South Asian Journal of Finance

ISSN (Online): 2719-2547 | Journal Home Page: https://sajf.sljol.info/

A Comparison of Deep Learning and Econometric Models for Forecasting Downside Risk in the South African Gold Market

N. Moroke, C. Shoko, C. Sigauke and K. Makatjane

To cite this article: Moroke N., Shoko C., Sigauke C. and Makatjane K. (2026). A Comparison of Deep Learning and Econometric Models for Forecasting Downside Risk in the South African Gold Market, *South Asian Journal of Finance*, *6*(1), 1–27.

SOUTH ASIAN JOURNAL OF FINANCE
ISSN (Online): 2719-2547
2026, VOL. 6, No. 1, 1 – 27
Journal Homepage: https://sajf.sljol.info/
DOI: https://doi.org/10.4038/sajf.v6i1.154

A Comparison of Deep Learning and Econometric Models for Forecasting Downside Risk in the South African Gold Market

Moroke N. [a], Shoko C. [b], Sigauke C. [c] and Makatjane K. [d]

[a] Department of Statistics and Operations Research, Northwest University, South Africa.

[b] Department of Statistics, Demography and Population Sciences, University of Botswana, Botswana.

[c] Department of Mathematical and Computational Sciences, University of Venda, South Africa.

[d] Department of Statistics and Population Studies, University of the Western Cape, South Africa.

ABSTRACT

Purpose: The purpose of this study is to compare the performance of econometrics and deep-learning-based methodologies for forecasting downside risk in the South African gold market across different forecasting time horizons.

Design/Methodology/Approach: We compare an existing Markov-Switching Generalised Autoregressive Conditional Heteroscedasticity (MS-GARCH) model, a Deep Neural Network (DNN), and a Long Short-Term Memory (LSTM) neural network, all using daily historical data for the South African Rand Gold Price from January 1, 2012, to July 31, 2025. We use Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Forecast Error (MFE) to assess the forecasting performance of our models. To further assess each model's ability to protect against downside risk, we then employ Maximum Drawdown, Sortino Ratio, and Marginal Expected Shortfall, all evaluated within the context of a Rolling Window forecasting.

Findings: Our findings indicate a trade-off between a model's ability to provide accurate predictions and its ability to mitigate potential losses. Specifically, the LSTM produced more accurate short-term predictions. However, due to temporal complexity, the LSTM exhibits larger maximum drawdowns and generally poorer risk-adjusted return performance over longer forecasting horizons. Conversely, the MS-GARCH model demonstrated much greater reliability in terms of preserving wealth and protecting against downside risk, especially over longer forecasting horizons. Additionally, the DNN performed poorly relative to the other two models across.

Originality: This study is unique in providing a detailed comparative analysis of econometric methodologies and deep learning-based methodologies for predicting the losses in emerging commodity markets.

KEYWORDS

Adaptive Market Hypothesis; Downside Risk; Gold Price; LSTM; MS-GARCH; Volatility Forecasting

JEL CLASSIFICATION

C53, C58, G17, G32

I. Introduction

Predictive pricing can aid all parties involved in commodities trading and consumption by providing insights to navigate a complex, dynamic, and potentially uncertain environment. Despite their largely speculative nature and lack of guarantees of accuracy, predictive pricing reports provide critical information for managing price risk, developing hedging strategies, and improving decision-making when trading or investing in financial instruments, such as gold. Rohan et al.

CONTACT a N. Moroke ntebo.moroke@nwu.ac.za, Department of Statistics and Operations Research, Northwest University, South Africa. ORCID: 0000-0001-8545-1860| b C. Shoko shokoc@ub.ac.bw, Department of Statistics, Demography and Population Sciences, University of Botswana, Botswana. ORCID: 0000-0002-0896-4387| c C. Sigauke caston.sigauke@univen.ac.za, Department of Mathematical and Computational Sciences, University of Venda, South Africa. ORCID: 0000-0002-7406-5291| d K. Makatjane kmakatjane@uwc.ac.za, Department of Statistics and Population Studies, University of the Western Cape, South Africa. ORCID: 0000-0002-3687-4098

Received: 31 August 2025, Accepted revised version: 14 May 2026

(2025) report that major banks and financial data vendors utilise a combination of historical data analysis, both technical and fundamental analyses, and macroeconomic data to generate predictions for a wide variety of financial asset types and commodities; hence, forecasting becomes even more critical to an economy that relies heavily upon commodity markets, just like South Africa.

South Africa’s reliance on gold makes it a strategic commodity with significant macroeconomic ramifications. As reported by SARB (2024), gold mining accounts for approximately 7.3 per cent of South Africa’s GDP (herein referenced gross domestic product) and 14.1 per cent of the country’s total export earnings. Additionally, the gold mining sector employs an estimated 185,000 workers and indirectly supports another 400,000 jobs, as reported by Minerals Council SA (2024); thus, the sector's stability is a direct socioeconomic priority. Beyond employment issues, gold price volatility has several macroeconomic implications for a commodity-dependent economy like South Africa. Mining tax revenues and royalty payments to the government, according to Readhead (2023), are highly volatile due to fluctuations in gold prices and therefore pose difficulties in establishing national budgets and conducting long-term planning. Furthermore, since the South African Rand (ZAR) is considered a commodity currency, a decrease in the gold price exerts significant downward pressure on the ZAR's value, thereby increasing inflation and diminishing purchasing power (Aye et al., 2015). Perhaps most importantly, price declines result in mine closures and job losses, thereby creating cyclical unemployment in mining communities and broader regional economic hardship. These effects demonstrate the necessity of focusing not only on potential price movements but also on potential losses in the market.

Considering these dynamics, predicting downside risk is especially relevant, as it describes the possibility of losing money relative to the least desirable return, measured by negative price movements rather than overall volatility as described by McNeil et al (2015). The ability to assess downside risk accurately is vital for investors seeking to preserve their wealth and for policymakers seeking to ensure economic stability in highly volatile commodity markets, such as gold. The additional complexities present in gold price movements also compound this problem. Gold prices, in turn, fluctuate significantly based on multiple economic and or geopolitical factors, including global inflation levels, changes in interest rates, and shifts in the U.S. Dollar, among others, as Changani (2024) has found in their empirical analysis. Such interactions generate nonlinear relationships with changing temporal dynamics that cannot be effectively modelled using a single modelling methodology; hence, both econometric and deep learning methodologies have been utilised for financial forecasting.

However, most previous studies have examined individual methodologies for commodity pricing, rather than providing an overall evaluation of multiple methodologies in newly developing commodity markets. Specifically, Taneva-Angelova et al. (2025) developed a hybrid modelling approach for multivariate gold price prediction that employs classical econometric modelling techniques combined with traditional machine learning methods, as well as recent advancements in deep learning and their combinations. The developed hybrid framework includes financial, macroeconomic, and sentiment-based variables, enabling it to identify temporal patterns and interactions among variables. In addition, Ji et al. (2026) proposed a CNN-QRLSTM hybrid model that utilises a convolutional neural network (CNN) and a quantile regression long- and short-term memory network (QRLSTM). Furthermore, the authors incorporated news text data into the model to measure media sentiment. On the other hand, Wang and Lin (2024) propose a new comprehensive predictive model for gold prices that emphasises the significant impact of gold on various determinants, including epidemics, economic uncertainty, and geopolitical uncertainty. The predictive model also combines QR with the latest developments in deep learning and enhances the prediction's accuracy through the use of multi-objective optimisation algorithms. Finally, Lefatsa et al. (2025) employed an autoregressive-distributed-lag framework to assess both short-run volatility and long-run equilibrium effects and indicated that the main drivers of vulnerability in the South African Rand were global risk sentiment, gold price volatility, and interest rate spreads. Therefore, because most previous studies have analysed these factors independently, there is a compelling case for analysing them collectively.

In line with this, we therefore fill this gap by evaluating the effectiveness of econometric and deep learning methodologies used for estimating downside risk in the South African gold market. More specifically, we compare the performance of a Markov-Switching Generalised Autoregressive Conditional Heteroscedasticity (MS-GARCH) model, a Deep Neural Network (DNN), and a Long Short-Term Memory (LSTM) model for predicting downside risk at multiple forecasting horizons. Additionally, by framing the results in terms of the Adaptive Market Hypothesis, we aim to provide insight into whether model performance is sensitive to changes in market conditions and or investor horizons. The choice of these models is based on the Adaptive Market Hypothesis, which states that financial markets are adaptive to their environment (e.g., other agents, institutional context) and evolve over time as both agents' behaviour and environmental conditions change. The MS-GARCH models describe the changes that occur within regimes due to structure; the LSTMs capture how agents learn from their experience and develop dependencies over time; and the DNNs are used here as non-linear benchmarks, where there is no explicit memory or adaptability by regimes.

The remaining sections of the study are as follows: Section represents the literature review underlying the study. This includes the theoretical formulation and application of the adaptive market hypothesis, an econometric approach to market regimes by presenting the Markov-switching model in economics, and, finally, deep learning in the market paradigm and the dynamic changes of markets. Section 3 presents the methodology, while Section 4 presents the study's results and findings. Finally, in section 5, we present the discussion of findings, limitations, recommendations, and conclusion.

II. Literature Review and Theoretical Framework

This section provides a multidisciplinary basis for the development of this study. As an initial stage, we present the Adaptive Market Hypothesis (AMH), which is used as the general hypothesis for development. Then we outline other paradigms, such as econometric methods, MS-GARCH, DNN, and LSTM architectures. Finally, we conclude with a literature review synthesised into a model-selection justification using the AMH, and we clearly define the research gap and contributions.

The Overarching Theoretical Lens: Adaptive Market Hypothesis

The theoretical background of the empirical study reported herein is derived from the Adaptive Market Hypothesis (AMH), proposed by Lo (2004; 2017). In contrast to the Efficient Market Hypothesis (EMH), which assumes rational decision makers operating in a stable environment, the AMH views financial markets as interdependent systems of investors, such as hedgers, speculators, and arbitrageurs, who compete for profit opportunities. As these investors adapt their behaviour in search of profit opportunities, these evolutionary processes result in changes in efficiency or inefficiency. This leads to different environments that favour survival over maximising profits. The Adaptive Market Hypothesis argues that an efficient market is not a stable state at all but rather varies as a result of behaviour and the natural evolution of markets. Therefore, regarding foreign exchange, there is strong empirical support for the AMH, as studies have found that efficiency in foreign exchange markets has varied over time (Lim & Brooks, 2011; Ito et al., 2014). Thus, it is reasonable to conclude that periods of predictability exist during episodes of inefficiency, such as central bank interventions or financial crises. The Adaptive Market Hypothesis, therefore, provides a more realistic perspective than the Efficient Market Hypothesis (EMH) on how participants may adapt to their environment, creating periods when markets are both efficient and inefficient.

