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

0009-0009-8708-2224

Supervisor: Dr Banjo Aderemi

November 2025

Abstract

This paper examines how traditional time series models perform under extreme economic volatility using Zimbabwe's GDP trajectory (2000-2023) as a natural experiment encompassing hyperinflation, dollarisation, currency reintroduction, and pandemic shocks.

We compare ARIMA, Holt's Exponential Smoothing, and VECM across three economic regimes (stable, unstable, transitional) using expanding window forecasts at one-step, three-step, and four-step horizons. Model accuracy is evaluated through MAE, MSE, RMSE, and MAPE. Rolling cointegration analysis examines assumption stability across forecast windows

No single model dominates across all conditions. ARIMA demonstrates superior overall reliability and crisis resilience, while Holt's Exponential Smoothing excels in short-term stable forecasts. VECM fails catastrophically during instability, with errors exceeding 400% MAPE. Critically, full sample cointegration evidence (trace statistic: 20.59) masked complete absence of equilibrium relationships in operational forecast windows (0% cointegration detection during 2009-2018). Forecast accuracy deteriorates significantly with longer horizons, particularly during unstable periods where multi-year forecasts become unreliable. Results demonstrate that structural model sophistication amplifies rather than reduces errors when underlying assumptions fail.

Small sample size (24 annual observations) limits statistical power. Annual data frequency constrains short-term dynamics capture. Geographic specificity to extreme volatility case may limit generalizability to moderately volatile economies.

Forecasters should adopt regime-conditional model selection rather than relying on single-model approaches. Implementing rolling assumption verification is essential, as full-sample diagnostics can be misleading. Simpler models may outperform complex alternatives during structural breaks. Forecasting in unstable economies has inherent limits that no model fully overcomes.

Systematic comparison of traditional forecasting models across documented extreme economic regimes. Demonstrates critical gap between full-sample and operational cointegration stability. Provides evidence-based model selection framework for developing economies experiencing structural volatility.

Table of Contents

Chapter 1: Introduction

Accurate GDP forecasting is essential for economic policy formulation, particularly in developing economies where volatility constrains planning horizons (Ning et al., 2010). Yet despite extensive theoretical development, significant gaps remain in understanding how traditional time series models perform under extreme economic volatility precisely when reliable forecasts become most critical (Hamilton, 2011).

Traditional forecasting models ARIMA, VECM, and Exponential Smoothing demonstrate robust performance in stable environments where statistical assumptions hold over time (De Gooijer & Hyndman, 2006). However, their reliability during structural breaks and regime transitions remains inadequately understood. Existing research evaluates models either in stable contexts or through simulated shocks, leaving a critical gap: how do these models actually perform across genuine regime cycles?

Zimbabwe as Natural Experiment: Between 2000 and 2023, Zimbabwe experienced hyperinflation (79.6 billion percent in 2008), dollarization (2009), bond note introduction (2016), RTGS dollar reintroduction (2019), and COVID-19 shocks. This trajectory encompasses three post-dollarization regimes: stable dollarization (2009-2013, growth 23.74%, volatility 12.39%), transition (2014-2018, growth 16.45%, volatility 28.63%), and renewed instability (2019-2023, growth 1.21%, volatility 23.34%).

Research Questions:

RQ1: Which models (ARIMA, Holt's, VECM) provide more accurate forecasts during stability, instability, and transition?

RQ2: How does forecasting accuracy vary between short-term (1-step) and long-term (4-step) predictions across different regimes?

Three Contributions:

1. systematic comparison using real-world data across documented extreme regimes rather than simulations
2. Rolling cointegration analysis showing full-sample tests can mislead—we find 0% cointegration during 2009-2018 operational windows despite significant full-sample evidence (trace: 20.59)
3. Regime-conditional model selection framework providing actionable decision rules for practitioners

Key Preview: No universal winner exists. Holts excels during stability (1.1% MAE), ARIMA dominates crises (10.2% MAE), while VECM fails catastrophically during instability (150.87% MAE, 462% MAPE).

Chapter 2: Literature Review

2.1. Forecasting under volatile environments

Economic forecasting in volatile environments confronts fundamental challenges beyond technical model specification. Hamilton (2011) establishes that economic time series exhibit regime-switching behaviour where distinct data-generating processes operate during stable versus unstable periods, fundamentally altering the statistical properties forecasting models rely upon. Stock and Watson (2007) document that forecast accuracy deteriorates significantly during high-volatility periods due to model misspecification, parameter instability, and structural shifts in economic fundamentals. Bloom (2014) demonstrates this deterioration occurs systematically as volatility increases, with traditional models particularly vulnerable to extreme shocks characterizing developing economy crises.

