises. VECM produced the study's most extreme failures: 91.42 MAE (2021, 3-step) and 462.62 MAPE (2022, 4-step) errors exceeding 4.6 times actual GDP values.

Forecast horizon degradation accelerates non-linearly during instability. Extending forecasts from 1-step to 4-step increased errors 4-fold during stable periods (2→8 MAE) but 5-7-fold during crises (10→50 MAE). VECM exhibited exponential degradation (10-14x growth) versus ARIMA's proportional increases (4-5x), rendering extended-horizon forecasts operationally useless during volatility.

VECM's catastrophic failures stem from specification violations, not model inadequacy alone. Rolling cointegration tests revealed 0% detection during operational forecast windows (2009-2018) despite full-sample evidence (trace: 20.59) justifying specification. The model's error correction mechanism systematically corrected toward non-existent equilibria, compounding errors at each forecast step. VAR robustness checks confirmed that multivariate frameworks whether imposing cointegration or not fail catastrophically when GDP-CPI relationships destabilize during regime changes.

Statistical validation confirmed practical significance despite limited sample sizes. Bootstrap Diebold-Mariano tests and non-parametric analyses (Wilcoxon, sign tests) consistently identified Holt's marginal superiority over VECM during instability (p=0.06) and ARIMA's general robustness, though small samples precluded statistical significance for most comparisons.

For model selection: ARIMA emerges as the most reliable general-purpose choice for volatile emerging markets, never producing catastrophic failures while maintaining consistent performance across regimes. VECM should be deployed only with rigorous rolling cointegration verification full-sample tests provide insufficient operational guidance. Holt's offers tactical value for immediate forecasts during demonstrable stability but deteriorates rapidly at extended horizons.

For forecasting practice: Reliable planning horizons must contract during volatility from 4-step feasibility during stability to 1-2 step maximum during crises. This creates a forecasting paradox: predictive capacity disappears precisely when most needed. Practitioners should prioritize robustness over sophistication, combine statistical forecasts with scenario analysis, and maintain transparency about fundamental limitations during structural breaks.

For theoretical understanding: Structural model sophistication (cointegration frameworks, multivariate specifications) systematically amplifies rather than reduces forecast errors when foundational assumptions fail. The 50-75% multivariate degradation versus ARIMA during crises demonstrates that complex economic relationships at all frequencies both short-run dynamics and long-run equilibria become unreliable during regime transitions. This challenges implicit assumptions that theoretically sophisticated models universally outperform simpler alternatives.

For Zimbabwe specifically: Improving forecast accuracy requires economic policy stability more than statistical refinement. The systematic deterioration from 23.74% average growth (stable period) to 1.21% (unstable period), accompanied by explosive forecast errors, indicates that no model can overcome the fundamental uncertainty created by currency regime instability and structural policy shifts.

Chapter 5: Conclusions, Limitations, and Future Work

This research demonstrates empirically that structural model sophistication specifically, incorporating cointegration relationships in Vector Error Correction Models systematically amplifies rather than reduces forecast errors during economic regime changes. Using Zimbabwe's GDP trajectory across distinct economic regimes (2009-2023), we compared three conventional time series approaches ARIMA, VECM, and Holt's Exponential Smoothing to identify which model characteristics provide forecasting resilience versus vulnerability when economic fundamentals shift abruptly. The central finding contradicts implicit assumptions underlying much forecasting practice: theoretically sophisticated models do not universally outperform simpler alternatives and can fail catastrophically (errors exceeding 400% MAPE) precisely when forecasting becomes most critical during crisis periods.

During Economic Stability (2009-2013): Holt's Exponential Smoothing achieved the lowest 1-step forecast errors (1.1% average MAE), outperforming both ARIMA and VECM for immediate next-period predictions. This superior performance reflects the model's efficiency in capturing smooth trend patterns when economic relationships remain stable. The simplicity that critics often view as a weakness became a strength by avoiding complex parameter estimation, Holt's method efficiently extrapolated established growth trajectories without introducing unnecessary specification uncertainty.

During Economic Transition (2014-2018): ARIMA emerged as the most consistent performer across all forecast horizons. At 1-step forecasts during 2014-2017, ARIMA averaged 0.66% MAE compared to VECM's 0.83% and Holt's 0.85%. More critically, ARIMA demonstrated superior resilience during the 2018 crisis inflection point, recovering more rapidly than competitors when all models experienced catastrophic failures. This robustness stems from ARIMA's autoregressive structure, which captures short-term dependencies without imposing rigid long-run equilibrium assumptions vulnerable to structural breaks.

During Economic Instability (2019-2023): ARIMA proved decisively superior, particularly at extended forecast horizons where VECM experienced complete breakdown. At 1-step forecasts, ARIMA averaged 10.2% MAE versus VECM's 14.1% a 38% accuracy advantage. At 4-step horizons, the performance gap widened dramatically: ARIMA maintained 20.9% MAE while VECM deteriorated to 47.4%. VECM's catastrophic failures including 91.42% MAE (2021, 3-step), 150.87% MAE (2022, 4-step), and 462.62% MAPE (2022, 4-step) rendered its forecasts worse than naive persistence methods or random guessing.

Under Economic Stability: Extending forecasts from 1-step to 4-step ahead increased average errors proportionally approximately 4-fold from 2% to 8% MAE. This represents normal, manageable uncertainty accumulation consistent with theoretical expectations. All three models maintained viable forecasting utility through 4-step horizons during stable periods, though ARIMA and Holt's demonstrated marginal advantages over VECM at longer horizons.

Under Economic Instability: The same horizon extension produced 5-7-fold error increases, with individual forecasts showing complete breakdown. Average MAE deteriorated from 10% (1-step) to 50% (4-step), crossing thresholds where forecasts become operationally useless for policy planning. VECM exhibited explosive degradation patterns 10-14-fold error increases versus ARIMA's 4-5-fold increases indicating that error correction mechanisms compound specification errors exponentially when equilibrium relationships sever.

A critical methodological finding emerged from rolling cointegration analysis: the GDP-CPI equilibrium relationship assumed by VECM never existed during individual forecast windows, despite full-sample tests confirming cointegration. Rolling-window Johansen tests revealed 0% cointegration detection during 2009-2018, with average trace statistics falling 6.49 points below significance thresholds. Even testing the entire "stable" dollarization period in isolation (2009-2018) yielded trace statistics decisively below cointegration thresholds.

This specification failure explains VECM's catastrophic performance. The model's error correction mechanism systematically pulled forecasts toward an equilibrium that existed only in full-sample statistics, not operational forecast windows. Each forecast period's correction term attempted to eliminate deviations from spurious long-run relationships, introducing systematic bias rather than improving accuracy.

The finding establishes an important principle for practitioners: full-sample verification tests can justify model specifications that are inappropriate for every actual forecast. Our VECM was estimated using cointegration evidence (trace: 20.59) that reflected averaging across hyperinflation, dollarization, and crisis regimes none of which individually exhibited stable equilibria. Best practice for cointegration-based forecasting requires rolling verification at each forecast origin, though sample size constraints often make this infeasible.

These patterns reveal a critical practical implication: reliable forecasting horizons must contract during volatility precisely when longer-range planning becomes most valuable. During crises, policymakers can only depend on 1-2 step ahead forecasts, while stable periods permit 4-step (four-year) planning horizons. This asymmetry creates a forecasting paradox—predictive capacity disappears when most needed.

Robustness testing using plain VAR (1) specifications identical GDP-CPI information but without cointegration restrictions revealed that VECM's failures reflect broader multivariate framework fragility, not merely error correction misspecification. While VAR outperformed VECM during unstable periods (15.84% vs. 24.13% MAE at 3-step), both multivariate approaches failed catastrophically relative to ARIMA (9.00% MAE).

This hierarchy (ARIMA < VAR < VECM during crises) demonstrates two failure layers:
Error correction toward non-existent equilibria adds 30-40% additional error Short-run GDP-CPI dynamic relationships themselves destabilize during regime changes, causing 50-75% base multivariate degradation.
The finding establishes that structural economic relationships at all frequencies both short-run dynamics and long-run equilibria break down during currency crises, making any approach relying on stable GDP-CPI interactions inherently vulnerable.