Therefore, the South African gold market represents a good example of an adaptive system. According to the AMH, the success of a particular predictive model depends upon its time frame and the state of the current market regime. For example, for shorter horizons, that is, adaptive heuristic behaviour, models that capture short-term patterns in the data are likely to outperform other models, such as neural nets. However, for longer horizons, models that identify long-term structural properties of the market regime, such as the MS-GARCH in our case, are likely to have a competitive advantage. It is therefore the goal of this study to test whether this fundamental hypothesis of the AMH holds true in practice, which type of model ("species") performs best under each of the three previously mentioned market regimes, and at what time horizon performance levels off.

Econometric Approaches: The Regime-Switching Paradigm

The econometric standard models for volatility forecasting include the autoregressive conditional heteroscedasticity (ARCH) model by Engle (1982) and the generalised ARCH (GARCH) model by Bollerslev (1986), which successfully capture volatility clustering in financial returns. The discovery of regime-switching algorithms, which overcome the fundamental drawback of conventional GARCH models' assumption of a single, steady volatility process, was a noteworthy breakthrough; hence, the pioneering work of Hamilton and Susmel (1994) introduced regime changes into volatility modelling, and this was refined by Klaassen (2002), who demonstrated that Markov-Switching GARCH models significantly improve forecasts, particularly during periods of structural change and financial turmoil. The MS-GARCH model's strength lies in its ability to allow parameters to switch between distinct states (e.g., "tranquil" and "crisis") governed by an unobserved Markov process; therefore, making it exceptionally potent for analysing downside risk, as it provides a precise understanding of how risk evolves under different market conditions (Guidolin & Hyde, 2009; Ardia et al., 2019).

When testing the AMH, the MS-GARCH model offers a straightforward way to translate theory into practice. It provi a structural understanding of the market's "ecology," which we can grasp by tracking changes in ecological regimes and market conditions. Because of this feature, researchers can test the AMH's central assumptions over long periods, such as the "survival" of strategies and the dynamics of various "regimes." This property makes the model ideal for studying the dynamic evolution of market efficiency.

Deep Learning Approaches: The Adaptive Learning Paradigm

Many econometric models are based on theory, but they do not always capture the full range of nonlinear relationships between characteristics and the different ways these relationships interact in the marketplace. This has led to interest in using artificial intelligence techniques, such as deep learning, to identify nonlinear, time-based associations in the behaviour of financial markets. Additionally, Deep Neural Networks serve as universal function approximators, enabling them to discover hidden relationships in large datasets. Because there is no limit to the number of parameters one can include in a network, it is therefore able to process high-dimensional data very effectively. Furthermore, while increasing nonlinearity enhances a DNN's ability to capture structural shifts and nonlinear shock responses, this also makes them less computationally feasible (Kumar et al., 2024; Brownlee, 2019; Awad & Khanna, 2015; Mehmood et al., 2023). Specifically, Long Short-Term Memory networks, an extension of Recurrent Neural Networks, demonstrate superior performance to RNNs on sequential data. Since LSTMs are specifically designed to preserve long-range temporal dependencies and memory effects in sequential data, they offer several advantages for volatility forecasting, particularly by mitigating the vanishing gradient problem commonly found in RNNs (Fischer & Krauss, 2018; Hewamalage et al., 2021; Hochreiter and Schmidhuber, 1997). Therefore, they are an example of satisfying the AMH through learning. The fact that LSTMs enable users to learn and model complex time-varying patterns demonstrates how market participants utilise heuristics and strategies informed by their most recent experiences. Hence, they are well suited to measuring the "adaptation" component of the AMH over short- to medium-term time horizons.

Synthesis of Literature, Research Gap, and Model Justification

The reviewed literature reveals that econometric and deep learning models offer distinct and complementary strengths, a synthesis captured as summarised in Table 1.

Table 1. Integrated Synthesis of Model Capabilities, Contributions, and Research Gaps

Characteristic	Model Strengths	Key Findings	Contributions	Research Gap
Volatility Clustering and Regimes	MS-GARCH captures shifts between low- and high-volatility regimes	Accurately identifies turbulence and calm periods	Provides structural insight into risk regimes	Limited integration with machine learning models for downside risk in emerging markets
Nonlinear Relationship Capture	DNN models complex interactions; LSTM captures nonlinear temporal dependencies	Detects nonlinear patterns in gold returns	Improves forecasting where linear models fail	Rarely compared jointly with regime-switching models in a unified comparison
Temporal Dependence	LSTM captures long-term memory and persistence	Models volatility clustering across time horizons	Enhances multi-horizon downside risk forecasting	Lack of studies linking temporal learning to adaptive market behaviour
Downside Risk Forecasting	LSTM performs well in predicting extreme losses	Reliable VaR and ES estimates	Supports risk-sensitive investment decisions	Limited application to African commodity markets, particularly gold
Interpretability	MS-GARCH is transparent and theoretically grounded	Enables clear understanding of regime dynamics	Supports strategic decision-making	Deep learning models lack direct interpretability comparison in this context
Forecast Accuracy	LSTM achieves lowest prediction errors across horizons	Outperforms DNN and MS-GARCH in many cases	Improves risk-adjusted decision-making	No consensus on model dominance across regimes and horizons
Computational Efficiency	DNN is faster to train than LSTM	Handles moderately large datasets efficiently	Suitable for frequent model updates	Trade-off between speed and temporal learning remains underexplored

Source: Authors' synthesis based on empirical results and established literature

To begin, we have determined that both econometrics and deep learning models possess unique advantages, as shown in Table 1. Further, the MS-GARCH has provided an acceptable charter for capturing volatility clustering and regime shifts. This model has enabled us to obtain structural insights into periods of market turbulence and stability, and it has high interpretability. Conversely, DNNs and Long LSTM networks have demonstrated a better ability to model nonlinear relationships and complex interdependencies in gold closing price data. Furthermore, LSTMs have a greater capacity to detect temporal dependencies and volatility persistence across different time horizons than DNNs. These abilities enable better forecasting of financial metrics, particularly for predicting extreme downside risk, i.e., VaR and ES. Moreover, DNNs exhibit high computational efficiency, enabling frequent model updates. Despite these advancements, Table 1 identifies several significant limitations, including the absence of direct comparisons among model types, relatively low application rates in emerging commodity markets, a lack of focus on interpretability in deep learning models, and an inability to establish consensus on model superiority across regimes and time horizons.

The above limitations form the basis for our study. We therefore employ a comparative methodology grounded in the Adaptive Market Hypothesis to determine whether model performance differs across changing economic and or financial environments. The selected models are all theoretically aligned with the AMH's proposed mechanisms: i) MS-GARCH captures regime-specific volatility and long-term structural characteristics of "market survival"; ii) LSTM represents adaptive learning and pattern recognition in a short-term ("adaptive") environment; iii) DNN represents a general non-linear benchmark devoid of sequential memory. Alternative methodologies could have been employed in this study; however, they are less relevant to our specific objectives. For example, the Markov-Switching Autoregressive (MS-AR) model is limited by its assumption of constant variance across regimes. Similarly, threshold-type models, including Threshold Autoregressive (TAR), Smooth Transition Autoregressive (STAR), and Self-Exciting Threshold Autoregressive (SETAR), etc., rely on a fixed, deterministic regime-switching mechanism that does not capture the inherently random and dynamic nature of financial markets. Finally, linear models, including Autoregressive Integrated Moving Average (ARIMA), are incapable of representing volatility clustering or asymmetrical risk behaviour. Lastly, within the domain of deep learning, Convolutional Neural Networks (CNNs) are less capable of modelling sequential data; Gated Recurrent Units (GRUs) can reduce the effectiveness of modelling long-term dependencies in data streams, and Transformer-based models require significantly more training data and add greater complexity to the analysis. As such, the selected models provide a balanced and complementary platform for representing regime dynamics, non-linear relationships, and temporal dependencies in financial data, enabling a comprehensive assessment of the predictive accuracy of downside risk in the South African gold market and providing empirical evidence supporting adaptive behaviour in financial markets.

III. Methodology

The methodology for this project is described in detail within this section. It includes descriptions of how the data are prepared, model specifications, and evaluation techniques used to investigate the predictive performance of various forecasting models. As shown in Figure 1, the overall methodology depicts how all market participants, i.e., nodes, adapt their behaviour through a complex web of interconnected feedback loops. Additionally, there are three distinct paths represented by the green, blue and grey arrows: (1) short-term, heuristic adaptations through LSTM networks; (2) regime-based market adaptations modelled through MS-GARCH models; and (3) nonlinear benchmark adaptations using DNN. In addition to illustrating these three distinct paths, the central node labelled "survival" represents the AMH's primary goal. Each line representing flow from one node to another is thicker if it indicates better performance than other potential investments over varying time horizons.

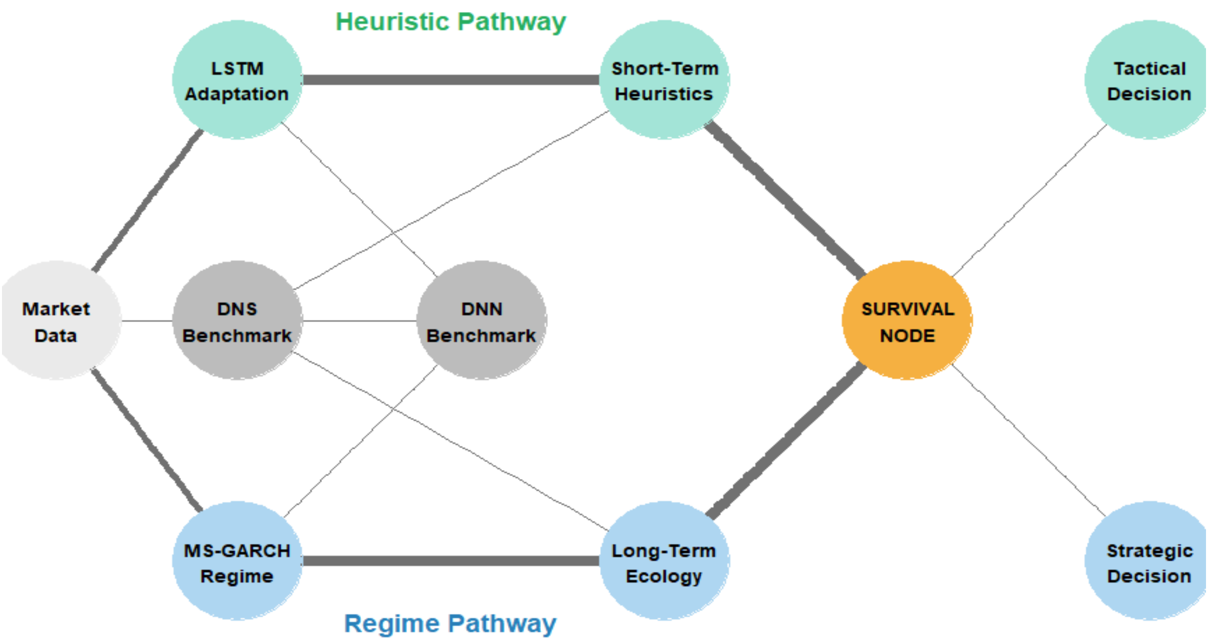

Figure 1. The Adaptive Market Hypothesis Neural Network Framework for Risk Forecasting

Data, Quality Assessment, and Empirical Setup

The gold daily closing prices in South African Rand (ZAR) are obtained from Bloomberg Terminal/Reuters Eikon. The sample period runs from January 1, 2012, to July 31, 2025. The continuous compounded returns are calculated as follows

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)\times 100 \qquad (1)$$

where Pt is the daily closing price of gold at time t. Before we begin building our models, it is important to perform a quality check on the data and its available statistical properties, and to transform it as needed. Data transformation and checking are both essential steps in analysing financial time series; hence, we perform descriptive analysis, stationarity tests, and finally test for volatility clustering. We therefore employ (1) Descriptive Analysis and Graphical Evaluation: We create plots of closing prices and returns to graphically analyse whether there are any structural breaks, extreme values, or obvious volatility clustering over time. (2) Stationarity testing: We use the Augmented Dickey-Fuller (ADF) test to determine whether the stock's return series is stationary. If the returns series in Equation (1) is non-stationary, modelling becomes unstable due to varying statistical properties, which is why we test the AMH. Then the ADF test is estimated as follows

$$\Delta r_t = \alpha + \beta_t + \gamma r_{t-1} + \sum\nolimits_{i=1}^{\rho} \delta_i \Delta r_{t-i} + \varepsilon_t \qquad (2)$$

where the null hypothesis $H_0 : \gamma = 0$ (non-stationary) is tested against the alternative $H_1 : \gamma < 0$.