Traditional time series models embed assumptions that become binding under volatility. ARIMA models assume stationarity that statistical properties remain constant over time an assumption violated during hyperinflation, currency reforms, or abrupt policy shifts (Poon, 2005). Vector Error Correction Models depend on stable cointegration relationships that may collapse during crises. Exponential smoothing methods lack theoretical grounding to distinguish meaningful structural shifts from temporary fluctuations (Hyndman & Athanasopoulos, 2018).

Empirical evidence confirms these theoretical vulnerabilities. Engle and Rangel (2004) show volatility varies dramatically across countries and time periods, particularly in emerging markets, while traditional models assume volatility reverts to a constant mean. Rossi (2011) demonstrates that rolling-window estimations reveal substantial fluctuations in forecast regression coefficients, confirming predictive relationships shift systematically during economic shocks. This instability renders conventional diagnostic tests like the Mincer-Zarnowitz framework unreliable (Giacomini & Rossi, 2010).

While the literature emphasizes methods designed for instability fluctuation tests and adaptive frameworks these remain underutilized in developing economies (Rossi & Sekhposyan, 2011). Critically, existing research examines either stable periods or simulated shocks rather than genuine regime cycles. No study systematically compares model performance across a complete regime cycle from stability through crisis to stabilization using real-world data from an economy experiencing all three phases.

2.2. ARIMA Models: Performance and Limitations in Volatile Contexts

ARIMA models demonstrate robust performance in stable economic environments, achieving forecast accuracies exceeding 85% when underlying data-generating processes remain constant (Mondal et al., 2014). Empirical applications across diverse contexts including China's GDP forecasting (Wang & Wang, 2011) and South African fuel price prediction (Atoyebi et al., 2023) confirm ARIMA's reliability during periods of consistency. The model's theoretical strength lies in its parsimony: combining autoregressive and moving average components with differencing to achieve stationarity enables complex temporal pattern capture using minimal parameters.

This parsimony becomes problematic during structural instability. Poon (2005) argues ARIMA models struggle with economic volatility because they assume stationarity can be achieved through differencing an assumption violated during hyperinflation or currency reforms where economic relationships transform continuously. De Gooijer and Hyndman (2006) suggest ARIMA may fail entirely when structural breaks define an economy, as differencing techniques presume transformed series achieve stable statistical properties.

Abonazel and Abd-Elftah (2019) demonstrate these challenges through Egypt's GDP analysis, where achieving stationarity required second-order differencing, yet transformation does not guarantee forecast reliability if new regime shifts occur. Mondal et al. (2014) found ARIMA exhibited larger predictive accuracy standard deviations for volatile sectors, suggesting performance deteriorates as structural complexity increases. Critically, studies demonstrating ARIMA effectiveness examine relative stability periods rather than genuine regime changes currency abandonment, hyperinflationary collapse, or complete economic restructuring most relevant for volatile developing economies.

2.3. Vector Autoregression Models: Multivariate forecasting under instability

Vector Error Correction Models incorporate economic relationships through cointegration constraints, exploiting long-run equilibrium structures between non-stationary variables to enhance forecast accuracy. Khan and Khan (2020) demonstrate this advantage empirically, showing VAR/VECM models consistently outperform ARIMA when forecasting strongly correlated economic indicators in Bangladesh. Abdul Razak et al. (2017) provide compelling Malaysian evidence, where VAR approaches achieved substantially lower MAPE values across multiple indicators during 1998 to 2016, encompassing the Asian Financial Crisis aftermath.

The theoretical appeal rests on distinguishing short-run fluctuations from long-run equilibrium relationships, potentially providing stability during temporary shocks. However, this advantage depends critically on cointegration relationships identified in estimation samples persisting during forecast periods. The literature provides extensive guidance on testing cointegration using full-sample data through Johansen's trace tests and Engle-Granger procedures, yet remarkably little attention addresses whether these relationships remain stable across time, particularly during regime transitions.