Several constraints shape appropriate interpretation of findings. Sample size limitations (24 annual observations, 5 per regime phase) restrict statistical power for formal hypothesis testing, particularly for period-specific comparisons. We address this through practical significance frameworks integrating multiple evidence sources (test statistics, confidence intervals, effect magnitudes, metric triangulation) rather than relying solely on p-value thresholds but acknowledge that some period-specific findings represent suggestive evidence requiring confirmation with larger samples. Annual data frequency prevents analysis of within-year dynamics and business cycle patterns, focusing inference instead on medium-term regime-level forecasting. This represents a feature rather than pure limitation for our research questions centred on structural regime changes but means findings may not extend to quarterly forecasting applications. Geographic specificity to Zimbabwe's extreme economic trajectory hyperinflation to dollarization to renewed crisis provides exceptional natural experimental variation but raises questions about generalizability to economies experiencing moderate volatility without currency abandonment. However, if models fail during extreme instability, they likely also struggle during moderate instability; our findings thus establish upper bounds on model vulnerabilities. Variable constraints limiting VECM to bivariate specification prevent testing whether richer information sets improve crisis performance, though our finding that even theoretically appropriate GDP-CPI cointegration fails during regime breaks suggests additional variables would not fundamentally alter conclusions about cointegration.

A critical limitation is that cointegration was tested once using the full sample rather than re-tested at each expanding window. This assumes stable long-run relationships throughout the forecast period an assumption likely violated during Zimbabwe's 2019 currency crisis. Ideally, rolling cointegration tests would identify when VECM becomes inappropriate, but the 24-observation sample makes this infeasible. This constraint means VECM failures during 2019-2023 may partly reflect specification issues rather than purely model inadequacy.

Future research should address key limitations through quarterly data analysis enabling within-year dynamics examination and rolling-window cointegration testing, multivariate extensions incorporating monetary and fiscal variables via FAVAR or Bayesian VAR approaches, systematic comparison against regime-switching models and machine learning methods (LSTM, ensemble techniques) specifically designed for structural breaks, real-time forecasting exercises confronting data revisions and operational constraints, and cross-country panel analysis across African economies experiencing currency crises to test generalizability beyond Zimbabwe's context. These extensions would advance both theoretical understanding and practical forecasting guidance for volatile developing economies.

Overall, this research shows that forecasting GDP growth in Zimbabwe is strongly affected by changes in economic stability. ARIMA was the most dependable model across different periods, while Holt’s and VECM worked only under specific conditions. The study highlights the importance of choosing forecasting methods that match current economic realities rather than expecting one model to always perform best. It also shows that reliable forecasting depends just as much on economic stability as on the statistical methods that are used.

References

Abdul Razak, N.A., Khamis, A. & Abdullah, M.A.A. (2017) ‘ARIMA and VAR modeling to forecast Malaysian economic growth’, Journal of Science and Technology, 9(3), pp. 16–24.

Abonazel, M.R. & Abd-Elftah, A.I. (2019) ‘Forecasting Egyptian GDP using ARIMA models’, Reports on Economics and Finance, 5(1), pp. 35–47. doi: 10.12988/ref.2019.81023.

Alsheheri, G. (2025) ‘Comparative analysis of ARIMA and NNAR models for time series forecasting’, Journal of Applied Mathematics and Physics, 13, pp. 267–280. https://doi.org/10.4236/jamp.2025.131012

Alsinglawi, O., Al Wadi, S., Aladwan, M. & Saleh, M.H. (2022) ‘Predicting Jordanian GDP based on ARIMA modelling’, Italian Journal of Pure and Applied Mathematics, 47, pp. 12–25.

Andrianady, J.R. (2023) Crunching the numbers: A comparison of econometric models for GDP forecasting in Madagascar. MPRA Paper No. 120698. Available at: https://mpra.ub.uni-muenchen.de/120698/
(Accessed: [insert date]).

Atoyebi, S.B., Olayiwola, M.F., Oladapo, J.O. & Oladapo, D.I. (2023) ‘Forecasting currency in circulation in Nigeria using Holt-Winters exponential smoothing method’, South Asian Journal of Social Studies and Economics, 20(1), pp. 25–41. doi: 10.9734/SAJSSE/2023/v20i1689.

Bloom, N. (2014) ‘Fluctuations in uncertainty’, Journal of Economic Perspectives, 28(2), pp. 153–176.

Bonga, W.G. (2020) ‘Understanding diaspora remittances levels in Zimbabwe (2009–2020) including future forecasts using ARIMA technique (2020–2022)’, European Journal of Research Development and Sustainability, 1(2), pp. 10–21.

Box, G.E.P. & Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Box, G.E.P. & Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Bryman, A. (2016) Social Research Methods. 5th edn. Oxford: Oxford University Press.

Collis, J. & Hussey, R. (2014) Business Research: A Practical Guide for Undergraduate and Postgraduate Studies. 4th edn. London: Palgrave Macmillan.

Coorey, S., Clausen, J.R., Funke, N., Muñoz, S. and Ould-Abdallah, B., 2007. Lessons from High Inflation Episodes for Stabilizing the Economy in Zimbabwe. IMF Working Paper WP/07/99. Washington, DC: International Monetary Fund.

De Gooijer, J.G. & Hyndman, R.J. (2006) ‘25 years of time series forecasting’, International Journal of Forecasting, 22(3), pp. 443–473.

Dritsaki, M. & Dritsaki, C. (2021) ‘Comparison of the Holt-Winters exponential smoothing method with ARIMA models: Forecasting GDP per capita post COVID in five EU Balkan countries’, Modern Economy, 12, pp. 1972–1998.

Engle, R.F. & Rangel, J.G. (2004) ‘The spline-GARCH model for low-frequency volatility and its global macroeconomic causes’, The Review of Financial Studies, 21(3), pp. 1187–1222.

Erkekoglu, H., Garang, A.P.M. & Deng, A.S. (2020) ‘Comparative evaluation of forecast accuracies for ARIMA, exponential smoothing and VAR’, International Journal of Economics and Financial Issues, 10(6), pp. 206–216.

Funde, Y. & Damani, A. (2023) ‘Comparison of ARIMA and exponential smoothing models in prediction of stock prices’, The Journal of Prediction Markets, 17(1), pp. 21–38.

Gujarati, D.N., Porter, D.C. & Gunasekar, S. (2015) Basic Econometrics. 5th edn. New Delhi: Tata McGraw-Hill.

Hamilton, J.D. (2011) ‘Calling recessions in real time’, International Journal of Forecasting, 27(4), pp. 1006–1026.

Hanke, S.H. & Kwon, A.K.F. (2009) ‘On the measurement of Zimbabwe’s hyperinflation’, Cato Journal, 29(2), pp. 352–364.

Holt, C.C. (1957) ‘Forecasting seasonals and trends by exponentially weighted averages’, ONR Research Memorandum, 52.

Holt, C.C. (2004) ‘Forecasting seasonals and trends by exponentially weighted moving averages’, International Journal of Forecasting, 20(1), pp. 5–10.

Hyndman, R.J. & Athanasopoulos, G. (2018) Forecasting: Principles and Practice. 2nd edn. OTexts. Available at: https://otexts.com/fpp2/
(Accessed: 26 February 2025).

Jafari-Samimi, A., Shirazi, B. & Fazlollahtabar, H. (2007) ‘A comparison between time series, exponential smoothing and neural network methods to forecast GDP of Iran’, Iranian Economic Review, 12(19).

Jere, S. & Siyanga, M. (2016) ‘Forecasting inflation rate of Zambia using Holt’s exponential smoothing’, Open Journal of Statistics, 6, pp. 363–372. doi: 10.4236/ojs.2016.62031.

Khan, M.S. & Khan, U. (2020) ‘Comparison of forecasting performance with VAR vs ARIMA models using economic variables of Bangladesh’, Asian Journal of Probability and Statistics, 10(2), pp. 33–47.
Kramarenko, V., Engstrom, L., Verdier, G., Fernandez, G., Oppers, S.E., Hughes, R., McHugh, J. and Coats, W., 2010. Zimbabwe: Challenges and Policy Options After Hyperinflation. African Departmental Paper AFR10/03. Washington, DC: International Monetary Fund.