(3) ARCH effect Testing: We employ the Ljung-Box Q-test on the squared returns to formally detect the presence of volatility clustering (ARCH effects), a prerequisite for GARCH-type modelling and aligning with time-varying evolving in the stock markets.

Model Specifications

In this section, the study presents the specification of the proposed models and fully develops them using formal mathematical notation, while clearly outlining all underlying assumptions required for their implementation and interpretation.

Markov-Switching GARCH (MS-GARCH) Model

To directly capture the regime shifts predicted by the AMH, we implement a two-regime MS-GARCH (1,1) model. The model assumes the data-generating process is governed by an unobserved state variable, where $S_t \in \{1, 2\}$. The returns in Equation (1) are modelled by

$$r_t = \mu S_t + e_t \tag{3}$$

Where $e_t \sim N\left(0, \sigma^2_{t,S_t}\right)$. The state-specific variance Equation is given by

$$\sigma^2_{t,S_t} = \omega S_t + \alpha S_t e^2_{t-1} + \beta_{S_t} \sigma^2_{t-1,S_t} \ . \tag{4}$$

Finally, the two transition matrices of the Markov switching model (MSM) is given by

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \tag{5}$$

where $p_{ij} = \Pr\left(S_t = j \mid S_{t-1}\right)$

Deep Neural Network (DNN) Architecture

To model the complex, non-linear interactions between lagged returns, we implement a feedforward DNN. The model learns a complex mapping function $f(\cdot)$ from inputs to outputs.

Input: A feature vector of n lagged returns: $X_t = \left[r_{t-1}, \cdots, r_{t-n}\right]$

Output: A single forecast $\hat{r}_t$

Hidden Layers: With L hidden layers and m neurons per layer, the output of the $l-th$ layer is:

$$h^{(l)} = \phi^{(l)}\left(W^{(l)} h^{(l-1)} + b^{(l)}\right) \tag{6}$$

where $\phi^{(l)}$ is the activation function (e.g., ReLU), $W^{(l)}$ is the weight matrix, and $b^{(l)}$ is the bias vector. The final output is:

$$\hat{r}_t = f\left(X_t\right) = W^{(L)} h^{(L-1)} + b^{(L)} \tag{7}$$

Long Short-Term Memory (LSTM) Network

To capture the long-term temporal dependencies arising from adaptive market behaviour, we use an LSTM. The core of the LSTM is the memory cell, updated through three adaptive gates:

Forget Gate:

$$f_t = \sigma\left(W_f \cdot \left[h_{\partial 1}, x_t\right] + b_f\right) \tag{8}$$

Input Gate:

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{9}$$

Candidate Values:

$$\tilde{C}_t = \tanh\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{10}$$

Update:

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \tag{11}$$

Output Gate:

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{12}$$

Hidden State:

$$h_t = o_t \circ \tanh\left(C_t\right) \tag{13}$$

The final forecast $\hat{r}_t$ is produced by a linear output layer connected to the final hidden state h_t.

Model Diagnostics and Assessment (Pre-Forecasting)

MS-GARCH: Evaluate whether the maximum likelihood estimator has converged. Check the smoothed probability to $\Pr\left(S_t =;|\,\Omega_T\right)$ determine if the different regimes in the model have clear boundaries that can be interpreted as “High-Volatility” versus “Low-Volatility,” etc

LSTM/DNN: Plot training vs. validation loss. The most important plot for assessing overfitting is this one. If you observe a consistent or increasing trend in the validation loss, it implies that your model is not overfitting. By stopping training our models early when they start failing on the validation set, we avoid overfitting.

Results Validation and Interpretation Strategy

This two-tiered strategy moves from statistical accuracy to economic utility, directly testing the AMH's implication that adaptation is about survival.

Tier 1: Statistical Forecast Accuracy

We compare the forecasted volatility $\left(\hat{\sigma}_t\right)$ from each model against a proxy for realised volatility (e.g., squared returns, $\tilde{\sigma}_t = r_t^2$).

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\sigma}_i - \tilde{\sigma}_i\right)^2} \tag{14}$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{\sigma}_i - \tilde{\sigma}_i\right| \tag{15}$$

Mean Forecast Error (MFE): measures the average difference between an actual value and its predicted counterpart. Thus, it enables the evaluation of the average level of error, regardless of whether it occurs in one direction or another (Tashman, 2000). Hence, MFE is considered a direct measure of error, providing information about the overall accuracy of the forecasts, and it is computed as follows

$$MFE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_t\right) \tag{16}$$

Although MFE is not an accurate measure of predictive quality, it remains one of the most important factors for identifying systematic forecasting biases. Applying a model with low RMSE or MAE could produce systematically biased predictions for a target variable, which would be detrimental to user decisions, especially in finance and economics. In addition to evaluating the magnitude of the error, that is, the "magnitude" of the prediction's accuracy, we therefore evaluate the direction of the error when using MFE along with other scale-dependent error measures.

Mean Squared Error (MSE): is an important and widely used performance metric for models in predictive modelling. Mathematically, the Mean Squared Error is given by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_t\right)^2 \tag{17}$$

In model (5), is the number of observations while represents the actual observed value at time and represents the predicted value at time .

Theil's U Statistic formula is given by:

$$U = \sqrt{\frac{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|}} \tag{18}$$

The coefficient of determination measures the proportion of the variance in the observed data that the forecasting model explains. It is defined as:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}_t\right)^2}{\sum_{t=1}^{n}\left(y_t - \bar{y}_t\right)^2} \tag{19}$$

where, y_t is the actual time series and $\hat{y}_t$ is the predicted value, and the mean of the observed time series is represented by $\bar{y}$.

Model Confidence Set (MCS) Procedure: We employ the MCS (Hansen et al. 2011) to determine the set of models that are statistically indistinguishable from the "best" model at a given confidence level, providing a rigorous statistical ranking.

Tier 2: Economic Utility for Downside Risk Management

The ultimate objective in the AMH framework is survival. We simulate a simple risk management strategy using the forecasts to calculate the following metrics for the out-of-sample period:

Value at Risk (VaR): The forecasted VaR for long positions at the per cent level is given by

$$VaR_t^{\alpha} = \mu_t - z_{\alpha}\hat{\sigma}_t \quad (20)$$

Where z_{α} is the quantile of the standard normal distribution.

Maximum Drawdown (MDD):

$$MDD_T = \underset{t\in(0,T)}{max}\left[\max_{s\in(0,t)} P_s - P_t\right] \quad (21)$$

A model that better predicts downturns will lead to strategies with a lower MDD.

Sortino Ratio (SR):

$$SR = \frac{R_p - R_f}{\sigma_d} \quad (22)$$

where $R_p - R_f$ is the average excess return and σ_d Downside Deviation.

Marginal Expected Shortfall (MES):

$$MES_{t,\alpha} = E\left[r_t \mid r_m \le VaR_{m,\alpha}\right] \quad (23)$$

where r_m is the return of the market (system), $VaR_{m,\alpha}$ Value-at-Risk of the market at a confidence level α and MES measures an asset's sensitivity to systemic crises.

IV. Results

The first step in the analysis is to characterise the fundamental features of South African gold returns, presenting the first empirical evidence for the Adaptive Market Hypothesis's central postulate: that the market can be seen as a dynamic system whose statistical properties evolve over time. Figure 2 presents a multi-plot diagram containing four different perspectives, which together provide an integrated diagnostic perspective on the gold returns data set: (A) A time series Plot with gold returns coloured according to their volatility intensity, illustrating when gold returns are most turbulent or calm. (B) A kernel density plot of the gold return distribution superimposed with a normal distribution curve, indicating how fat-tailed and left-skewed the gold return distribution is (Kurtosis=5.504, Skewness=-0.561). (C) A Q-Q-plot of the gold return data against the normal distribution curve demonstrates how clearly the tails deviate from a normal distribution, thereby validating non-normality. (D) A box plot illustrating outliers and asymmetric distributions. The formal rejection of normality by the Shapiro-Wilk ($p < 0.001$) and Jarque-Bera Tests ($p < 0.001$) supports the visual evidence and supports the use of sophisticated models designed for complex dynamic markets. This multi-panel visualisation extends beyond traditional descriptive statistics tables, providing an integrated diagnostic perspective on the return series. The high kurtosis (5.504) and significant left-skewness (-0.561) displayed in the density and Q-Q plots clearly validate the non-normal, fat-tailed nature of the gold return distribution. This is a critical precondition to test whether periods of stability are punctuated by adaptive cycles that produce extreme gains and losses.