Studies demonstrating VECM superiority examine stable or moderately volatile periods within intact institutional frameworks. Abdul Razak et al. covered post-crisis recovery rather than fundamental regime change, while Khan and Khan analysed routine macroeconomic fluctuations instead of currency abandonment or hyperinflation. Andrianady (2023) shows VAR underperformance in Madagascar's developing economy, suggesting theoretical advantages may not materialize in structurally unstable environments. Critically, no existing study examines whether cointegration relationships identified through full-sample testing persist during specific forecast generation windows, assuming temporal stability most likely violated during regime shifts.

2.4. Exponential Smoothing: Adaptive Approaches to economic volatility

Exponential smoothing methods offer computational simplicity and intuitive appeal through emphasizing recent observations. Liu et al. (2016) highlight key advantages including low computational cost, exceptional adaptability across diverse applications, and efficiency in incorporating new information. The preference for recent data theoretically provides responsiveness to changing conditions, potentially advantageous in volatile environments. Lyu et al. (2023) demonstrate this potential, finding Holt-Winters exponential smoothing outperformed

ARIMA for US GDP forecasting, achieving smaller mean errors and effectively capturing long-term cyclical patterns.

However, exponential smoothing's empirical success lacks strong theoretical foundation for understanding when the method will succeed or fail. Unlike ARIMA models with explicit stochastic process assumptions or VECM approaches incorporating economic theory through cointegration constraints, exponential smoothing relies primarily on ad hoc weighting schemes. Hyndman and Athanasopoulos (2018) acknowledge these methods prove effective for short-term forecasting due to heavy weighting of recent observations but may underperform during sudden economic shocks, particularly in hyperinflationary environments where exponential growth patterns exceed adaptive capacity.

While comparative studies show exponential smoothing can match or exceed ARIMA performance in specific contexts, conditions determining relative performance remain underspecified. The method's responsiveness to recent data could prove beneficial during regime shifts by rapidly incorporating new information or detrimental by overreacting to temporary shocks, leaving practitioners without principled bases for determining appropriate methodology under volatility.

2.5. Research Gap

The literature reveals three critical gaps constraining practical forecasting in volatile developing economies. First, existing studies evaluate model performance using full-sample data without examining whether foundational assumptions hold during actual forecast windows. Standard practice tests for stationarity and cointegration using all available observations, then applies specifications throughout forecast periods, presuming temporal stability that regime shifts violate. No research implements rolling assumption verification to determine whether cointegration relationships or stationarity conditions exist during operational forecast periods rather than merely historical data.

Second, the literature lacks systematic comparison across complete regime cycles. Studies show ARIMA performs well during stability (Mondal et al., 2014), VECM excels when correlations are strong (Khan & Khan, 2020), and exponential smoothing provides computational efficiency (Liu et al., 2016), yet these findings emerge from distinct contexts without testing whether performance rankings shift as regimes change.

Third, empirical evidence concentrates on moderate volatility rather than extreme structural instability. Even studies acknowledging shocks examine recovery phases (Abdul Razak et al., 2017) or routine fluctuations (Khan & Khan, 2020). Zimbabwe's 2000 to 2023 trajectory, encompassing hyperinflation reaching 79.6 billion percent, currency abandonment, dollarization, and pandemic shock, represents volatility orders of magnitude beyond existing literature, providing natural experimental conditions unavailable through simulation or moderate volatility analysis.

Chapter 3: Research Methodology

3.1. Data and Economic Context

We analyse annual GDP growth data for Zimbabwe (2000-2023) from the IMF World Economic Outlook database, selected for its comprehensive coverage of the hyperinflation period that Zimbabwe's national statistical agency inadequately documents. The 24-year timeframe provides a natural experiment across three distinct post-dollarization regimes:

Table 1: Economic Regime Characteristics

Period	Regime	GDP Growth Mean%	Volatility (std Dev)	Key Features
2009-2013	Stable	23.74	12.39	Dollarization, price stability, recovery
2014-2018	Transition	16.45	28.63	Declining growth, deterioration
2019-2023	Unstable	1.21	23.34	Local Currency reintroduction, global pandemic

Annual rather than quarterly data reflects availability constraints during hyperinflation and filters short-term noise to focus on persistent structural changes relevant for medium-term policy planning.