Liu, B. et al. (2016) ‘The prediction of secondary exponential smoothing of the total GDP of Mongolia’, in ICEME 2016 and EBM 2016. ISBN: 978-1-60595-365-6.

Lyu, G., Lyu, X., Sun, N. & Zhao, X. (2023) ‘Application of ARIMA model and exponential smoothing in GDP prediction of the United States’, BCP Business & Management, 38.

Makochekanwa, A. (2016) Zimbabwe to Introduce Bond Notes: Reactions and Perceptions of Economic Agents Within the First Seven Days After the Announcement. MPRA Paper No. 71695. Available at: https://mpra.ub.uni-muenchen.de/71695/
(Accessed: 9 October 2025).

McIndoe, T. (2009) ‘Hyperinflation in Zimbabwe: Money demand, seigniorage and aid shocks’, Draft, 13 March.

Mohamed, A.O. (2022) ‘Modeling and forecasting Somali economic growth using ARIMA models’, Forecasting, 4(4), pp. 1038–1050.

Mondal, P., Shit, L. & Goswami, S. (2014) ‘Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices’, International Journal of Computer Science, Engineering and Applications, 4(2), pp. 13–29.

Mpofu, R.T. (2015) ‘Dollarization and economic development in Zimbabwe: An interrupted time-series analysis’, Risk Governance and Control: Financial Markets and Institutions, 5(4).

Mutale, J. & Murape, C. (2024) ‘Modelling and forecasting the foreign aid–poverty nexus in Zimbabwe: Empirical evidence from a Box–Jenkins ARIMA approach’.

Nkomazana, L. & Niyimbanira, F. (2014) ‘An overview of the economic causes and effects of dollarisation: Case of Zimbabwe’, Mediterranean Journal of Social Sciences, 5(7), pp. 69–83.

Nyamunda, T. (2019) ‘Currencies, the “new dispensation” and the Zimbabwean economic crisis’, Review of African Political Economy, 46(161), pp. 1–16.

Poon, S.-H. (2005) A Practical Guide to Forecasting Financial Market Volatility. Chichester: John Wiley & Sons.

Reserve Bank of Zimbabwe (RBZ) (2016) Press Statement: Introduction of Bond Notes and Coins. Harare: RBZ.

Reserve Bank of Zimbabwe (RBZ) (2019) Monetary Policy Statement: Establishment of the RTGS Dollar. Harare: RBZ.

Sims, C.A. (1980) ‘Macroeconomics and reality’, Econometrica, 48(1), pp. 1–48.

Southall, R. (2017) ‘Bond notes, borrowing, and heading for bust: Zimbabwe’s persistent crisis’, Canadian Journal of African Studies, 51(1), pp. 145–158.

Stock, J.H. & Watson, M.W. (2007) ‘Why has US inflation become harder to forecast?’, Journal of Money, Credit and Banking, 39(s1), pp. 3–33.

Umar, G. & Kilishi, A.A. (2010) ‘Oil price shocks and the Nigeria economy: A variance autoregressive model’, International Journal of Business and Management, 5(8), pp. 41–49.

Vafin, A. (2020) ‘Forecasting macroeconomic indicators for seven major economies using the ARIMA model’, Sage Science Economic Reviews, 3(1).

Winters, P.R. (1960) ‘Forecasting sales by exponentially weighted moving averages’, Management Science, 6(3), pp. 324–342.

World Bank (2019) Zimbabwe Economic Update: The State in the Economy. Washington, DC: World Bank. Available at: https://openknowledge.worldbank.org/
(Accessed: 9 October 2025).

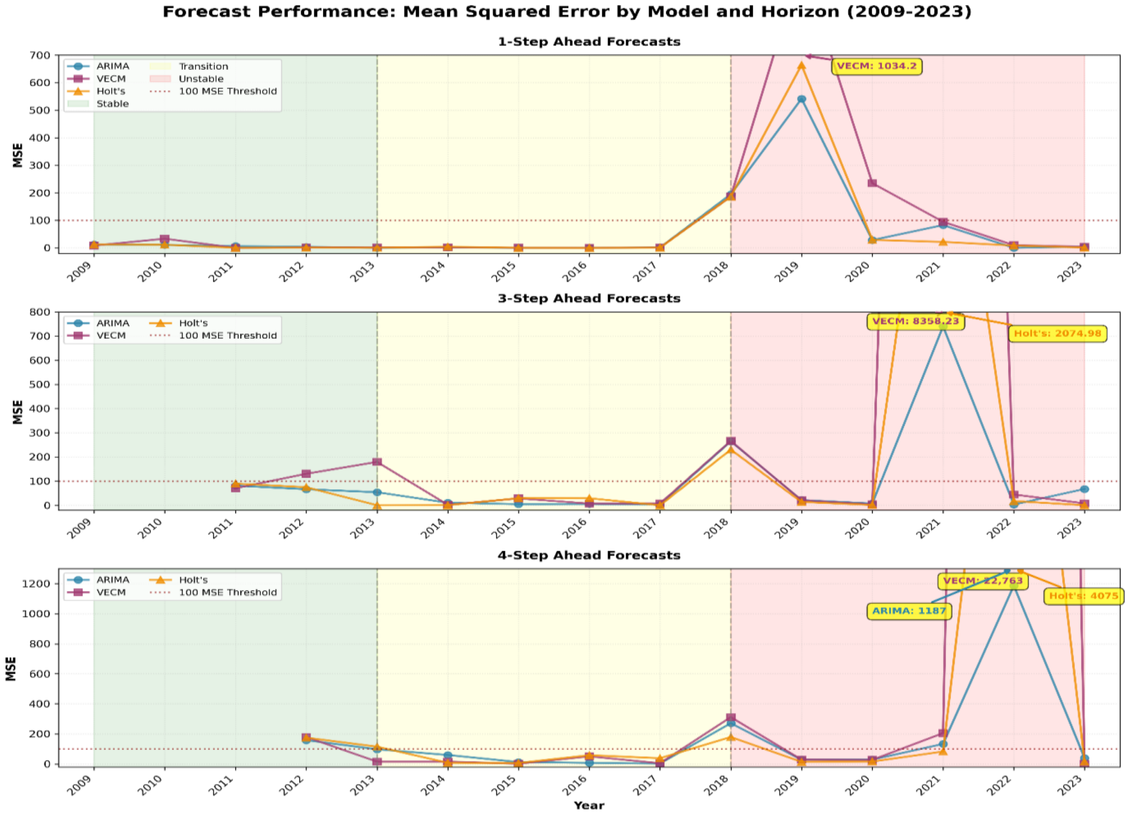
Zhang, H. (2013) Modeling and forecasting regional GDP in Sweden using autoregressive models. Master’s thesis, Dalarna University.

Appendix

Appendix 4.2.2.A: MSE Results

MSE	1 Step			3 step			4 step		
Year	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's
2009	12.23	8.22	12.63						
2010	10.77	33.24	11.4						
2011	6.14	0.09	0.19	81.35	70.68	89.56			
2012	3.56	1.45	0.87	66.42	130.43	74.72	158.1	179.28	174.59
2013	0.01	1.3	0.53	54.2	179.47	0.33	97.45	15.14	114.85
2014	1.39	1.77	3.73	10.64	2.08	0.71	59.42	15.14	7.47
2015	0.01	0.34	0.3	4.92	28.39	29.34	12.4	1.48	6.01
2016	0.07	0.11	0	5.62	7.26	29.11	7.17	49.89	57.92
2017	1.24	1.14	0.82	2.93	7.52	0.24	2.76	4.45	37.98
2018	194.69	185.92	186.54	264.64	266.38	230.26	271.23	311.55	179.08
2019	541.66	1034.2	664.14	21.61	20.34	13.84	27.79	28.27	13.65
2020	28.15	235	28.88	7.57	4.19	1.36	28.73	26.88	15.81
2021	82.64	94.7	21.59	740.2	8358.23	2074.98	133.1	205.14	82.44
2022	0.1	9.66	8.41	3.13	44.71	17.47	1187.19	22763.16	4075.16
2023	4.49	3.95	1.07	67.68	7.89	0.46	38.81	0.54	15.37