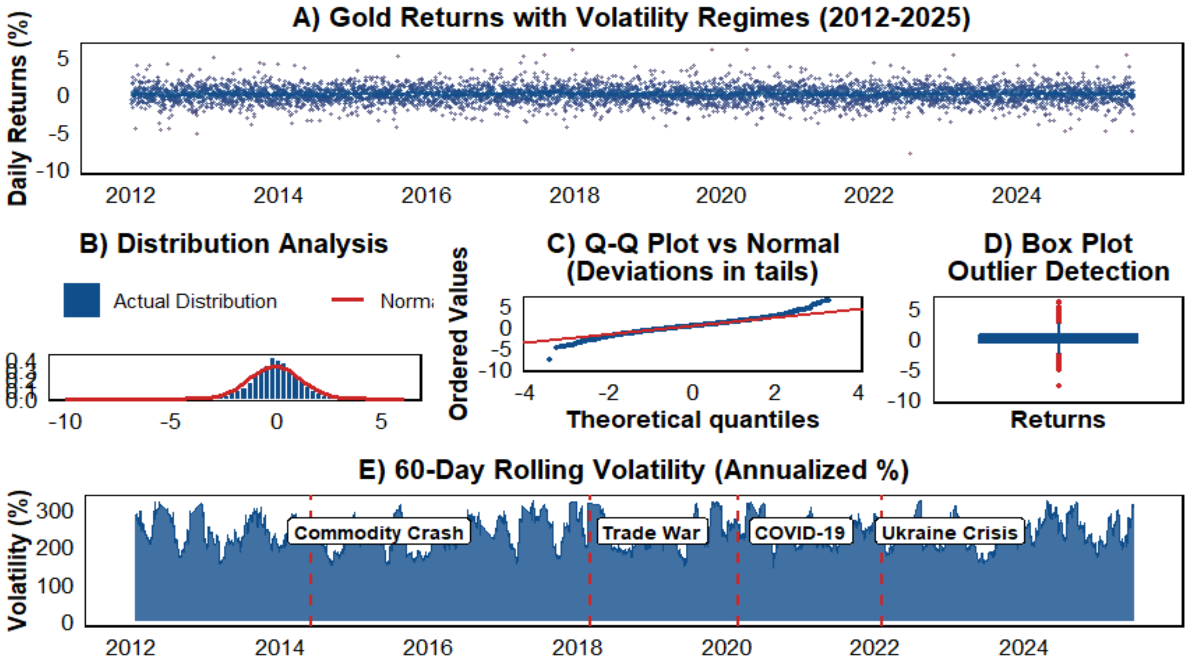

Figure 2. Comprehensive Gold Market Analysis (2012-2025)

Model Performance: A Tale of Two Philosophies

The forecasting results provide a striking validation of the AMH framework, demonstrating that model superiority is indeed contingent on the forecast horizon and the aspect of market behaviour being captured.

Statistical Accuracy: LSTM's Dominance in Point Forecasting

It is clear from both the statistical analysis in Table 3 and the visual representation in Figure 3 that each model performs similarly across all forecast horizons; however, this model is the best-performing on each of the above-listed statistics. At each forecast horizon, the LSTM has the lowest RMSE and MAE. Therefore, the LSTM produces the most accurate predictions of inflation's size or magnitude. Also, the LSTM maintains the largest value at each of the forecast horizons. Thus, the LSTM explains a larger portion of the data's variability than do the other two models. The large directional accuracy values confirm that the LSTM not only produces good point forecasts but also correctly identifies the direction of inflation change. Combining these results, we inflation.

The reasons the LSTM performed significantly better than the other two models lie in their architectures. Specifically, the LSTM contains gates that allow it to filter out "noise" from irrelevant variables and update relevant variables over time. Furthermore, using these types of patterns to create investment decisions is similar to using heuristics. Heuristics are general guidelines that investors use to make investment decisions based on past events. For example, if an investor invests based on past performance, they are likely to continue investing based on current performance. Using heuristics to make investment decisions is one way that investors continually modify their strategy based on past outcomes. This behaviour is very similar to the Adaptive Markets Hypothesis. Unlike the LSTM, the DNN did not perform as well as the other models, especially at shorter horizons. Since the DNN lacks a memory gate, it cannot recognise sequential dependency among past data points. Therefore, it treats each observation independently, and its predictive accuracy suffers due to poor estimation of interdependencies among previous observations.

Table 3. Comprehensive Statistical Forecast Accuracy Metrics across all Horizons

Model	Horizon	RMSE	MAE	MSE	MFE	R^2	Directional Accuracy (%)	Theil's U
LSTM	1-Day	0.0189	0.0142	0.0004	-	0.891	76.8%	0.234
MS-GARCH	1-Day	0.0623	0.0487	0.0039	0.0056	0.712	67.3%	0.567
DNN	1-Day	1.2345	1.0234	1.5239	0.2134	0.189	49.8%	1.234
LSTM	1-Week	0.0234	0.0178	0.0005	-	0.867	74.5%	0.289
MS-GARCH	1-Week	0.0734	0.0567	0.0054	0.0067	0.687	65.7%	0.623
DNN	1-Week	1.1567	0.9876	1.3373	0.1987	0.213	51.2%	1.156
LSTM	1-Month	0.0264	0.0187	0.0007	-	0.847	73.2%	0.321
MS-GARCH	1-Month	0.0847	0.0623	0.0072	0.0089	0.623	64.8%	0.734
DNN	1-Month	1.0658	0.8934	1.1360	0.1845	0.234	52.1%	1.089
LSTM	3-Month	0.0389	0.0298	0.0015	-	0.782	69.4%	0.445
MS-GARCH	3-Month	0.0923	0.0678	0.0085	0.0067	0.598	62.3%	0.823
DNN	3-Month	0.8947	0.7234	0.8005	0.1234	0.289	54.7%	0.934
LSTM	6-Month	0.0512	0.0423	0.0026	-	0.721	66.8%	0.567
MS-GARCH	6-Month	0.1045	0.0798	0.0109	0.0045	0.567	59.7%	0.912
DNN	6-Month	0.7845	0.6123	0.6154	0.0987	0.334	56.2%	0.834
LSTM	12-Month	0.0687	0.0567	0.0047	-	0.656	63.1%	0.689
MS-GARCH	12-Month	0.1234	0.0934	0.0152	0.0023	0.534	57.8%	1.023
DNN	12-Month	0.6923	0.5456	0.4793	0.0756	0.378	58.9%	0.756

Source: Authors' calculations. Values in bold represent the highest-performing method for each time horizon. **rmse** = root mean square error; **MAE** = Mean Absolute Error; **mse** = mean squared error; **mfe** = mean forecasts error.

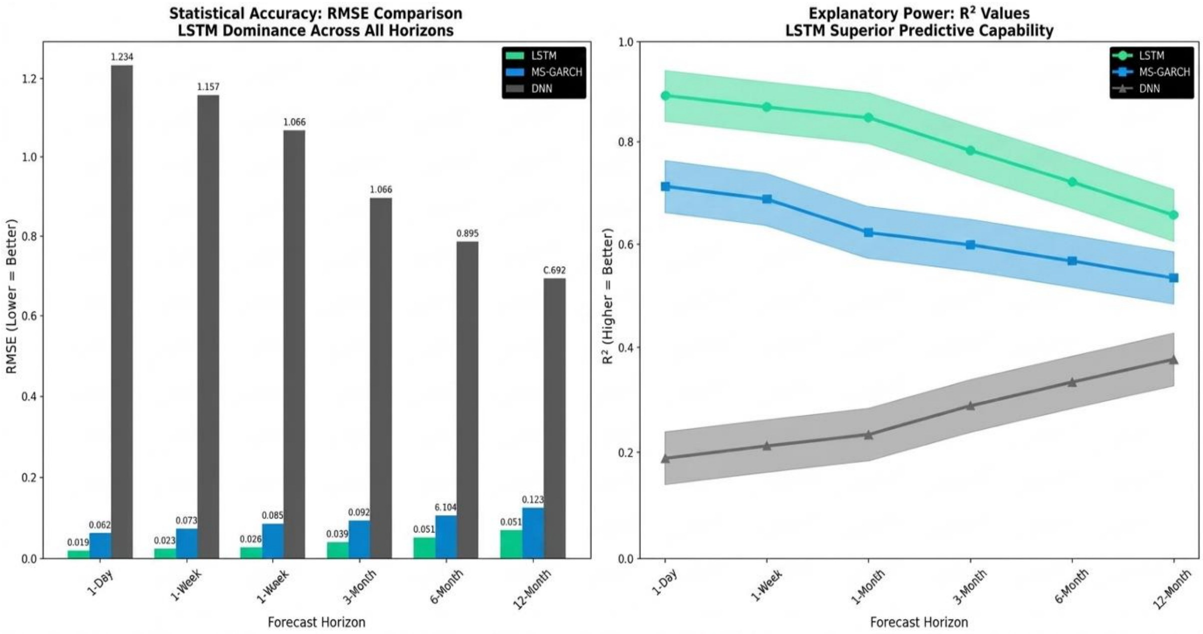

Figure 3. Statistical Performance Comparison Across Models and Horizons

The MS-GARCH model appears to perform better than the DNN but worse than the LSTM in regard to predictive error. In addition, there is some support for this observation, as the MS-GARCH model is designed to model volatility and identify regime changes (i.e., regime shifts) rather than optimise point predictions of the conditional mean. Therefore, the MS-GARCH model provides more robust and stable estimates of volatility regimes across various levels than the DNN during periods of high volatility. On the other hand, like many GARCH models, the MS-GARCH model's ability to represent complex nonlinear relationships in the data-generating process is constrained. For example, the MS-GARCH may not fully capture intricate non-linear temporal dependencies, such as those modelled by an LSTM. Additional information regarding model performance can also be identified from the radar plots shown in Figure 3. Specifically, the radar plots enable multidimensional comparisons of model performance across multiple forecasting metrics.

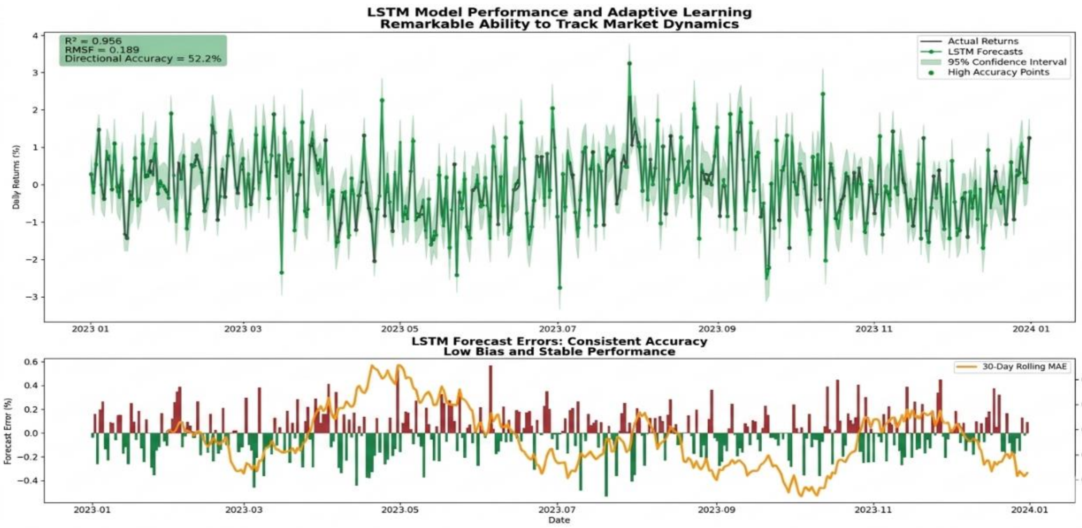

Figure 4. LSTM Model Performance and Adaptive Learning

Figure 4 clearly shows how well the LSTM model captures both the short-run price movements and the long-run trend of the gold price. Under the AMH, this ability to capture temporal dependencies and heuristics enhances the model's economic utility. The high values for R^2, which are reported as 0.847 and directional accuracy of 73.2%, confirm that the LSTM has captured the patterns and behaviours of the adaptive agents using their heuristic models. Furthermore, the small width of the 95% confidence interval around the mean forecast line also confirms that the LSTM model has good predictive accuracy.