3.2. Model Specifications

ARIMA (1,1,1): Augmented Dickey-Fuller testing on raw GDP growth yielded inconclusive stationarity (test statistic: -2.87, $p=0.052$). First differencing achieved stationarity (ADF: -5.34, $p<0.01$), confirming integration order $d=1$. ACF/PACF diagnostics suggested ARIMA(1,1,1), which outperformed AIC-optimal ARIMA(0,1,1) in out-of-sample forecasts (MAE: 3.49 vs 3.77, 7.4% improvement). We prioritized forecast accuracy over in-sample fit.

Holt's Exponential Smoothing: Selected over simple exponential smoothing due to visible trend in ACF patterns, and over Holt-Winters due to absence of seasonal patterns in annual data. The dual-parameter structure (level + trend) enables adaptation to Zimbabwe's evolving growth dynamics without unnecessary complexity.

VECM (Bivariate GDP-CPI): Consumer Price Index proved optimal among candidate variables based on: (1) matching integration order I(1), (2) strong theoretical justification (inflation-output nexus central to Zimbabwe's trajectory), and (3) sample size constraints. Johansen cointegration testing on full sample (2000-2023) confirmed one cointegrating relationship: trace statistic for $r=0$ (20.59) exceeded critical value (15.49), while trace for $r\leq1$ (0.82) did not exceed its threshold (3.84). The bivariate specification satisfies Brooks' (2019) minimum 10:1 observations-to-parameters ratio.

Critical Limitation: VECM was estimated once using complete 2000-2023 sample, with this fixed specification applied throughout evaluation. Ideally, cointegration would be re-tested at each expanding window, but 24 observations make this infeasible (reliable testing requires 10-15 observations minimum). Section 4.4 presents retrospective rolling analysis revealing full-sample evidence masked complete absence of cointegration during operational windows.

3.3. Forecast Evaluation Framework

Expanding Window Approach: We employ expanding windows that progressively increase estimation sample size, beginning with 2000-2008 (9 observations) for 2009 forecasts and culminating with 2000-2022 for 2023 forecasts. This mimics operational conditions where practitioners utilize all available historical information.

Multi-Horizon Structure: Each window generates forecasts at three horizons:

- **1-step ahead:** Immediate-year budget planning
- **3-step ahead:** Medium-term expenditure frameworks
- **4-step ahead:** Strategic infrastructure planning

The 4-step maximum reflects both sample constraints and practical limits—forecast reliability beyond 4-5 periods becomes questionable even in stable environments (Hyndman & Athanasopoulos, 2018).

Accuracy Metrics: We evaluate performance using four complementary metrics:

- **MAE/RMSE:** Absolute error magnitude in original units
- **MAPE:** Scale-independent comparison across periods
- **MSE:** Mathematical foundation, sensitive to outliers

Evaluation Period: Forecasts cover complete 2009-2023 out-of-sample period (15 years, 45 total forecasts across three horizons), distributed across all three economic regimes for regime-specific accuracy assessment.

3.4. Statistical Validation

Bootstrap Diebold-Mariano Tests: We test whether observed performance differences reflect genuine forecasting superiority rather than sampling variation. Bootstrap implementation addresses small-sample concerns (5 observations per regime), relaxing asymptotic assumptions. Pairwise comparisons test all model combinations separately for each regime and horizon.

Non-Parametric Tests: Wilcoxon signed-rank tests evaluate whether median forecast errors differ between models, while sign tests examine consistency (frequency of wins). These distribution-free approaches ensure conclusions don't depend on parametric assumptions potentially violated under extreme volatility.

Rolling Cointegration Analysis: To investigate whether cointegration underlying VECM specification remains stable during operational periods, we retrospectively test GDP-CPI cointegration using successive expanding windows matching forecast generation. Beginning with 2000-2008 (for 2009 forecasts) and expanding through 2000-2018, we apply Johansen's trace test at each window to quantify the proportion exhibiting significant cointegration. This tests the critical but typically unexamined assumption that full-sample relationships persist during individual forecast windows.

Chapter 4: Research Findings and Discussion

4.1. Overall Performance Patterns

No single model dominates across all conditions. Table 2 presents condensed MAE results revealing three distinct phases: comparable performance during 2009-2017, catastrophic simultaneous failures during 2018-2019, and subsequent divergence where model rankings separated dramatically.

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

Figure 1 visualizes temporal evolution, revealing critical inflection points where regime changes overwhelmed model capabilities. All models-maintained MAE below 4% through 2017, but the 2018 structural break triggered simultaneous failures (MAE ≈14%) regardless of theoretical sophistication. The 2019 RTGS dollar reintroduction produced the period's worst 1-step errors: VECM (32.16%), ARIMA (23.27%), and Holt's (25.77%).

