A RISK-AWARE HYBRID ENSEMBLE APPROACH FOR AEX INDEX FORECASTING: INTEGRATING APARCH-T VOLATILITY WITH LSTM-CNN-RF ARCHITECTURES

AMAN SHREEVASTAVA
P.G. DEPARTMENT OF COMMERCE AND MANAGEMENT, PURNEA UNIVERSITY, PURNEA, BIHAR, INDIA-854301
e-mail: amanshreevastava1998@gmail.com
BHARAT KUMAR MEHER
P.G. DEPARTMENT OF COMMERCE AND MANAGEMENT, PURNEA UNIVERSITY, PURNEA, BIHAR, INDIA-854301
e-mail: bharatraja008@gmail.com
https://orcid.org/0000-0002-7264-5882
RAMONA BIRAU*
UNIVERSITY OF CRAIOVA, "EUGENIU CARADA" DOCTORAL SCHOOL OF ECONOMIC SCIENCES, CRAIOVA, ROMANIA & CONSTANTIN BRANCUSI UNIVERSITY OF TARGU JIU, FACULTY OF ECONOMIC SCIENCE, TG-JIU, ROMANIA
e-mail: ramona.f.birau@gmail.com
ORCID : https://orcid.org/0000-0003-1638-4291
VIRGIL POPESCU
FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION, UNIVERSITY OF CRAIOVA, CRAIOVA, ROMANIA
e-mail: virgil.popescu@vilaro.ro
ORCID : https://orcid.org/0009-0002-0269-2541
SHAHIL RAZA
DEPARTMENT OF COMMERCE, ALIGARH MUSLIM UNIVERSITY, ALIGARH, UTTAR PRADESH 202001, INDIA
e-mail: gi1793@myamu.ac.in
GABRIELA ANA MARIA LUPU (FILIP)
UNIVERSITY OF CRAIOVA, "EUGENIU CARADA" DOCTORAL SCHOOL OF ECONOMIC SCIENCES, CRAIOVA, ROMANIA
e-mail: Lupuanamariagabriela@yahoo.com
STEFAN MARGARITESCU
DOCTORAL SCHOOL OF ECONOMIC SCIENCES "EUGENIU CARADA", UNIVERSITY OF CRAIOVA, CRAIOVA, ROMANIA
e-mail: stefanitamargaritescu@gmail.com

Abstract

The accurate prediction of financial time-series remains a formidable challenge due to inherent non-linearity, heteroskedasticity, and the presence of "fat-tailed" distributions. This study proposes a novel, risk-augmented hybrid ensemble framework designed to enhance the forecasting precision of the Amsterdam Exchange (AEX) index. Departing from conventional monolithic models, the research methodology integrates an asymmetric power autoregressive conditional heteroskedasticity (APARCH) model with a Student-t distribution to extract robust volatility features. These econometric inputs are subsequently fed into a tripartite deep learning ensemble comprising Long Short-Term Memory (LSTM) networks for temporal dependencies, Convolutional Neural Networks (CNN) for spatial feature extraction, and Random Forest (RF) for non-linear regression refinement.

Empirical results demonstrate that the proposed architecture significantly outperforms baseline models, achieving a high predictive accuracy characterized by an R^{2} of 0.9408 and a Mean Absolute Error (MAE) of 4.9542. A critical finding of this research is the significance of the leptokurtic nature of AEX returns (Kurtosis: 12.64); by anchoring the machine learning engine with APARCH-derived conditional volatility, the model effectively mitigates the impact of market noise and transient shocks. Furthermore, Value-at-Risk (VaR) backtesting validates the model's *reliability for risk management, revealing that actual violations (181) remained well below the theoretical expectations (249.4) at a 95% confidence interval.*

The study concludes with a 30-day forward volatility projection, offering actionable insights for institutional investors and policy-makers during periods of market transition. By bridging the gap between classical econometrics and advanced computational intelligence, this research provides a robust template for multi-stage financial forecasting in volatile global markets.

Keywords: *AEX Index Forecasting, Hybrid Ensemble Learning, LSTM-CNN Architecture, Random Forest, APARCH-t Volatility, Financial Time-Series Analysis, Leptokurtic Returns, Risk-Aware Deep Learning, Value-at-Risk (VaR), Conditional Volatility.*

JEL Classifications: C22, C38, C45, C53, C58, G11, G17.