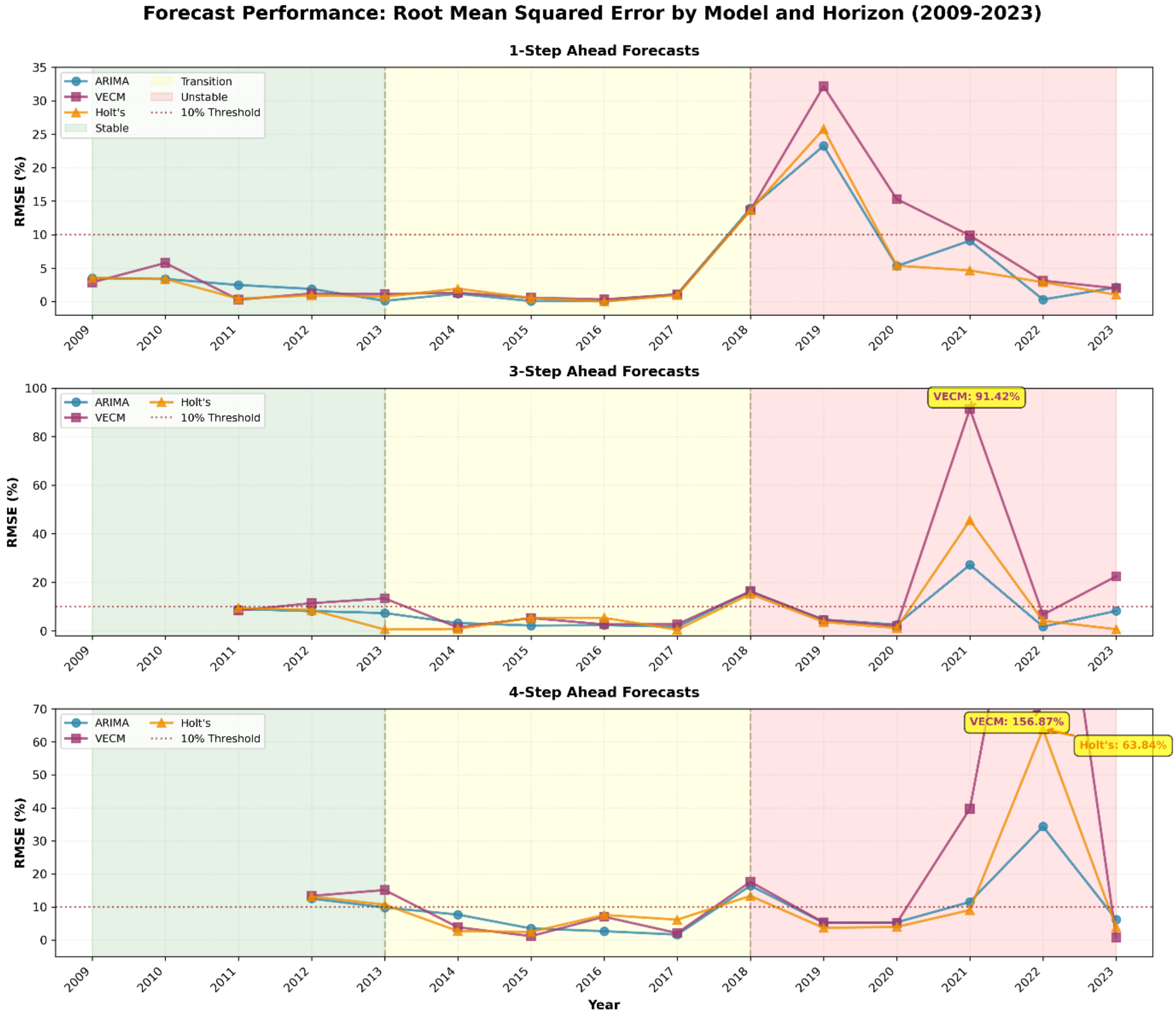
Appendix 4.2.2.B: Forecast Performance: MSE by Model and Horizon



Appendix 4.2.2.C: RMSE Results

RMSE	1 Step			3 step			4 step		
Year	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's
2009	3.5	2.87	3.55						
2010	3.38	5.77	3.38						
2011	2.48	0.29	0.44	9.02	8.41	9.46			
2012	1.89	1.2	0.93	8.15	11.42	8.64	12.57	13.38	13.21
2013	0.11	1.14	0.73	7.36	13.39	0.58	9.87	15.14	10.72
2014	1.18	1.33	1.93	3.26	1.44	0.84	7.71	3.89	2.73
2015	0.08	0.58	0.54	2.22	5.32	5.42	3.52	1.18	2.45
2016	0.08	0.33	0.01	2.37	2.69	5.4	2.68	7.06	7.61
2017	1.11	1.07	0.96	1.71	2.74	0.49	1.66	2.12	6.16
2018	13.95	13.64	13.66	16.27	16.32	15.18	16.47	17.65	13.38
2019	23.27	32.16	25.77	4.65	4.51	3.72	5.27	5.32	3.69
2020	5.31	15.33	5.37	2.57	2.04	1.17	5.36	5.18	3.98
2021	9.09	9.88	4.65	27.21	91.42	45.55	11.54	39.77	9.08
2022	0.31	3.12	2.9	1.77	6.69	4.18	34.37	156.87	63.84
2023	2.12	1.99	1.03	8.23	22.47	0.68	6.23	0.73	3.92

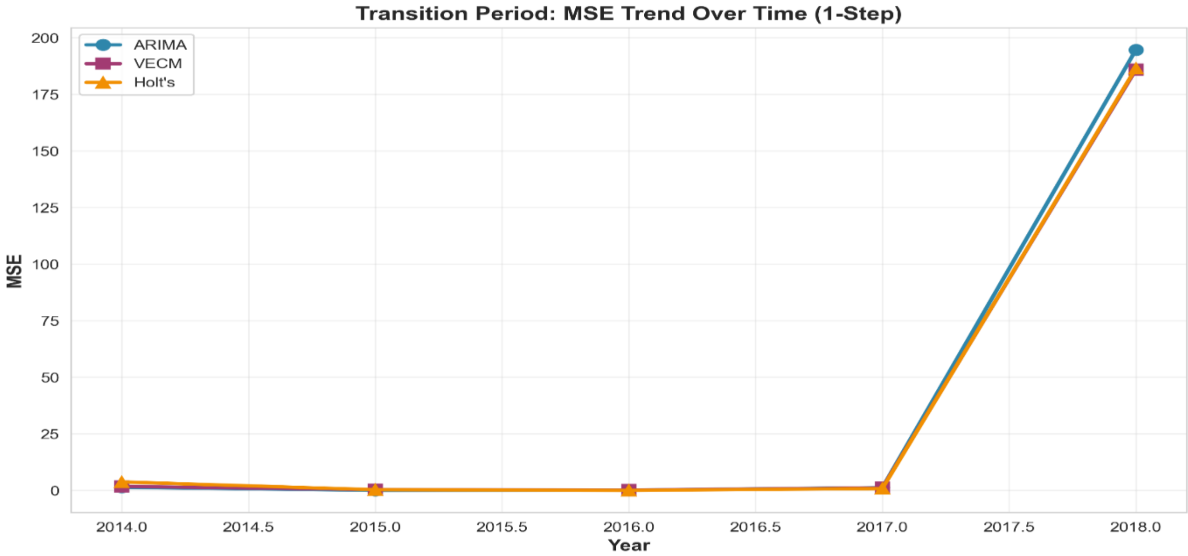
Appendix 4.2.2.D: Forecast Performance: RMSE by Model and Horizon