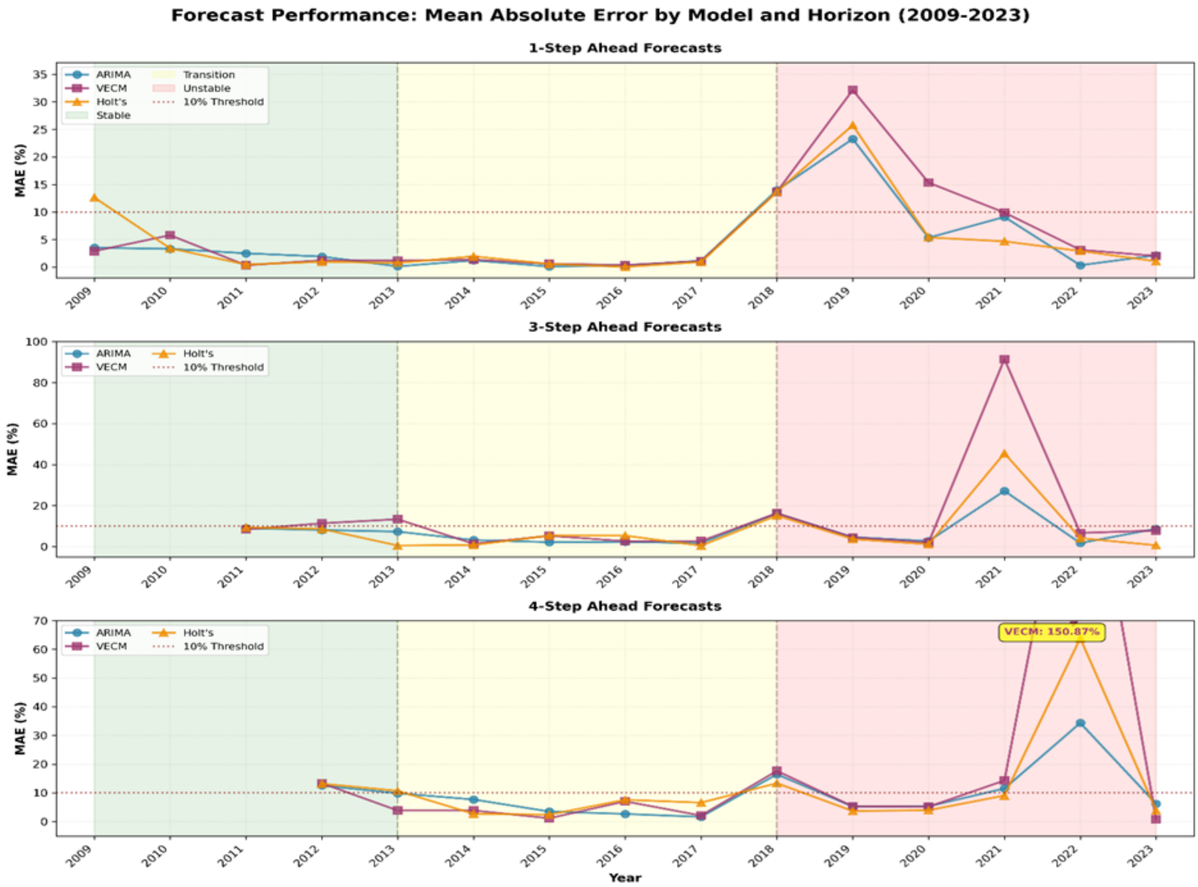

Figure 1: Forecast Performance by MAE

Post-crisis divergence reveals fundamental resilience differences: ARIMA recovered quickly (0.31% MAE by 2022), demonstrating ability to adapt without carrying mis specified assumptions. Holt's exhibited similar recovery (1.03% by 2023). VECM struggled persistently (15.33% in 2020, 9.88% in 2021), only normalizing by 2022-2023.

4.2. Regime Specific Performance

Stable Period (2009-2013): Holt's excelled at 1-step horizons (1.10% MAE), outperforming ARIMA (2.25%) and VECM (2.57%). Mechanical trend extrapolation proved optimal during smooth dollarization-era growth. At 3-4 step horizons, all models struggled (8-13% MAE), with VECM exhibiting erratic performance despite improving conditions its isolated 2013 excellence (3.89% vs 13% for alternatives) likely reflected overfitting rather than robust specification.