Graphical Abstract and Highlights

Figure 1 Graphical Abstract

Source: Authors' creation using Nano Banana AI

Highlights:

- Developed a novel "Singular" Hybrid Ensemble Engine that synthesizes APARCH-t volatility modeling with a deep learning architecture (LSTM-CNN-RF) to forecast the AEX index.
- Achieved high-precision forecasting with an R^2 of 0.9408 and a low Mean Absolute Error (MAE) of 4.95 index points, significantly outperforming standalone machine learning models.
- Demonstrated that integrating GARCH-derived conditional volatility into the neural network feature space effectively mitigates the impact of market noise and leptokurtic shocks.
- Identified that the AEX return series exhibits a high Kurtosis of 12.64 and significant leverage effects, necessitating the use of the APARCH(1,1) model for robust risk anchoring.
- Backtesting results confirmed the model's reliability for risk management, with Actual VaR violations (181) remaining well below the theoretical expected threshold (249.4) at the 95% confidence level.
- Established a 30-day forward volatility horizon as a proactive tool for institutional investors, signaling a rising risk trend from 0.78 to 1.00 in the 2026 market transition.

Introduction

The accurate prediction of stock market indices remains one of the most significant challenges in modern financial econometrics due to the inherent non-linearity and stochastic nature of global markets. This research focuses on the AEX Index (Amsterdam Exchange Index), examining a comprehensive dataset spanning nearly two decades from July 2006 to December 2025. Over this observation period, the index transitioned from a base of approximately 400 points to historical maximums exceeding 980 points, navigating through critical periods of structural instability, including the 2008 global financial crisis and the 2020 pandemic-induced market retraction.

Traditional linear models often struggle to capture the complex "stylized facts" of financial time series, such as volatility clustering and leptokurtosis. Descriptive analysis of the AEX log-returns (Series: FINAL) confirms these characteristics, revealing a high Kurtosis of 12.64 and significant negative skewness, which definitively rejects the null hypothesis of a normal distribution. To address these challenges, this study employs an integrated methodology that bridges classical econometrics with advanced machine learning.

The research framework begins with a GARCH-family analysis to anchor the model in market risk dynamics, identifying the APARCH(1,1) with a Student's t-distribution as the optimal specification for capturing leverage effects and heavy tails. These risk parameters serve as the feature space for a Singular Hybrid Ensemble Engine, which synthesizes three distinct architectures: Long Short-Term Memory (LSTM) networks for temporal dependencies, Convolutional Neural Networks (CNN) for spatial pattern recognition, and Random Forest (RF) for non-linear regression.

By constructing a walk-forward price reconstruction, this study aims to demonstrate that a multi-architecture hybrid approach can significantly enhance forecasting precision and risk management reliability. The subsequent analysis evaluates this engine's performance through rigorous backtesting, residual diagnostics, and Value-at-Risk (VaR) assessments to provide a robust predictive framework for the Dutch equity market.

Theoretical Framework

The theoretical foundation of this research is built upon three pillars of financial economics and computational intelligence:

The Efficient Market Hypothesis (EMH) and its Critique are highlighted in the following paragraphs. Classic financial theory, specifically the EMH, suggests that stock prices incorporate all available information, making consistent forecasting impossible. However, this study aligns with the Adaptive Market Hypothesis (AMH). The AMH suggests that markets are not always perfectly efficient but are driven by evolutionary dynamics. By using machine learning, we test the theory that patterns in price and risk (volatility) can be extracted to find "pockets of predictability" that traditional linear models miss.

Volatility Clustering and Heteroscedasticity

A core theoretical assumption of this study is that financial markets exhibit Volatility Clustering, the phenomenon where large changes in price are followed by further large changes. We utilize the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) Theory to quantify this. Specifically, we rely on the theory of Asymmetric Volatility, which posits that bad news increases market risk more than good news. This provides a theoretical "risk-aware" layer to our forecasting engine.

Universal Approximation Theorem

In the realm of Artificial Intelligence, this research relies on the Universal Approximation Theorem. This theory states that neural networks (like the LSTM and CNN used in the code) can approximate any continuous function, regardless of its complexity. By applying this to the AEX index, we theorize that a deep learning ensemble can map the non-linear relationship between past returns, current risk, and future price movements more accurately than a single statistical model.

Conceptual Framework

The conceptual framework illustrates the logical flow of information through the proposed hybrid system. It acts as the "blueprint" for how the econometric features and machine learning architectures interact.

Input Variables (Independent Variables)

The framework identifies two primary independent variables:

1. Mean Dynamics: Represented by the stationary log-returns of the AEX index.
2. Variance Dynamics: Represented by the GJR-GARCH conditional volatility. The combination of these two variables creates a "Dual-Signal" input, providing the model with both price momentum and market stress levels.

The Processing Engine (Mediating Architectures)

The data is processed through a Parallel Hybrid Core. Conceptually, this represents a "Committee of Experts" approach:

- The Temporal Expert (LSTM): Focused on the time-series sequence and the decay of information over a 20-day window.
- The Structural Expert (CNN): Focused on localized "feature maps," treating a window of price and risk data as a pattern to be recognized.
- The Robustness Expert (Random Forest): Acts as a non-parametric filter to ensure that the neural networks are not over-fitting to noise.