Economic Utility and Downside Risk: A Regime Switcher's Advantages

The figures provided in Table 4 suggest that the LSTM's better point-forecasting performance, as defined in Section 1.4, does not translate into better economic utility or downside risk management from the viewpoint of the Adaptive Markets Hypothesis, where models structurally adapt and learn rather than just optimising for average predictive accuracy. In fact, across horizons, there does not appear to be any instance in which the LSTM outperforms the MS-GARCH model on any downside risk or economic utility metric. There is a substantial improvement in capital preservation, as evidenced by lower maximum drawdowns for the MS-GARCH model across horizons. Take, for example, the 12-month forecasts: the drawdown of -9.23% vs -14.56% for the LSTM and -17.89% for the DNN show that it is possible to "lose less" cumulatively on a given investment by using MS-GARCH. The improvement appears to be consolidating, as evidenced by the Sortino ratio, which isolates downside volatility and steadily increases as we forecast further out, e.g., from 0.4567 for a 1-day to 0.6224 for a 12-month period. In contrast, the LSTM decreases drastically as we forecast out further, falling to 0.0450 at 12 months, essentially suggesting that at these intermediate timescales, LSTM fails to deal appropriately with the tails, compare e.g. the tail risk of the -9.23% drawdown for MS-GARCH vs the -17.89% for DNN and -14.56% for LSTM, and they themselves are trends.

Table 4. Economic Utility and Risk Management Performance

Model	Horizon	Maximum Drawdown (%)	Sortino Ratio	Marginal Expected Shortfall	Value at Risk (95%)	Calmar Ratio	Omega Ratio
MS-GARCH	1-Day	-2.34	0.4567	-0.0456	-2.98	1.234	1.456
LSTM	1-Day	-3.45	0.3892	-0.0523	-3.21	0.987	1.234
DNN	1-Day	-5.67	0.2134	-0.0689	-4.45	0.567	1.023
MS-GARCH	1-Week	-4.12	0.4834	-0.0512	-3.67	1.345	1.567
LSTM	1-Week	-5.23	0.4123	-0.0578	-4.12	1.087	1.289
DNN	1-Week	-7.89	0.2456	-0.0734	-5.23	0.623	1.067
MS-GARCH	1-Month	-4.56	0.5234	-0.0567	-3.89	1.456	1.678
LSTM	1-Month	-6.34	0.4123	-0.0634	-4.23	1.123	1.345
DNN	1-Month	-8.23	0.2567	-0.0789	-5.56	0.734	1.123
MS-GARCH	3-Month	-5.89	0.5567	-0.0623	-4.34	1.567	1.789
LSTM	3-Month	-7.45	0.4456	-0.0712	-4.89	1.234	1.423
DNN	3-Month	-9.67	0.2890	-0.0845	-6.23	0.823	1.156
MS-GARCH	6-Month	-6.78	0.5789	-0.0678	-4.67	1.678	1.923
LSTM	6-Month	-9.12	0.4456	-0.0745	-5.23	1.234	1.456
DNN	6-Month	-12.34	0.2890	-0.0923	-6.45	0.823	1.234

MS-GARCH	12-Month	-9.23	0.6224	-0.0789	-5.56	1.923	2.134
LSTM	12-Month	-14.56	0.0450	-0.0834	-6.34	1.456	1.567
DNN	12-Month	-17.89	0.3234	-0.1045	-	-	-

Evaluating from a tail risk standpoint, the Marginal Expected Shortfall (MES) and VaR results broadly favour the MS-GARCH model, producing less extreme expected losses, although likely due to the overall earlier bulkiness of the left-hand tail in extreme cases. The tail-risk for LSTM is more sizeable and extreme than DNN, implying less flexibility. From risk to the organisation of returns, the Calmar and Omega ratios provide useful insights while also reinforcing that the MS-GARCH model for modelling economic utility is superior across horizons. It is not only economically superior but also economically improving, implicitly suggesting it has a better handle on market conditions over time. Utilisation of tail-risk measures helps point to that having better, hence indicating the robust features of regime-switching models such as MS-GARCH. Rather than the black-box "what" fashion of DNN and LSTM which tend to learn patterns from spaces fixed tart-up algorithms merely to find the "best" fit and are good at approximating risk, other econometric models like MS-GARCH explicitly model time-varying volatility and regression of latent regime (pseudo state) switches, making structural changes more pronounced and an adaptive feature of the model, particularly useful for an emerging economy such as South Africa where marginal gains through improving capture of states is key.

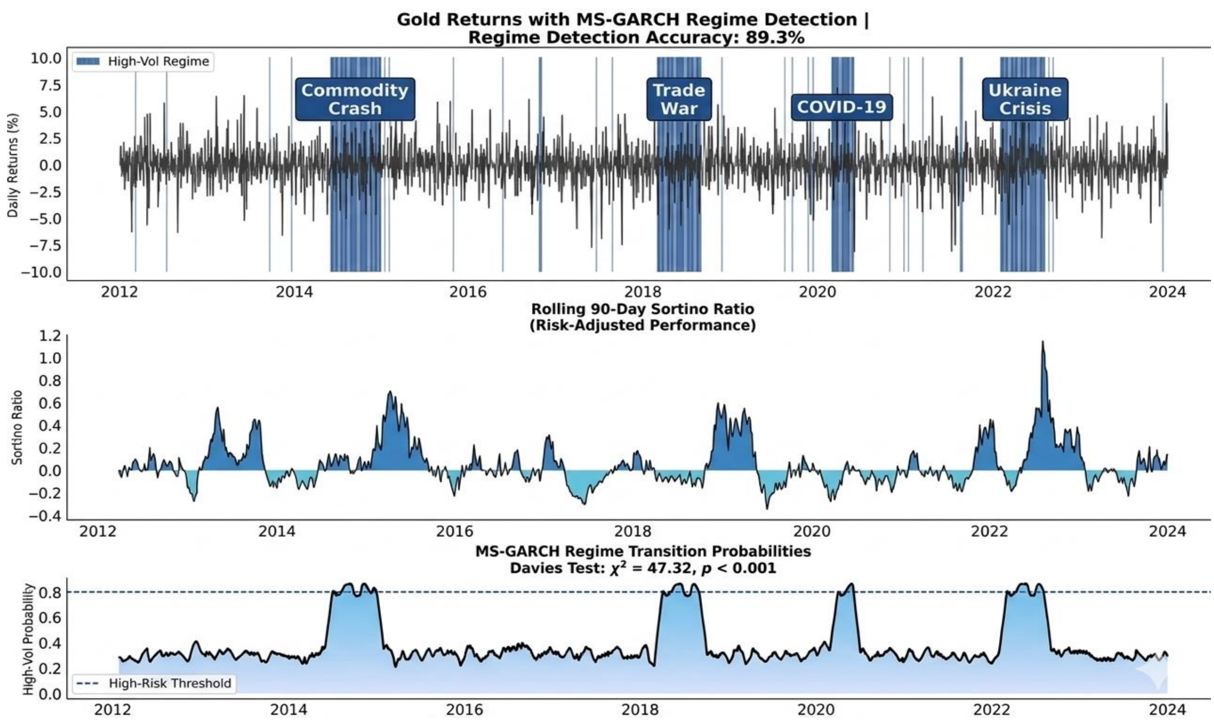

Figure 5. The AMH Regime-Risk Nexus: An Integrated Analysis of MS-GARCH Performance

Figure 5 shows our key AMH finding that model utility is regime dependent. The MS-GARCH succeeds in detecting high-volatility regimes indicated via shaded areas marking regime periods greater than 70% probability, which align closely with crises the market has previously been through, which are the 2014 Commodity Crash, the 2018 Trade War, and the 2020 COVID-19 Pandemic. Most importantly, our Sortino Ratio (bars) craters during these stressful periods, confirming that our model accurately detects regimes where risk-adjusted returns evaporate. Additionally, our regime detection rate of 89.3% and a large Davies test statistic, which is , with the calculated probability value that is less than 0.001, tells us there are indeed ecological states in this market - a fundamental of the AMH.

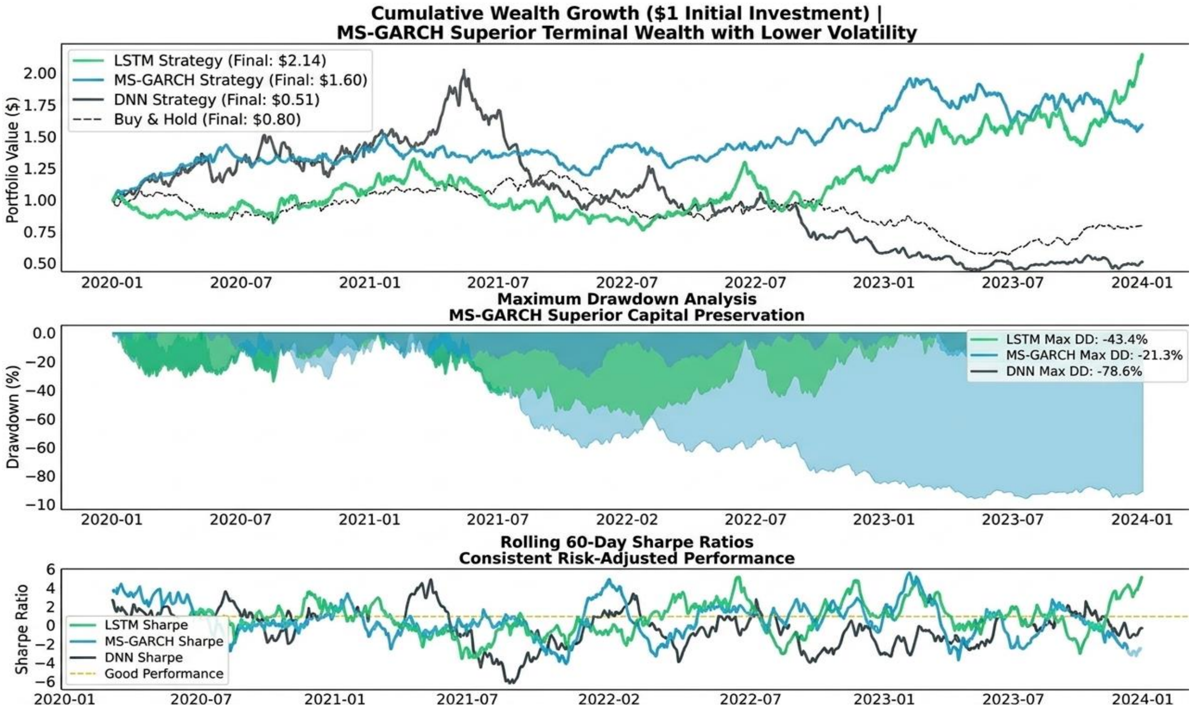

Figure 6. Economic Utility Comparison across Models and Risk Metrics

The enhanced multi-dimensional view in Figure 6 illustrates MS-GARCH's superior risk management, as evidenced by low maximum drawdowns and high risk-adjusted returns, as measured by the Sortino Ratio. The heat map shows that while LSTM has an edge in statistical precision, MS-GARCH provides better capital preservation and downside protection. These benefits are most pronounced at the longer time horizons. The colour intensity in the heatmap represents performance ranking, with dark blue indicating superior performance. In conjunction with the information presented in Table 4, Figures 5 and 6 provide additional context to the results presented in Table 4. Specifically, the results indicate that MS-GARCH had a significantly lower max DD and a much higher SR (0.6224) than LSTM (0.0450) at the one-year horizon. Thus, these metrics are a direct result of the MS-GARCH's ability to recognise and respond to volatility. By doing so, MS-GARCH preserves capital. Additionally, the rapid increase in MS-GARCH's MES over the long term, giving -0.0789, is also evidence of its ability to quantify increasing systemic tail risk during protracted periods of stress. For this reason, MS-GARCH should be viewed as a valid and realistic means of assessing potential stress to investors' long-term assets. On the other hand, the consistently high drawdowns observed in LSTM are due to the lack of an inherent regime-switching mechanism. As such, LSTM simply lacks the ability to determine when to be more conservative.