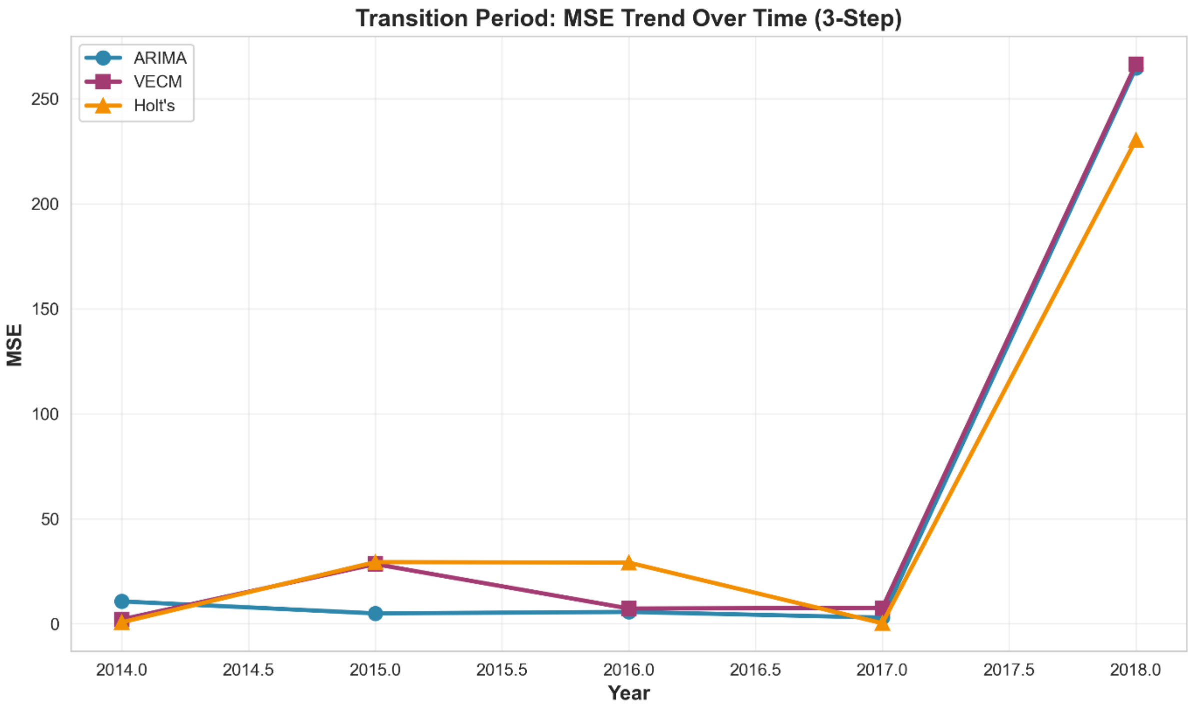
Appendix 4.3.2.A: MSE Transition Results

MSE	1 Step			3 step			4 step		
Year	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**
2014	1.39	1.77	3.73	10.64	2.08	0.71	59.42	15.14	7.47
2015	0.01	0.34	0.3	4.92	28.39	29.34	12.4	1.48	6.01
2016	0.07	0.11	0	5.62	7.26	29.11	7.17	49.89	57.92
2017	1.24	1.14	0.82	2.93	7.52	0.24	2.76	4.45	37.98
2018	194.69	185.92	186.54	264.64	266.38	230.26	271.23	311.55	179.08

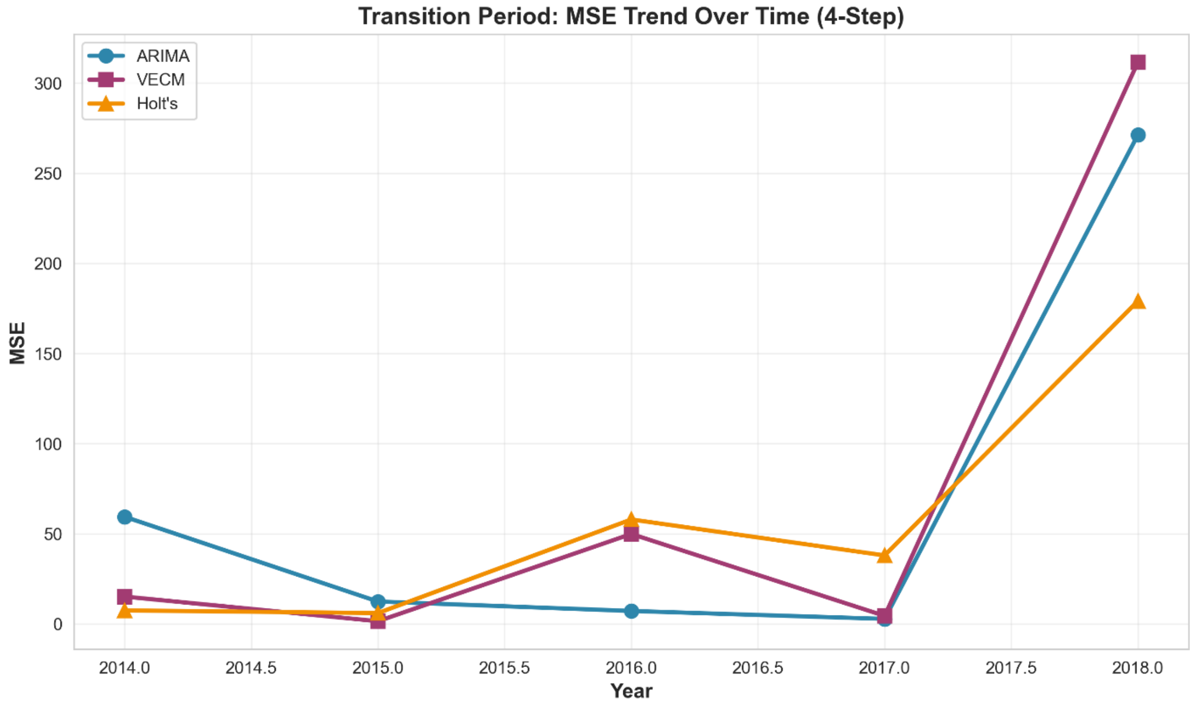
Appendix 4.3.2.B: Transition Period: MSE Trend Over time (1-step)



Appendix 4.3.2.C: Transition Period: MSE Trend Over Time (3-step)



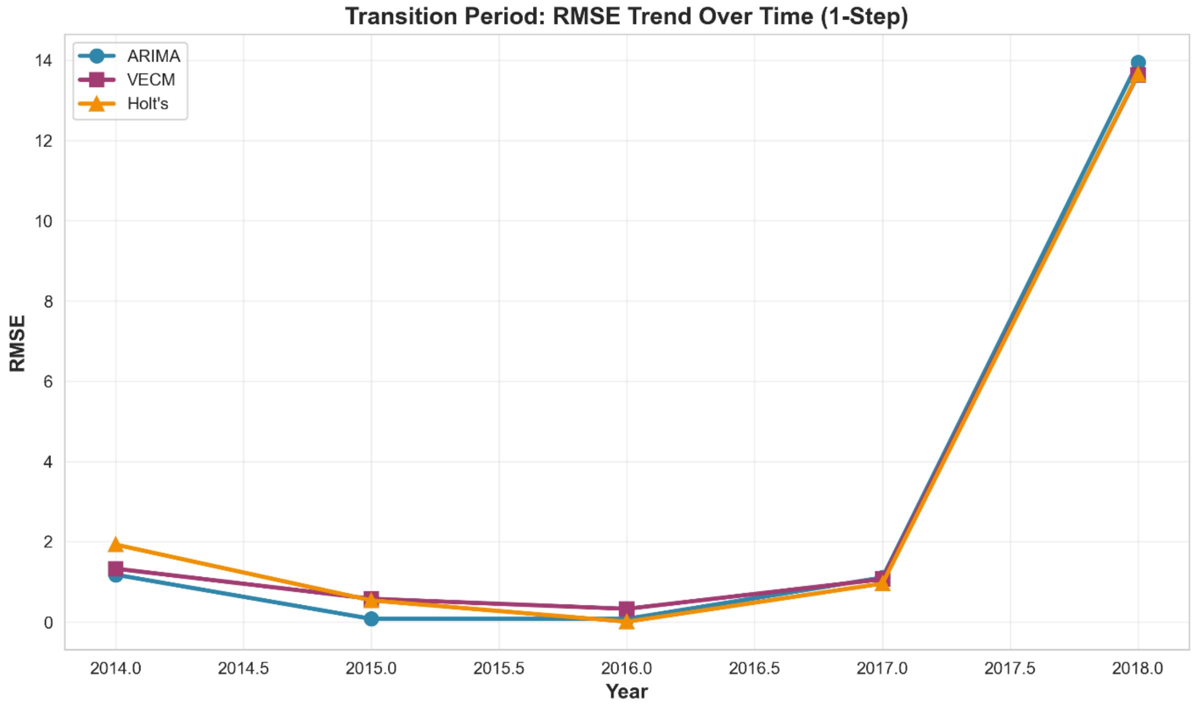
Appendix 4.3.2.D: Transition Period: MSE Trend Over Time (4-step)