Transition Period (2014-2018): During 2014-2017, models achieved indistinguishable 1-step performance (ARIMA: 0.66%, VECM: 0.83%, Holt's: 0.85%). Gradual bond note deterioration preserved short-term accuracy despite accumulating distortions. The critical 2018 inflection exposed universal forecasting limits: all models failed simultaneously with 20-fold error deterioration, demonstrating the fundamental impossibility of forecasting through regime changes using historical patterns.

Unstable Period (2019-2023): VECM's catastrophic 32.16% initial 1-step error worst across all periods reflects complete GDP-CPI cointegration breakdown following RTGS dollar introduction. ARIMA (23.27%) and Holt's (25.77%) started elevated but declined steadily to 0.31% and 1.03% by 2023, revealing 3-4 year adaptation periods.

At **3-step horizons**, the 2021 critical failure exposed fundamental robustness differences:

- ARIMA: 27.21% MAE (elevated but bounded)
- **VECM: 91.42% MAE** (catastrophic—errors nearly equal actual GDP)

- Holt's: 45.55% MAE (severe but less extreme)

At **4-step horizons**, VECM's 2022 catastrophe represents the study's most extreme failure:

- **VECM: 150.87% MAE, 462% MAPE** (forecasts 4.6× actual GDP values)
- ARIMA: 34.37% MAE (bounded, minimally useful)
- Holt's: 63.84% MAE (intermediate)

VECM's error correction mechanism didn't merely fail but actively worsened forecasts, with cumulative bias pulling predictions progressively further toward non-existent equilibria.

4.3. Forecast Horizon Effects

Table 3 quantifies error growth patterns from 1-step to 4-step forecasts, revealing non-linear relationships between horizon extension and accuracy loss.

Table 3: 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

Stark mean-median divergence reveals occasional catastrophic horizon failures. Holt's demonstrates most balanced profile (lowest mean and median), suggesting consistent error accumulation. During stable periods, 4-step forecasts remain viable (8-12% MAE). During unstable periods, reliable forecasting contracts to 1-2 steps maximum ARIMA averages 14.59% at 4-step while VECM's 38.12% renders extended horizons operationally useless.

.

4.4. Cointegration Specification Failure

Full-sample testing using 2000-2023 data produced strong cointegration evidence: trace statistic (20.59) decisively exceeded critical value (15.49), seemingly validating VECM specification. Rolling-window tests tell a dramatically different story.

Table 4: 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

During supposedly stable dollarization years, GDP and CPI exhibited zero detectable cointegration when tested using expanding windows actually available for forecast generation. The unstable period's 50% detection occurred precisely when VECM produced catastrophic failures, suggesting spurious relationships from parallel market distortions rather than genuine equilibria.

Critical finding: Pooling heterogeneous periods generated sufficient power to detect "cointegration" that never existed within any operational forecast window a fundamental hazard in cointegration-based forecasting that standard practice fails to address.

4.5. Statistical validation

Bootstrap Diebold-Mariano Results: Stable and transition periods show statistical equivalence across models (all p>0.96), confirming rankings emerge specifically during volatility. The unstable period produces substantial test statistics (ARIMA vs VECM 4-step: -22.73) with confidence intervals directionally favouring ARIMA [-69.92, 2.77], providing directional evidence despite small-sample power limitations.

Non-Parametric Tests: During unstable periods, Holt's achieves perfect 100%-win rates against VECM at 1-step and 3-step horizons (Wilcoxon p=0.06, Sign test p=0.06). This cross-test replication strengthens confidence that simple exponential smoothing systematically outperforms cointegration-based approaches when equilibrium relationships destabilize.

Practical vs Statistical Significance: Following Ziliak & McCloskey (2009), we emphasize that VECM's catastrophic failures (91-462% MAPE) exceed any decision-making threshold regardless of formal significance levels. Triangulation across multiple testing methods consistently identifies Holt's superiority over VECM during crises despite small samples, demonstrating robust empirical regularities with clear operational implications.