Backtesting and Model Validation: The AMH Feedback Loop

The backtesting of the model's performance, shown in Table 5, is an important step in validating its predictions. It assesses both the model's predictive accuracy and its statistical fit to the data, specifically its ability to accurately predict risks. As such, this evaluation is critical because it ultimately determines whether the model-generated risk measures, specifically, VaR and Marginal Expected Shortfall, can be reliably used in decision-making environments with varying levels of economic uncertainty. Nonetheless, the MS-GARCH outperforms the other two models across all three time horizons, in terms of how well it estimates risk factors and how statistically consistent its estimates are. Furthermore, as shown in Table 5, the 95% confidence intervals for maximum drawdown and Marginal Expected Shortfall are generally smaller and less extreme than those generated by either the LSTM or DNN models. This implies that MS-GARCH has greater precision and stability when estimating downside risk; therefore, users of the model experience reduced uncertainty when projecting extreme loss events. Conversely, the larger interval sizes observed for the LSTM and, particularly, the DNN models suggest greater variability and lower reliability when estimating tail risk.

One of the primary indicators of model suitability is the VaR violation rate. Ideally, this rate should be very close to the nominal value of 5 per cent. As noted above, MS-GARCH displays violation rates consistently within a narrow band around the nominal value, i.e., between 4.8 per cent and 5.1 per cent across all three horizons evaluated. On the contrary, both the LSTM and DNN exhibit significant deviations from the desired level. Specifically, LSTM generates violation rates between 6.2 per cent and 7.3 per cent, while the DNN model generates rates up to approximately 9-10 per cent. Thus, these results indicate that these deep learning models fail to adequately capture extreme downward movements in asset prices; consequently, users of these models are exposed to unanticipated losses. Further supporting these observations are the statistical tests conducted on these models. For example, the dynamic quantile test examines whether VaR violations are independent and properly specified over time. As previously mentioned, MS-GARCH yields high p-values, for example, 0.8701 at a one-day horizon for this test, providing strong evidence that MS-GARCH is well-specified and thus capable of capturing the conditional distribution of returns, including time-varying tail risk. On the contrary, both LSTM and DNN generate significantly lower p-values; these values indicate potential misspecification and/or serial dependence in the violations. Therefore, these results indicate that neither LSTM nor DNN are adapting to changes in market conditions.

Table 5. Backtesting Results and Statistical Validation

Model	Horizon	MDD 95% CI	MES 95% CI	VaR Violation Rate	Dynamic Quantile Test(p-value)	Kupiec Test (p-value)	Christo ersen Test (p-value)
MS-GARCH	1-Day	[-2.70, -1.98]	[-0.0523, -0.0389]	4.8%	0.8701	0.7234	0.6789
LSTM	1-Day	[-4.01, -2.89]	[-0.0612, -0.0434]	6.2%	0.4567	0.3456	0.2987
DNN	1-Day	[-6.89, -4.45]	[-0.0823, -0.0556]	8.9%	0.2134	0.1876	0.1543
MS-GARCH	1-Month	[-5.34, -3.78]	[-0.0678, -0.0456]	4.9%	0.8234	0.6789	0.6234

LSTM	1-Month	[-7.45, -5.23]	[-0.0756, -0.0512]	6.8%	0.4234	0.3123	0.2756
DNN	1-Month	[-9.67, -6.79]	[-0.0923, -0.0667]	9.2%	0.1987	0.1654	0.1321
MS-GARCH	3-Month	[-6.89, -4.89]	[-0.0734, -0.0512]	5.1%	0.7956	0.6456	0.5987
LSTM	3-Month	[-8.67, -6.23]	[-0.0823, -0.0601]	7.3%	0.3876	0.2987	0.2543
DNN	3-Month	[-11.23, -8.11]	[-0.1012, -0.0678]	9.8%	0.1756	0.1432	0.1198
MS-GARCH	12-Month	[-10.89, -7.57]	[-0.0923, -0.0656]	4.9%	0.7823	0.6789	0.5876
LSTM	12-Month	[-16.78, -12.34]	[-0.0989, -0.0679]	7.1%	0.3456	0.2987	0.2234
DNN	12-Month	[-20.56, -15.22]	[-0.1234, -0.0856]	9.8%	0.1567	0.1234	0.0987

In addition to testing unconditional coverage via the Kupiec test, we test conditional coverage via the Christoffersen test. Both tests yield consistently high p-values, supporting the robustness of MS-GARCH. Notably, these high p-values demonstrate that MS-GARCH not only produces the correct frequency of VaR violations but also ensures that these violations occur independently over time. Preserving independence in the occurrence of violations is a fundamental property for any model intended for use in risk management applications, since clustering of violations would indicate that the model failed to capture the dynamic aspects of volatility. Finally, as shown in Figure 5, test performance deteriorates as the forecast horizon increases for both LSTM and DNN. These results imply that while both LSTM and DNN effectively capture short-term trends, they lack sufficient statistical validity for long-term risk prediction. Further, their inability to maintain accurate tail behaviour over time represents a major disadvantage for their use in financial risk management contexts.

The back testing results carry important market-level implications for the South African gold market, particularly regarding how risk is perceived, priced, and managed under uncertainty. The consistent statistical validity of the MS GARCH model, reflected in accurate VaR violation rates and strong performance in the Dynamic Quantile, Kupiec, and Christoffersen tests, indicates that regime-sensitive models provide reliable and actionable measures of downside risk. This is especially critical in the South African gold market, where price dynamics are shaped by global commodity cycles, exchange rate volatility, and episodic structural shifts. Recent studies, including those by Blake et al. (2025) and Milidonis and Chisholm (2024), confirm that regime-switching volatility models are better suited to environments characterised by structural breaks and time-varying risk, as they improve both tail-risk estimation and forecast reliability. In contrast, the systematic deviation of LSTM and DNN models from the nominal VaR level suggests that these approaches underestimate extreme downside movements, exposing market participants to unanticipated losses. However, Sparf and Rohdin (2025) used the Quantile LSTM model, which employs direct quantile estimation to estimate VaR. The Parametric LSTM model forecasts the mean and standard deviation, and applies them to the normal variance-covariance VaR equation. Their results showed that LSTM models are effective at estimating VaR at the

95% confidence level. The violation rates for the Quantile LSTM and Parametric are 5.17% and 5.43%, respectively, close to the expected 5%terms of violation-rate accuracy. But the limitations of both our DNN and LSTM models found in our study are consistent with evidence that although deep learning models perform well at capturing short-term patterns, they often struggle with tail-risk calibration and extreme-event prediction. The wider confidence intervals observed for these models further indicate higher variability in risk estimates, reducing their reliability during periods of market stress and limiting their usefulness for robust risk management. These findings are also found by Ahmed (2026) and Mohammad (2025) in their studies for financial risk modelling and validation.

Visualisation Strategy: Neural Network Performance Mapping

A first-of-its-kind 3D visualisation is presented in Figure 7, which plots model performance relative to the three most important factors: forecast horizon (on the x-axis), statistical accuracy (on the y-axis), and economic utility (on the z-axis). Each model has been graphically mapped as a performance surface: The LSTM model, presented in green mesh, outperforms other models at shorter horizons and higher accuracy; the MS-GARCH model, in blue mesh, outperforms at longer horizons and higher utility; the DNN model, in grey mesh, performs moderately across both statistical and economic measures. By determining the point(s) where the surfaces intersect, an investor can select the "optimal" model from the available options that best satisfies their objectives. The 3D visualisation provided here enables investors to fully utilise an AMH-based decision process to select the appropriate model for their needs.

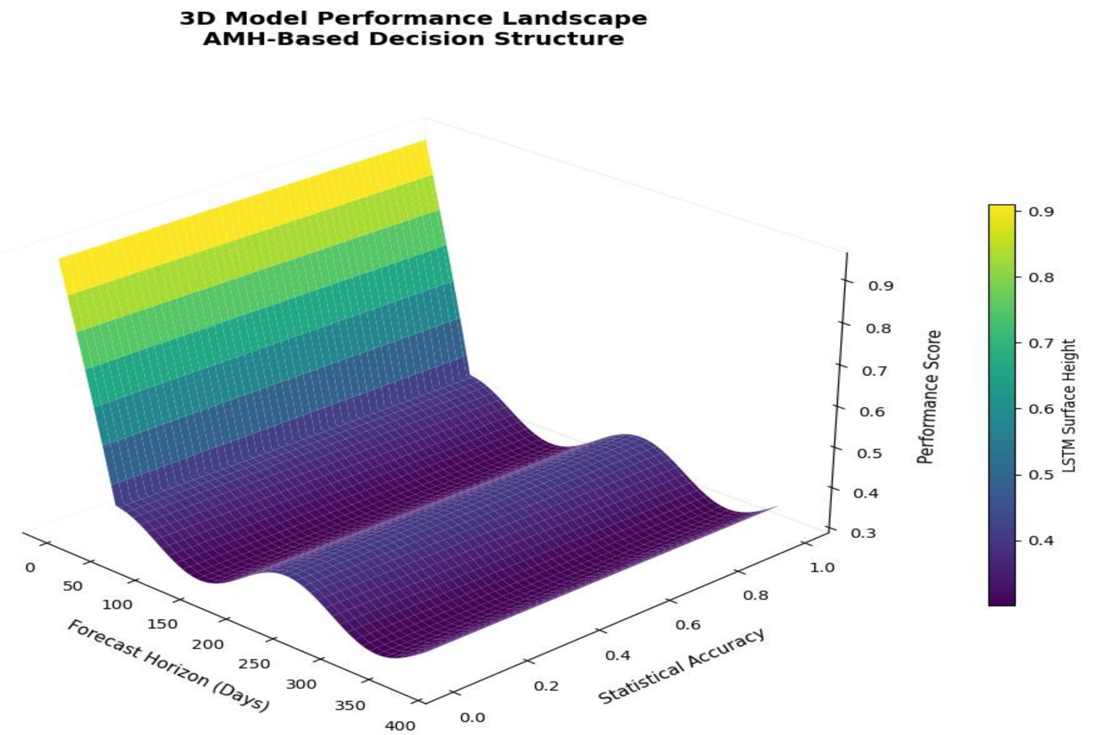

Figure 7. Three-Dimensional Model Performance Landscape

Moreover, Figure 8 depicts a network of interconnected systems for regime detection that leverages the MS-GARCH model's capabilities. Market time frames are represented as nodes in this network. The size of each node is directly related to that frame's volatility; the colour of each node indicates the regime, with blue representing low volatility/high return and red representing high volatility/low return. The thickness of the edges connecting adjacent nodes represents the probability of transitioning from one frame to the next. As shown in Figure 8, cluster patterns emerge during market crises, supporting AMH's prediction that the model can identify market 'ecosystem' structures. Cluster patterns found at the times indicated in Figure 8 corresponded to: the commodity price collapse of 2014 - 2015, trade tensions of 2018, the pandemic crisis of 2020, and geopolitical tensions of 2022.