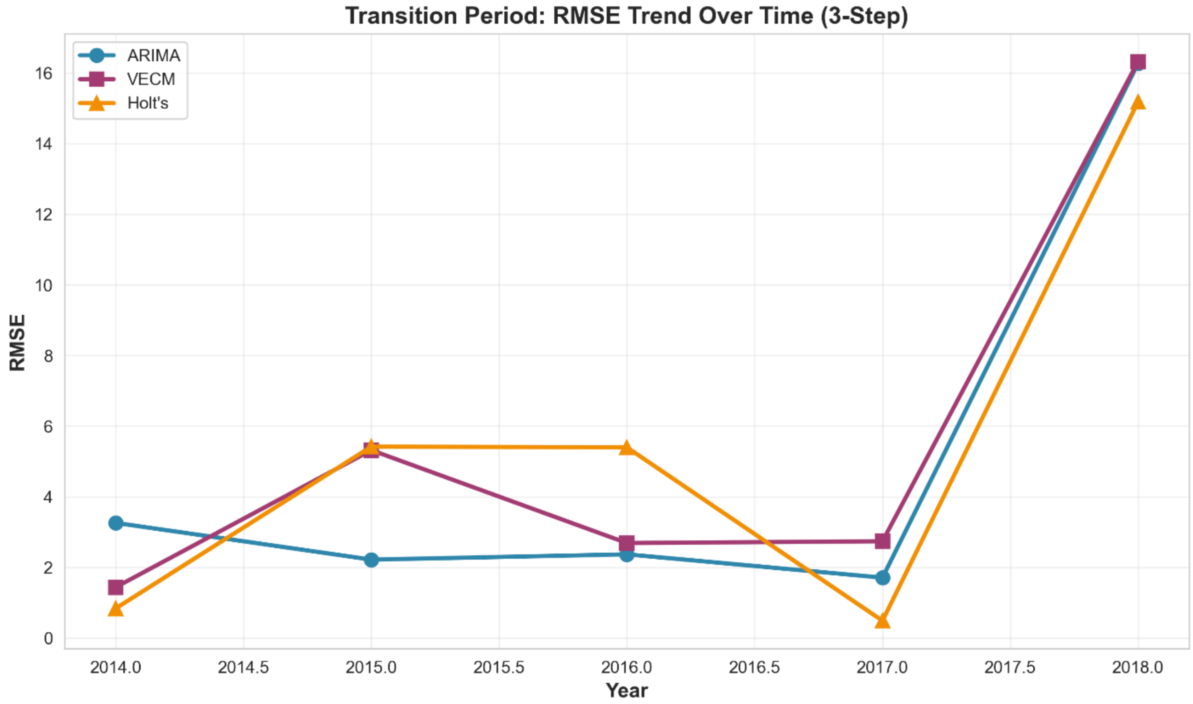
Appendix 4.3.2.E: RMSE transition Period

RMSE	1 Step			3 step			4 step		
Year	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's
2014	1.18	1.33	1.93	3.26	1.44	0.84	7.71	3.89	2.73
2015	0.08	0.58	0.54	2.22	5.32	5.42	3.52	1.18	2.45
2016	0.08	0.33	0.01	2.37	2.69	5.4	2.68	7.06	7.61
2017	1.11	1.07	0.96	1.71	2.74	0.49	1.66	2.12	6.16
2018	13.95	13.64	13.66	16.27	16.32	15.18	16.47	17.65	13.38

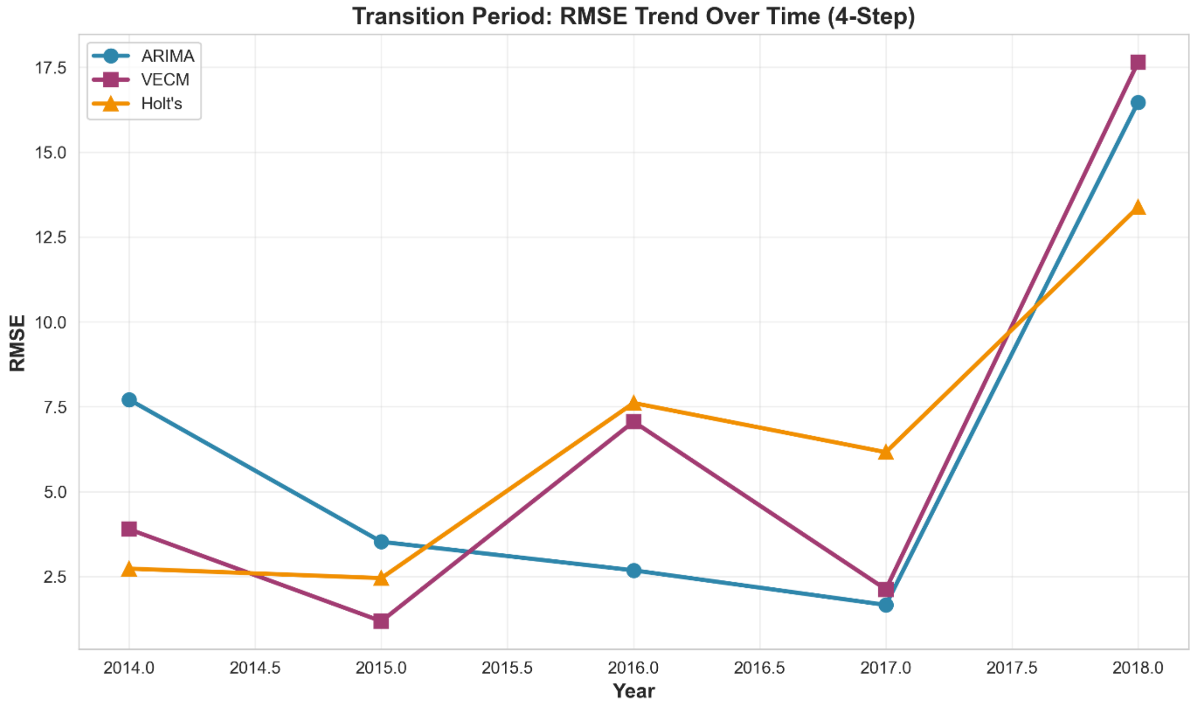
Appendix 4.3.2.E1: Transition Period: RMSE Trend Over Time (1-Step)



Appendix 4.3.2.E2: Transition Period: RMSE Trend Over Time (3-step)



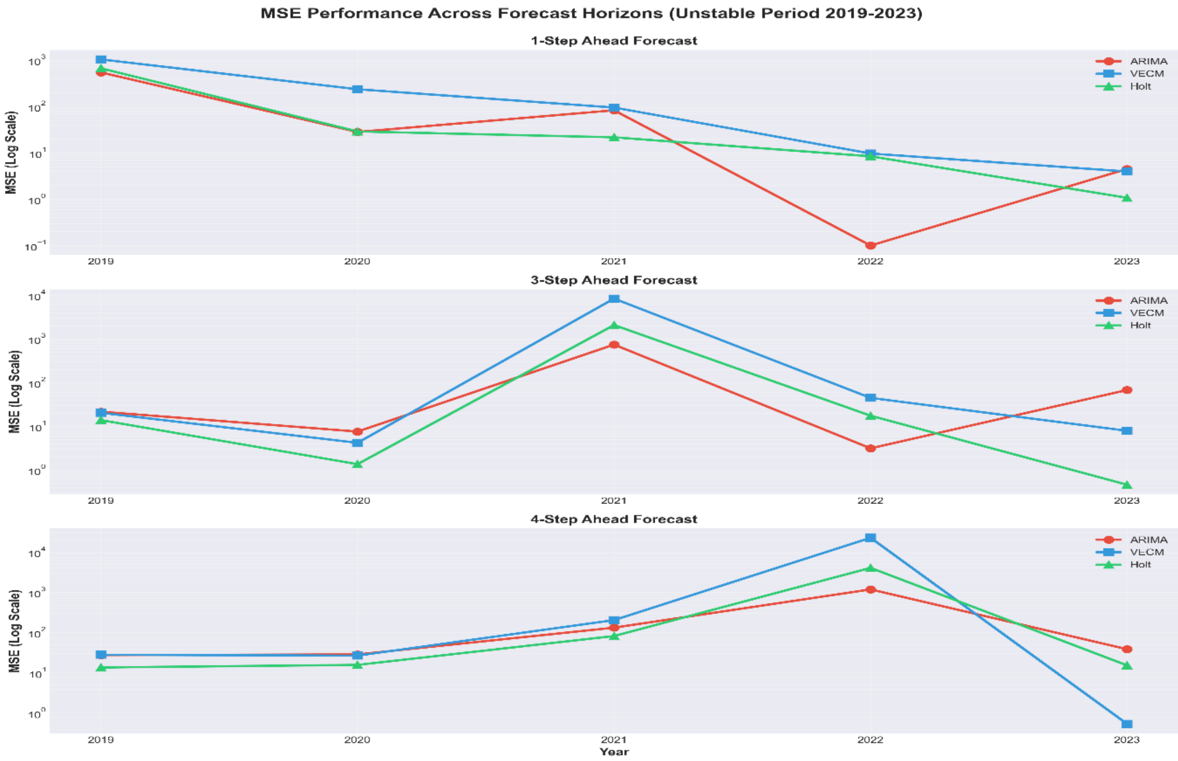
Appendix 4.3.2.E3: Transition Period: RMSE Trend Over Time (4-step)