4.6. Robustness Check: VAR vs VECM

Plain VAR (1) estimation without cointegration restrictions consistently outperforms VECM during instability (3-step: 15.84% vs 24.13% MAE, 34% error reduction), confirming error correction mechanisms compound specification errors. However, VAR still fails catastrophically relative to ARIMA (3-step: 15.84% vs 9.00%, 76% worse), producing errors 2.4× larger during 2021.
Interpretation: Multivariate frameworks fail hierarchically error correction adds 30-40% additional error, while base multivariate relationships destabilizing accounts for 50-75% degradation versus ARIMA. When GDP-CPI relationships sever during currency crises, both long-run equilibria and short-run dynamics become unreliable simultaneously.

Chapter 5: Discussion

5.1. Theoretical Implications

This research challenges a fundamental tenet of modern econometric forecasting: that theoretically sophisticated models incorporating economic structure necessarily deliver superior predictive performance. Our findings demonstrate that structural sophistication systematically amplifies rather than attenuates forecast errors when foundational assumptions fail.

VECM's catastrophic failures (150.87% MAE, 462.62% MAPE) versus ARIMA's bounded errors (20.9% MAE) during Zimbabwe's 2019-2023 currency crisis directly contradict implicit hierarchies favouring cointegration-based approaches. Where existing literature positions VECM as superior for correlated indicators (Khan & Khan, 2020; Abdul Razak et al., 2017), we find these rankings reverse systematically with economic conditions Holt's excels during stability (1.1% MAE), ARIMA dominates crises (10.2% MAE), VECM fails catastrophically during regime breaks.

5.2. Temporal Aggregation Bias in Assumption Verification

Our most critical methodological contribution addresses why existing literature failed to document VECM's instability-period failures. The rolling cointegration analysis exposes a dangerous verification practice: full-sample tests justified specifications inappropriate for every operational forecast window.

- Full sample (2000-2023): Trace statistic 20.59 (significant, suggesting cointegration exists)
- Operational windows (2009-2018): 0% cointegration detection, average trace 6.49 points below threshold
- Result: VECM deployed with violated assumptions throughout evaluation period

This "temporal aggregation bias" represents more than Zimbabwe-specific artifact it reflects systematic hazard in cointegration-based forecasting. Pooling heterogeneous regimes generates statistical power to detect relationships absent within any single forecasting period. Standard practice tests assumptions once using all available data, then applies specifications continuously, presuming temporal stability that regime shifts violate.

Implication for forecasting methodology: Assumption robustness should supersede theoretical elegance when structural stability is uncertain. ARIMA's superior crisis resilience reflects not superior economic theory but structural advantage making no long-run equilibrium assumptions, specification errors affect periods independently rather than compounding through correction mechanisms.

5.3. Practical Implications: Regime Conditional Model Selection

Our findings necessitate **regime-conditional model selection protocols** rather than universal forecasting frameworks:

Three-Tier Verification System:

1. *Rolling diagnostic monitoring:* Conduct cointegration tests at each forecast update using only historically available information, not full-sample data. Flag VECM specifications when trace statistics approach thresholds (within 3-5 points) our evidence shows proximity precedes catastrophic failures.

2. *Parallel forecasting systems:* Maintain multiple model classes (ARIMA, exponential smoothing, VECM) with ex-post-performance tracking to identify regime-dependent patterns. Implement automatic model switching when accuracy deteriorates.
3. *Automatic horizon contraction:* When 1-step forecast errors exceed institutional thresholds (e.g., MAE >10%), restrict published forecasts to 1-2 periods while flagging extended horizons as unreliable.

Decision Rules for Practitioners:

- *Stable conditions* (low volatility, policy consistency): Holt's for 1-step, ARIMA for 3-4 step, consider VECM only with continuous verification
- *Transition/uncertainty:* Default to ARIMA regardless of horizon—univariate robustness outweighs multivariate potential gains
- *Crisis/regime breaks:* Contract to 1-step forecasts only, use ARIMA or Holt's, abandon VECM entirely

The Forecasting Paradox: Predictive capacity disappears precisely when most needed. Our results demonstrate 3–4-year adaptation periods following regime changes forecasting utility returns only after acute phases pass. This cannot be solved through methodological refinement but must be acknowledged transparently in policy communications.

5.4. Limitations & Generalizability

Sample size constraints: Twenty-four annual observations impose statistical power limitations, particularly for period-specific analysis (5 observations per regime). While we address this through practical significance frameworks and multiple metric triangulation, some period-specific conclusions represent suggestive evidence requiring larger-sample confirmation.

Annual frequency limitations: Prevents analysis of within-year dynamics and business cycle patterns. However, if models fail at annual frequency during extreme instability, they likely struggle equally at quarterly frequency our findings establish upper bounds on model vulnerabilities.