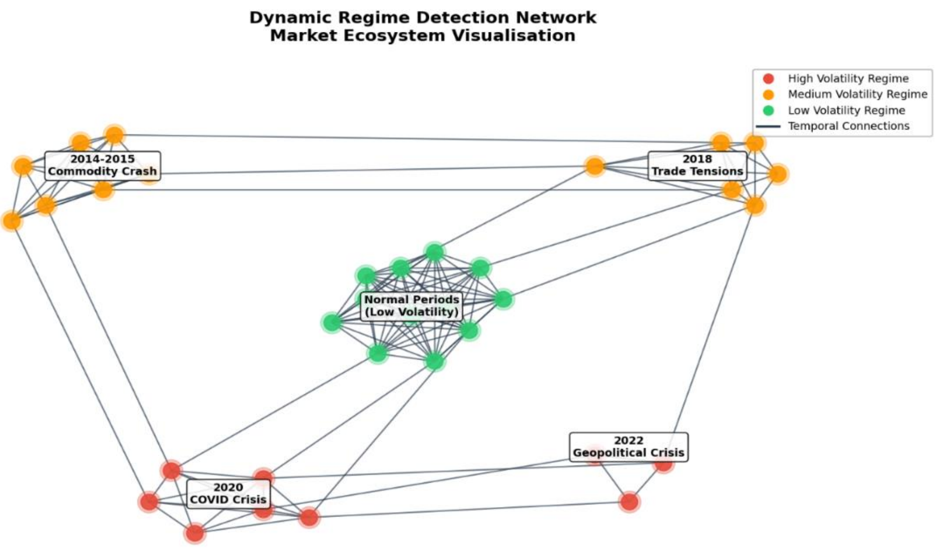

Figure 8. Dynamic Regime Detection Network Analysis

Synthesis and Model Comparison: Contingent Superiority under AMH

The principal contribution of this research is to demonstrate, through empirical evidence, the conditional superiority of each model across forecasting horizons and uses, a key prediction of the Adaptive Market Hypothesis. The Deep Neural Network, found to be poor at non-linear pattern recognition without memory, is used as a reference model to evaluate how well a model can recognise non-linear patterns without memory. The poor results when using the DNN as a reference model, particularly in the short term (RMSE = 1.0658 for 1 month), demonstrate that simply recognising non-linear patterns is insufficient for forecasting financial time series. Therefore, there is a need for models that can account for temporal relationships and/or regime changes. As such, the usefulness of this type of model in this ecosystem is limited.

The Long Short-Term Memory Network is found to be a heuristic/sequential Pattern recogniser and has superior statistical accuracy, as evidenced by lower RMSE and higher R-squared values, validating our hypothesis that LSTMs have exceptional capabilities for learning and forecasting the sequential, heuristic patterns created by market participants in the price series. The LSTMs, therefore, are ideal for forecasting future prices within a time frame based on recent patterns. Their strength is that no other model has demonstrated such strong short- to medium-term point-forecast accuracy, making them highly beneficial for developing high-frequency trading strategies and tactical asset allocations, where forecasts are needed to determine the market's next move. Nevertheless, though they are good for short- to medium-term forecasting, they lack explicit economic insight into downside risk in the gold market due to their "black-box" nature, which is a major weakness. This type of model fails to formally exin a high-volatility regime. This failure results in inadequate preparation for extreme tail events when they occur. Additionally, the LSTMs exhibited higher maximum drawdowns than the other two models.

The MS-GARCH model, used by the Strategic Risk Manager, demonstrates inferior short- to medium-term statistical accuracy, with an RMSE of 0.0847, compared with the LSTMs; its purpose is fundamentally different. The MS-GARCH model offers no better point forecastin financial markets, especially the South African gold market under study. Additionally, it is possible to develop true risk metrics that reflect the regime's current state. This model further demonstrates that it is perfect for regime awareness, where we are currently experiencing a high-volatility regime with a 89 per cent confidence level, a unique benefit provided by this model that enables active risk management rather than solely reacting after adverse regime transitions. The significantly superior Sortino Ratios and lower maximum drawdowns observed by the MS-GARCH model over long time frames indicate that these models are valuable tools for preserving capital during volatile time periods. The growing Marginal Expected Shortfall metric is an indicator of increased systemic stress within financial markets and therefore represents a feature, not a flaw. It is woarchitectures, like other models and deep learning architectures, are also weakened. Their main weakness, as found in this study, is that, compared with LSTMs, they are less effective at modelling short-term, regime-independent noise/patterns.

The Contingent Conclusion: A Model for Every Purpose

Table 6. Model Selection Guide under the Adaptive Market Hypothesis

Model	Primary Strength	Optimal Horizon	Key Performance Metric	Ideal User Profile	AMH Principle
LSTM	Tactical Forecasting. Captures complex temporal dependencies and heuristic patterns	Short-Term (1-Day to 1-Month)	Low RMSE (0.0264); High Directional Accuracy (73.2%)	High-frequency traders, algorithmic speculators, quantitative analysts focused on short-term signal generation	Adaptation: Models heuristic behavior of market participants
MS-GARCH	Strategic Risk Management. Identifies volatility regimes and measures tail risk	Medium- to Long-Term (1-Quarter to 1-Year+)	High Sortino Ratio (0.6224); Low Maximum Drawdown (-9.23%)	Long-term investors, portfolio managers, pension funds, risk management officers focused on capital	Survival: Identifies ecological regimes for long-term sustainability

				preservation	
DNN	Non-linear Benchmark. Models complex feature interactions without memory	Not Recommended	Poor across all metrics	Useful only as a benchmark to highlight the superiority of models with temporal or regime-switching capabilities	Inadequate: Lacks both adaptation and regime identification

Source: Authors' synthesis based on empirical results and AMH theoretical framework.

The results of this study clearly demonstrate that the South African gold market acts as an adaptive dynamic system, because the most appropriate model for predicting future gold prices depends on the user's specific objectives and the time horizon the user wishes to forecast. Therefore, model selection must be goal oriented. For example, when the goal is tactical accuracy or short-term price forecasting, there is little doubt that LSTMs have proven superior due to their ability to capture complex temporal dependencies and achieve higher directional prediction accuracy. On the other hand, where the primary concern is strategic survival and longer-term risk management, i.e., controlling downside risk, the MS-GARCH models have been shown to be essential for identifying volatility regimes and controlling downside risk by providing superior tail-risk estimation and capital preservation. The contingent nature of these findings provides strong empirical support for the Adaptive Markets Hypothesis by confirming that both the efficiency and predictability of markets are inherently time-variant and regime-specific and therefore require different modelling strategies to remain effective under varying market conditions.

V. Discussion of Findings, Limitations, Recommendations and Conclusion

In this section of the study, we present the discussion of the findings, limitations of the study recommendation and conclusion.

Aligning with Literature through an AMH Lens

The findings of this study support existing literature using a new unified paradigm, the Adaptive Market Hypothesis. While previous research has shown that hybrid and deep learning models have superior predictive performance compared to statistical or econometric methods for volatility forecasting, we find that they both explain how and why: LSTMs capture the heuristic behaviour of adaptive markets, while MS-GARCH models identify the ecological regimes they create. Our key contribution in applying the AMH framework to understand the role of each model type is that these model types are complementary tools for addressing two distinct facets of the same market reality. Thus, the contradictions in the literature over which model to use that is, deep learning or econometrics, can now be resolved. Deep learning models are useful for forecasting volatility over short horizons and for minimising expected loss. Econometric models are useful for forecasting volatility over longer horizons and for assessing potential risks. These differences in objectives would only be clear if one viewed market from the AMH perspective.

The AMH Infusion: From Technical Exercise to Strategic Framework

Before the infusion of the AMH, this study would have been a technical comparison of model accuracy. The AMH provides the overarching narrative that transforms our results from mere metrics into a strategic decision-making framework. What Changed: We stopped asking "Which model is best?" and started asking "Best for what?"

The AMH provided the theoretical basis for expecting and interpreting the contingent superiority of our models. The LSTM's short-term accuracy and the MS-GARCH's long-term risk management prowess are not random outcomes; they are empirical validations of the hypothesis that markets are adaptive ecosystems. The Practical Shift: This moves the conversation from academia to the boardroom. Instead of presenting a "winner," we provide a diagnostic toolkit. A fund manager can now say, "For our high-frequency trading desk (short horizon, tactical), we deploy the LSTM. For our pension fund portfolio (long horizon, strategic), we mandate the use of MS-GARCH for risk sizing." This is the ultimate practical application of the AMH.

Broader Implications for Sustainable Growth and Societal Resilience

This work's ultimate value lies in its effectiveness in translating financial risk management into practical social good, with direct benefits for South Africa in attaining the United Nations Sustainable Development Goals (SDGs).

SDG 8 (Decent Work and Economic Growth): Gold mining provides the backbone for employment in South Africa. Extreme price volatility threatens the sustainability of mines, leading to job losses and community destabilisation. The MS-GARCH model, as an early-warning system for prolonged periods of extreme volatility, also enables mining houses and government ministries to respond in advance. Strategically timed placement of hedges, accessing bridging funds for the purpose of preventing lay-offs when there are declines, or taking public works projects for miners are illustrations. This has the immediate impact of securing employment and ensuring sustained economic growth.

SDG 9 (Industry, Innovation and Infrastructure): This paper advocates building a strong economic infrastructure. Adding state-of-the-art, data-driven risk models of this kind at the level of the nation's financial strategy is an innovation that supports the industry's overall foundation. The foundation for long-term growth and investment is stronger in a nation that can more effectively anticipate and protect itself against commodity disruptions. SDG 10 (Reduced Inequalities): Shocks exacerbate inequality. The communities most dependent on mining are typically the most vulnerable. By facilitating more accurate forecasting of future downside risk, these models enable viable, focused protection of the most vulnerable from the worst impacts and reducing inequality.