Appendix 4.3.3.A1: MSE Unstable Results

MSE	1 Step			3 step			4 step		
Year	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**
2019	541.66	1034.2	664.14	21.61	20.34	13.84	27.79	28.27	13.65
2020	28.15	235	28.88	7.57	4.19	1.36	28.73	26.88	15.81
2021	82.64	94.7	21.59	740.2	8358.23	2074.98	133.1	205.14	82.44
2022	0.1	9.66	8.41	3.13	44.71	17.47	1187.19	22763.16	4075.16
2023	4.49	3.95	1.07	67.68	7.89	0.46	38.81	0.54	15.37

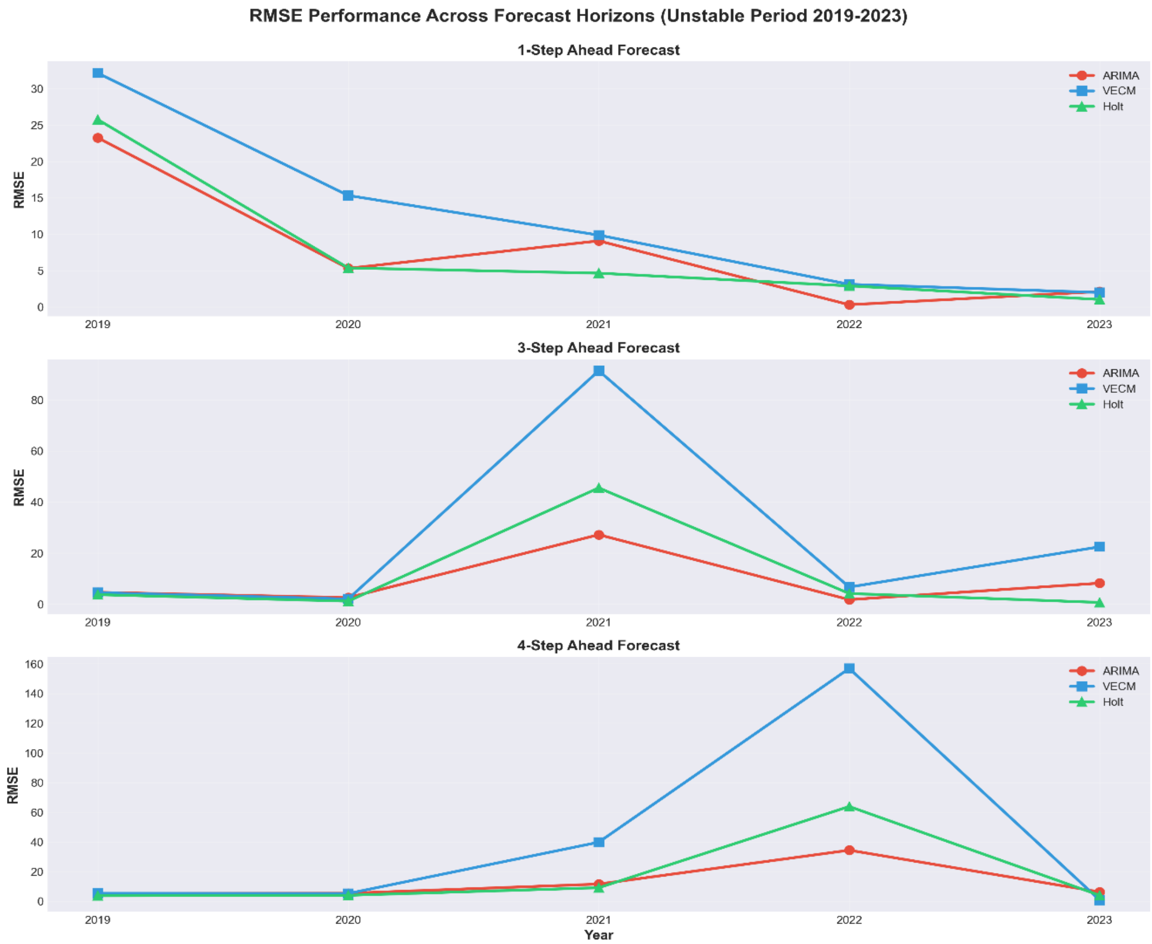
Appendix 4.3.3.A2: MSE Performance Across Forecast Horizons (Unstable Period 2019-2023)



Appendix 4.3.3.B1: RMSE Unstable Results

RMSE	1 Step			3 step			4 step		
Year	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's	ARIMA	VECM	Holt's
2019	23.27	32.16	25.77	4.65	4.51	3.72	5.27	5.32	3.69
2020	5.31	15.33	5.37	2.57	2.04	1.17	5.36	5.18	3.98
2021	9.09	9.88	4.65	27.21	91.42	45.55	11.54	39.77	9.08
2022	0.31	3.12	2.9	1.77	6.69	4.18	34.37	156.87	63.84
2023	2.12	1.99	1.03	8.23	22.47	0.68	6.23	0.73	3.92

Appendix 4.3.3.B2: RMSE Performance Across Forecast Horizons (Unstable Period 2019-2023)



Appendix 4.3.1.A1: MSE stable Results

MSE	1 Step			3 step			4 step		
Year	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**
2009	12.23	8.22	12.63						
2010	10.77	33.24	11.4						
2011	6.14	0.09	0.19	81.35	70.68	89.56			
2012	3.56	1.45	0.87	66.42	130.43	74.72	158.1	179.28	174.59
2013	0.01	1.3	0.53	54.2	179.47	0.33	97.45	15.14	114.85

Appendix 4.3.1.A2: RMSE stable Results

RMSE	1 Step			3 step			4 step		
Year	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**	**ARIMA**	**VECM**	**Holt's**
2009	3.5	2.87	3.55						
2010	3.38	5.77	3.38						
2011	2.48	0.29	0.44	9.02	8.41	9.46			
2012	1.89	1.2	0.93	8.15	11.42	8.64	12.57	13.38	13.21
2013	0.11	1.14	0.73	7.36	13.39	0.58	9.87	15.14	10.72

Appendix 4.5.3.A1: 1-step stable Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	7	7	6
P-value	1	1	0.8125
Effect Size(rank-biserial)	0.0667	0.0667	-0.2
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.3.A2: 1-step transition Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	5	6	6
P-value	0.625	0.8125	0.8125
Effect Size(rank-biserial)	0.3333	0.2	-0.2
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.3.A3: 1-step unstable Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	1	7	0.00
P-value	0.1250	1	0.0625
Effect Size (rank-biserial)	0.8667	0.0667	-1
Significance	Not Significant	Not Significant	Marginally Significant