Generalizability to moderate volatility: Zimbabwe's extreme trajectory (hyperinflation >79 billion percent, complete currency abandonment) raises legitimate questions for economies experiencing moderate instability. Yet this extreme case provides critical theoretical leverage: if theoretically sophisticated models fail catastrophically under maximum assumption stress, the same structural vulnerabilities likely operate at smaller magnitudes during moderate instability. The principles we document specification vulnerability compounding, regime-dependent rankings, temporal aggregation bias represent general phenomena whose magnitudes scale with instability severity.

VECM specification constraints: Our inability to re-test cointegration at each expanding window (using full-sample specification throughout) represents significant methodological limitation. Ideally, operational forecasting would verify assumptions continuously, but 24 observations make this infeasible (reliable testing requires 10-15 observations minimum). This means VECM's documented failures partly reflect specification based on misleading full-sample evidence rather than purely model inadequacy though rolling analysis confirms assumptions were violated throughout evaluation periods.

Chapter 6: Conclusions

This paper examined traditional time series models under extreme volatility using Zimbabwe's 2009-2023 GDP trajectory across three regimes: stable dollarization, transition, and crisis. We compared ARIMA, VECM, and Holt's Exponential Smoothing using expanding window evaluation across multiple horizons.

Three Principal Findings

1. Model rankings are regime-dependent, not invariant. Holt's excelled during stability (1.1% MAE at 1-step), ARIMA dominated transitions and crises (0.66%, 10.2% MAE), while VECM produced catastrophic failures: 32.16% (2019, 1-step), 91.42% (2021, 3-step), 150.87% (2022, 4-step with 462% MAPE forecasts 4.6× actual GDP).

2. Rolling cointegration exposed temporal aggregation bias. Evidence suggests that full-sample tests (trace: 20.59, significant) justified VECM specifications inappropriate for every operational window we detected **0% cointegration during 2009-2018**, with average trace 6.49 points below thresholds. Pooling heterogeneous regimes generated power to detect relationships absent in any single forecasting period.

3. Horizon degradation accelerates during instability. Extending 1-step to 4-step increased errors 4-fold during stability (2%→8%) but exploded 5-7-fold during crises (10%→50%), with VECM exhibiting 10-14-fold degradation versus ARIMA's 4-5-fold. Reliable forecasting contracted from 4-year to 1-2-year horizons during crises.

Theoretical Contribution:

We document that structural sophistication amplifies errors when assumptions fail**.** VECM's catastrophic performance versus ARIMA's bounded errors contradicts hierarchies favouring cointegration-based approaches. ARIMA's crisis resilience reflects structural advantage making no equilibrium assumptions, errors affect periods independently rather than compounding through correction mechanisms.

The Forecasting Paradox

Predictive capacity disappears when most needed. Following regime changes, accuracy requires 3-4 years of post-crisis data by which point acute phases have passed. This cannot be solved through statistical refinement but must be acknowledged transparently.

Core lesson: Improving forecast accuracy in volatile economies requires economic stability more than statistical innovation. Zimbabwe's deterioration from reliable 4-step forecasts during dollarization to useless predictions during crisis reflects underlying volatility no model can overcome. For policymakers, the path to reliable projections runs through institutional stability and policy consistency.

Our 24-observation annual dataset imposes power limitations (5 observations per regime), though we address this through practical significance frameworks. Zimbabwe's extreme trajectory raises generalizability questions yet provides theoretical leverage: if sophisticated models fail under maximum assumption stress, same vulnerabilities likely operate at smaller magnitudes during moderate instability.

Four extensions: (1) Quarterly data replication, (2) Cross-country validation (Ghana, Nigeria, Egypt), (3) Comparison with machine learning methods, (4) Real-time forecasting with operational constraints.

Final Takeaway

In structurally unstable environments characterizing many developing economies simplicity becomes strength. Models making fewer assumptions maintain utility when sophisticated alternatives collapse. Match model complexity to environmental stability, verify assumptions continuously, contract horizons upon deterioration, and acknowledge uncertainty transparently when structural shifts overwhelm historical patterns.

Reliable forecasting rests on institutional foundations that statistical methods cannot substitute. Zimbabwe's experience reveals econometric sophistication's limits when confronting genuine regime change a lesson with profound implications for developing economies navigating volatile transformations.

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.