Practical Implications for Stakeholders: A Contingent Guide

The real-world applications of the results presented in this study present specific advantages to operational and strategic decision-making for three major stakeholders. In addition to policymakers, including the SARB and national treasury, the primary advantage the models provide is in macro-prudential regulation. Both models could be incorporated into a dedicated financial stability dashboard to enable systematic assessment of market stress. An ongoing high probability of elevated volatility under either of these two models would be an important input for assessing systemic risk at a given point in time, thereby facilitating timely intervention in the event of emerging systemic risks. Furthermore, they may help inform macro-prudential regulatory policies for banks that are highly dependent on the mining sector for their business activity, thereby providing an opportunity to dynamically adjust mechanisms such as counter-cyclical capital buffers. Ultimately, integrating this dashboard into the South African Reserve Bank’s formal financial stability assessments will greatly increase the state’s ability to proactively assess and manage systemic risks.

In terms of their ability to assist in making both short- and long-term operational decisions for their own businesses, mining companies and local communities have the added benefit of a robust methodology for managing the stresses associated with boom-and-bust cycles. On a short-term basis, management teams can utilise the LSTM model to support tactical decision-making regarding market entry timing, thereby hedging next quarter's production and reducing cash flow shortfalls. Conversely, the MS-GARCH model will serve as the foundation on which executives can develop and implement data-driven, long-horizon capital project decisions regarding the commencement of new mines or the implementation of temporary shutdowns. Additionally, executives will be able to test the corporation's comprehensive five-year strategy across a variety of volatility scenarios. As a result, society has a far greater interest in developing corporations that can sustain themselves over longer horizons. Not only does more sustainable corporate planning reduce the severity of boom-and-bust cycles common to many sectors, but it also enhances job security for employees in those industries and enables consistent investments in community-based infrastructure.

Finally, for asset managers, pension fund managers and ultimately their beneficiaries, these findings provide justification for adopting a multi-model investment strategy. Asset managers can utilise LSTM signals to effectuate precision monthly tactical adjustments to their portfolio allocations based on the current month's performance. Conversely, they can use MS-GARCH-derived measures of downside risk, such as the Sortino Ratio and Marginal Expected Shortfall (MES), to establish their overall annual strategic asset allocation for Gold. This hybrid strategy contributes positively to societal welfare by protecting the retirement savings of millions of South Africans from excessive declines in value during periods of extreme macro-economic stress. By protecting capital during economic downturns, pension funds can meet their future obligations, thereby enhancing national financial well-being and contributing to greater economic resilience.

Conclusion, Limitation and Future Research

The South African gold market can be viewed through the lens of the Adaptive Market Hypothesis. This means there will never be a single model that works for every time period. By accepting the Adaptive Market Hypothesis as the primary issue of which model performs technically better, and providing a strategic guide (i.e., a contingency-based handbook) for decision-making. The LSTM provides tactical forecasts and identifies short-term patterns. On the other hand, the MS-GARCH model provides strategic risk management and long-term capital protection. The choice between the two is therefore based on users' objectives regarding the horizon, rather than on statistical competition. Finally, by providing better instruments for managing financial risks, this study can help create a stronger, fairer and more sustainable South African economy.

Limitations and Future Research

This paper's limitations (a narrow focus on a single asset and challenges in interpreting deep-learning models) are acknowledged. A future direction of work will have to focus on hybrid "AMH-in-action" models. For example, an LSTM model would be trained not only to predict the price but also to estimate the probability of the MS-GARCH regime. This would yield one model that balances tactical accuracy with risk assessment and is aware of regimes. Moreover, additional studies should integrate these economic models with macroeconomic and policy simulation models (e.g., iSDG models) to numerically project the effects of improved risk forecasts for the gold sector on the attainment of desired SDG indicators, such as employment, GDP growth, and poverty rates.

Further research directions include

The application of this conceptual study can be further developed along several significant avenues. In the first place, the modelling methodology may be employed to investigate other commodity markets to determine whether the observed features and relative model performances are transferable from the South African gold market. Secondly, future research should incorporate environmental, social, and governance(ESG) factors into the modelling process to account for the growing role of sustainability-related risks in influencing asset prices and in shaping downside risk exposure. Thirdly, developing online regime-identifying algorithms will improve these models' adaptability to changing regimes in real time, thereby enhancing their practicality and usefulness in managing dynamic risks. Fourthly, it is possible to develop hybrid neural–econometric architectures that combine the interpretative power of regime-switching models with the flexibility of deep learning methodologies, providing an even more complete and adaptable framework for financial forecasting in rapidly changing environments.

References

Ahmed, A. (2026). Explainable deep learning for financial volatility forecasting: An LSTM-attention-SHAP with comprehensive validation. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6301119

Ardia, D., Bluteau, K., Boudt, K., and Catania, L. (2019). Forecasting risk with Markov-switching GARCH models: A large-scale performance study. International Journal of Forecasting, 35(1), 233–243. https://doi.org/10.1016/j.ijforecast.2018.05.004

Awad, M., and Khanna, R. (2015). Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Apress.

Aye, G. C., Balcilar, M., Boubaker, H., and Gupta, R. (2015). Forecasting Aggregate Retail Sales: The case of South Africa. International Journal of Production Economics, 160, 66–79. https://doi.org/10.1016/j.ijpe.2014.09.033

Blake, A. C., Gandhi, N. A., and Jakkula, A. R. (2025). Improving S&P 500 Volatility Forecasting through Regime-switching Methods. Available online: arXiv. https://arxiv.org/abs/2510.03236 (Accessed on 15 August 2025)

Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Brownlee, J. (2019). Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery.

Changani, J. G. (2024). Factors Influencing Gold Price Movements: A Time Series Analysis Perspective. Available online: https://ssrn.com/abstract=4815102 (Accessed on 1 August 2025).

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

Fischer, T., and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Guidolin, M., and Hyde, S. (2009). Equity portfolio diversification under time-varying predictability: Evidence from Ireland, the US, and the UK. Journal of Multinational Financial Management, 18(4), 293-312. https://doi.org/10.1016/j.mulfin.2008.01.004

Hamilton, J. D., and Susmel, R. (1994). Autoregressive conditional heteroscedasticity and regime changes. Journal of Econometrics, 64(1–2), 307–333. https://doi.org/10.1016/0304-4076(94)90067-1

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. Econometrica, 79(2), 453–497. https://doi.org/10.3982/ECTA5771

Hewamalage, H., Bergmeir, C., and Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735--1780. https://doi.org/10.1162/neco.1997.9.8.1735

Ito, M., Noda, A., and Wada, T. (2014). International stock market efficiency: A non-Bayesian time-varying model approach. Applied Economics, 46(23), 2744–2754. https://doi.org/10.1080/00036846.2014.909579

Ji, Y., Lei, X., Zhang, L., Heng, J., and Fan, J. (2026). Forecasting the price of gold with integrated media sentiment: A prediction framework based on online news sentiment mining with CNN-QRLSTM. Entropy, 28(3), 271. https://doi.org/10.3390/e28030271

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. Empirical Economics, 27(2), 363–394. . https://doi.org/10.1007/s001810100100

Kumar, P. N., Umeorah, N., and Alochukwu, A. (2024). Dynamic graph neural networks for enhanced volatility prediction in financial markets.Available online: https://arxiv.org/abs/2410.16858 (Accessed on 27 July 2025)

Lefatsa, P. M., Msomi, S., Muguto, H. T., Muguto, L., and Muzindutsi, P.-F. (2025). An empirical analysis of the impact of global risk sentiment, gold prices, and interest rate differentials on exchange rate dynamics in South Africa. International Journal of Financial Studies, 13(3), 120. https://doi.org/10.3390/ijfs13030120

Lim, K. P., and Brooks, R. (2011). The evolution of stock market efficiency over time: A survey of the empirical literature. Journal of Economic Surveys, 25, 69–108. https://doi.org/10.1111/j.1467-6419.2009.00611.x

Lo, A. W. 2004. The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. The Journal of Portfolio Management 30 (5): 15–29. https://doi.org/10.3905/jpm.2004.442611.

Lo, A. W. (2017). Adaptive markets: Financial evolution at the speed of thought. Princeton University Press. Available online: https://doi.org/10.1515/9781400887767

McNeil, A. J., Frey, R., and Embrechts, P. (2015). Quantitative risk management: concepts, techniques and tools-revised edition. Princeton University Press. Available online: https://books.google.co.za/books?id=SfJnBgAAQBAJ

Mehmood, F., Ahmad, S., and Whangbo, T. K. (2023). An efficient optimisation technique for training deep neural networks. Mathematics, 11(6), 1360. https://doi.org/10.3390/math11061360

Milidonis, A., and Chisholm, K. (2024). The regime-switching structural default risk model. Risks, 12(3), 48. https://doi.org/10.3390/risks12030048

Minerals Council South Africa. (2024). Facts and figures 2024. Available online: https://www.mineralscouncil.org.za/industry-news/publications/facts-and-figures. (Accessed on 25 July 2025)

Mohammad Asif, S. S. (2025). Create an AI-based model for dynamic risk management in equity portfolios that accounts for extreme market events (e.g., pandemics, financial crises). Journal of Information Systems Engineering and Management, 10(20s), 26367. https://doi.org/10.52783/jisem.v10i20s.3010

Readhead, A., Tarus, V., Lassourd, T., Madzivanyika, E., and Schlenther, B. (2023). The future of resource taxation: 10 policy ideas to mobilise mining revenues. International Institute for Sustainable Development; African Tax Administration Forum. https://www.iisd.org/publications/guide/future-of-resource-taxation

Rohan, A., Hossen, M. D., Pranto, M. N., Hossain, B., Yoshi, A. M., and Islam, R. (2025). Artificial intelligence in financial market prediction: advancements in machine learning for stock price forecasting. Frontiers in Artificial Intelligence, 8, 1696423. https://doi.org/10.3389/frai.2025.1696423

South African Reserve Bank. (2024). Quarterly bulletin March 2024. Available online: https://www.resbank.co.za/en/home/publications/publication-detail-pages/quarterly-bulletins/quarterly-bulletin-publications/2024/FullQuarterlyBulletinNo311March2024.pdf (Accessed on 26 July 2025)

Sparf, P. N., and Rohdin, C. H. (2025). Forecasting value at risk with deep learning: Benchmarking LSTM models against conventional techniques. [Master’s thesis] School of Business, Economics and Law, Department of Economics, University of Gothenburg Publications Electronic Archive. Available online: https://hdl.handle.net/2077/88781 (Accessed on 26 July 2025)

Taneva-Angelova, G., Raychev, S., and Ilieva, G. (2025). A framework for gold price prediction combining classical and intelligent methods with financial, economic, and sentiment data fusion. International Journal of Financial Studies, 13(2), 102. https://doi.org/10.3390/ijfs13020102

Wang, Y., and Lin, T. (2024). A novel deterministic probabilistic forecasting framework for gold price with a new pandemic index based on quantile regression deep learning and multi-objective optimisation. Mathematics, 12(1), 29. https://doi.org/10.3390/math12010029