Appendix 4.5.3.B1: 3-step stable Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	1	3	1
P-value	0.5	1	0.5
Effect Size (rank-biserial)	0.6667	0.0	-0.6667
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.3.B2: 3-step transition Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	4	6	6
P-value	0.4375	0.8175	0.8125
Effect Size (rank-biserial)	0.4667	0.2	-0.2
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.3.B3: 3-step unstable Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	6	7	0
P-value	0.8175	1	0.0625
Effect Size (rank-biserial)	0.2	0.0667	-1
Significance	Not Significant	Not Significant	Marginally Significant

Appendix 4.5.3.C1: 4-step stable Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	1	0.0	1
P-value	1	0.5	1
Effect Size (rank-biserial)	-0.3333	1	0.3333
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.3.C2: 4-step transition Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	7	7	6
P-value	1	1	0.8125
Effect Size (rank-biserial)	-0.0667	-0.0667	0.2
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.3.C3: 4-step unstable Wilcoxon signed rank results

	ARIMA vs VECM	ARIMA vs Holts	VECM vs Holts
Test statistic	6	5	3
P-value	0.8125	0.625	0.3125
Effect Size (rank-biserial)	0.2	-0.333	-0.6
Significance	Not Significant	Not Significant	Not Significant

Appendix 4.5.4.A: 1-step sign test

Period	Model Comparison	ARIMA wins	VECM wins	Holts wins	Ties	Win %	P-value	significance
Stable	ARIMA vs VECM	2	3	-	0	ARIMA 40%	1	Not Significant
Stable	ARIMA vs Holts	3	-	2	0	ARIMA 60%	1	Not significant

Stable	VECM vs Holts	-	2	3	0	VECM 40%	1	Not significant
Transition	ARIMA vs VECM	3	2	-	0	ARIMA 60%	1	Not significant
Transition	ARIMA vs Holts	2	-	3	0	ARIMA 40%	1	Not significant
Transition	VECM vs Holts	-	2	3	0	VECM 40%	1	Not significant
Unstable	ARIMA vs VECM	4	1	-	0	ARIMA 80%	0.375	Not significant
Unstable	ARIMA vs Holts	3	-	2	0	ARIMA 60%	1	Not significant
Unstable	VECM vs Holts	-	0	5		VECM 0%	0.0625	Marginally significant

Appendix 4.5.4.B: 3-step sign test

Period	**Model Comparison**	**ARIMA wins**	**VECM wins**	**Holts wins**	**Ties**	**Win %**	**P-value**	**significance**
Stable	ARIMA vs VECM	2	1	-	0	ARIMA 66.7%	1	Not Significant
Stable	ARIMA vs Holts	2	-	1	0	ARIMA 66.7%	1	Not significant
Stable	VECM vs Holts	-	1	2	0	VECM 33.3%	1	Not significant
Transition	ARIMA vs VECM	4	1	-	0	ARIMA 80%	0.375	Not significant
Transition	ARIMA vs Holts	2	-	3	0	ARIMA 40%	1	Not significant
Transition	VECM vs Holts	-	2	3	0	VECM 40%	1	Not significant
Unstable	ARIMA vs VECM	2	3	-	0	ARIMA 40%	1	Not significant
Unstable	ARIMA vs Holts	2	-	3	0	ARIMA 40%	1	Not significant
Unstable	VECM vs Holts	-	0	5		VECM 0%	0.0625	Marginally significant

Appendix 4.5.4.C: 4-step sign test

Period	Model Comparison	ARIMA wins	VECM wins	Holts wins	Ties	Win %	P-value	significance
Stable	ARIMA vs VECM	1	1	-	0	ARIMA 50%	1.5	Not Significant
Stable	ARIMA vs Holts	2	-	0	0	ARIMA 100%	0.5	Not significant
Stable	VECM vs Holts	-	1	1	0	VECM 50%	1	Not significant
Transition	ARIMA vs VECM	3	2	-	0	ARIMA 60%	0.375	Not significant
Transition	ARIMA vs Holts	2	-	3	0	ARIMA 40%	1	Not significant
Transition	VECM vs Holts	-	3	2	0	VECM 60%	1	Not significant
Unstable	ARIMA vs VECM	3	2	-	0	ARIMA 60%	1	Not significant
Unstable	ARIMA vs Holts	1	-	4	0	ARIMA 20%	0.375	Not significant
Unstable	VECM vs Holts	-	1	4		VECM 20%	0.375	Not significant

Appendix 4.5.4.D: Summary of Non-Parametric test results (Wilcoxon rank and sign tests)

Period	Horizon	Model Comparison	Wilcoxon Signed-Rank Test	Sign Test	Interpretation
Stable	**1-Step**	ARIMA vs VECM	ARIMA better (p = 1.00) – Not Sig	ARIMA wins 40% – Not Sig	Models equivalent
		ARIMA vs Holt's	ARIMA better (p = 1.00) – Not Sig	ARIMA wins 60% – Not Sig	Slight ARIMA edge

		VECM vs Holt's	VECM better (p = 0.81) – Not Sig	VECM wins 40% – Not Sig	Comparable results
	3-Step	ARIMA vs VECM	ARIMA better (p = 0.50) – Not Sig	ARIMA wins 67% – Not Sig	ARIMA minor advantage
		ARIMA vs Holt's	ARIMA better (p = 1.00) – Not Sig	ARIMA wins 67% – Not Sig	Consistent with above
		VECM vs Holt's	Holt's better (p = 0.50) – Not Sig	VECM wins 33% – Not Sig	Not significant
	4-Step	ARIMA vs VECM	VECM better (p = 1.00) – Not Sig	ARIMA wins 50% – Not Sig	No difference
		ARIMA vs Holt's	ARIMA better (p = 0.50) – Not Sig	ARIMA wins 100% – Not Sig	ARIMA slightly dominant
		VECM vs Holt's	VECM better (p = 1.00) – Not Sig	VECM wins 50% – Not Sig	Models equivalent
Transition	**1-Step**	ARIMA vs VECM	ARIMA better (p = 0.63) – Not Sig	ARIMA wins 60% – Not Sig	No meaningful gap
		ARIMA vs Holt's	ARIMA better (p = 0.81) – Not Sig	ARIMA wins 40% – Not Sig	Comparable
		VECM vs Holt's	VECM better (p = 0.81) – Not Sig	VECM wins 40% – Not Sig	Comparable
	3-Step	ARIMA vs VECM	ARIMA better (p = 0.44) – Not Sig	ARIMA wins 80% – Not Sig	ARIMA minor advantage
		ARIMA vs Holt's	ARIMA better (p = 0.99) – Not Sig	ARIMA wins 40% – Not Sig	Not significant
		VECM vs Holt's	Holt's better (p = 0.98) – Not Sig	VECM wins 40% – Not Sig	Comparable
	4-Step	ARIMA vs VECM	Mixed (p = 1.00) – Not Sig	ARIMA wins 60% – Not Sig	Equivalent
		ARIMA vs Holt's	ARIMA better (p = 1.00) – Not Sig	ARIMA wins 40% – Not Sig	Similar performance
		VECM vs Holt's	VECM better (p = 0.81) – Not Sig	VECM wins 60% – Not Sig	Similar performance
Unstable	**1-Step**	ARIMA vs VECM	ARIMA better (p = 0.13) – Not Sig	ARIMA wins 80% – Not Sig	ARIMA minor advantage
		ARIMA vs Holt's	ARIMA better (p = 1.00) – Not Sig	ARIMA wins 60% – Not Sig	No significant difference
		VECM vs Holt's	Holt's better (p = 0.06) – *Marginally Sig*	Holt's wins 100% – *Marginally Sig*	Holt's performs better
	3-Step	ARIMA vs VECM	ARIMA better (p = 0.82) – Not Sig	ARIMA wins 40% – Not Sig	Similar
		ARIMA vs Holt's	ARIMA better (p = 1.00) – Not Sig	ARIMA wins 40% – Not Sig	Similar
		VECM vs Holt's	Holt's better (p = 0.06) – *Marginally Sig*	Holt's wins 100% – *Marginally Sig*	Holt's advantage persists
	4-Step	ARIMA vs VECM	ARIMA better (p = 0.81) – Not Sig	ARIMA wins 60% – Not Sig	Comparable
		ARIMA vs Holt's	ARIMA better (p = 0.63) – Not Sig	ARIMA wins 20% – Not Sig	Comparable

		VECM vs Holt’s	VECM better (p = 0.31) – Not Sig	VECM wins 20% – Not Sig	Comparable