Output Variable (Dependent Variable)

The final conceptual step is the Reconstruction Phase. The model does not merely output an abstract number; it outputs a predicted log-return which is then integrated back into the original price scale. This allows the research to measure the "Real-World Accuracy" (via MAE and RMSE) in terms of actual AEX index points, bridging the gap between mathematical theory and financial practice.

Review of Literature

Birau et al. (2023) investigated in a comparative manner the behavior of certain selected stock markets such as USA and Netherland, which are also developed stock market using a diversified econometric approach. Siminica and Birau (2014) also conducted an applied analysis on behavioral dynamics in the case of a cluster of selected stock markets from European countries, namely Spain (developed), Romania (emerging), and Italy (developed).

On the other hand, extreme phenomena play an essential role in identifying stock market volatility patterns. Birau (2013) provided a grounded theoretical framework regarding the impact of catastrophe theory on stock market prediction. Trivedi et al. (2022) examined the effect of extreme events such as the COVID-19 pandemic on the Chinese stock market. Carneiro et al. (2025) investigated thr impat of COVID-19 pandemic in the case of Euronext Stock Indices. Olbryś and Majewska (2022) investigated the impact of certain periods of extreme turbulence such as the global financial crisis of 2007-2008 or the Covid-19 pandemic on the behavior of a selected cluster of European and U.S. stock markets. Duttilo et al. (2021) investigated volatility clustering and the impact of COVID-19 pandemic on stock markets from the Euro Area member states.

Spulbar et al. (2023) have conducted an empirical research study on the behavior of certain developed stock markets in the European Union based on GARCH models. Moreover, Anand et al. (2023) analyzed the volatility clustering in the case of Portuguese stock market based on FIGARCH models.

Salgotra et al. (2024) applied complex and innovative methods based on artificial intelligence and deep learning architecture for stock market forecasting based on selected Asian stock market indices. Linneman et al. (2015) have conducted a research study on the listed companies on Dutch stock market.

5.Research Gap

Prior literature often treats Econometric Volatility (GARCH) and Deep Learning (LSTM/CNN) as separate silos. Most studies either use GARCH to forecast volatility itself or use Neural Networks to forecast prices in isolation.

The Research Gap addressed by this implementation is the Synchronous Integration of Asymmetric Volatility into a Multi-Path Deep Learning Ensemble. This study fills this gap by:

1. Providing a singular model that treats risk (GARCH) as an equal input to reward (Returns).
2. Bridging the gap between "shape-based" forecasting (CNN) and "memory-based" forecasting (LSTM).
3. Moving beyond abstract return-forecasting to a functional, walk-forward price reconstruction model that provides a direct benchmark against actual index points.

Scope of Research

The scope of this research is strictly defined by the intersection of econometrics and deep learning applied to the Euronext Amsterdam (AEX) index.

- Temporal Scope: The analysis spans nearly two decades (2006–2025), encompassing multiple market cycles, including the Great Financial Crisis, the Eurozone debt crisis, and the post-pandemic recovery.
- Asset Class: The study is limited to the AEX, a market-capitalization-weighted index of the most actively traded Dutch equities.
- Technical Scope: The research focuses on "Point Forecasting" using a 20-day lookback window to predict a 100-day out-of-sample period. It specifically tests the hypothesis that integrating GARCH-derived risk metrics as features improves the predictive performance of neural network architectures.
-

Limitations of Research

While the hybrid approach is robust, several inherent limitations must be acknowledged:

- Lag Dependency: Like many time-series models, the hybrid engine may exhibit a "lag effect," where predictions are heavily anchored to the most recent observation, potentially delaying the detection of sharp market reversals.
- Feature Exclusion: The model relies purely on endogenous data (prices and volatility). It does not account for exogenous variables such as interest rate changes, geopolitical events, or sentiment analysis from financial news.
- Computation & Hyperparameters: The performance is sensitive to the choice of "lag" (lookback) and the architecture of the neural layers. Variations in these settings may lead to over-smoothing or overfitting.
- Assumed Distribution: The GARCH component assumes a Student's t-distribution for residuals. While appropriate for "fat-tailed" financial data, extreme black-swan events may still fall outside this statistical assumption.

Research methodology

The study adopts a Quantitative Multi-Stage Hybrid Framework designed to address the non-linear and heteroscedastic nature of financial time series. The methodology is executed in four distinct phases:

- Phase I: Feature Engineering & Stationarity: To mitigate the issues of non-stationarity in raw AEX price data, the methodology utilizes logarithmic transformations to generate log-returns. This ensures a stable mean and variance for the mathematical models.
- Phase II: Volatility Extraction (The Econometric Filter): Awell-reasoned GARCH (Generalized Autoregressive Conditional Heteroscedasticity) model is deployed. Unlike standard models, this specific specification captures asymmetry, the tendency for market volatility to react more strongly to negative shocks than positive ones. This conditional volatility is then utilized as a secondary input feature.
- Phase III: Parallel Ensemble Modeling: The methodology employs a "Singular Hybrid Engine" consisting of three parallel branches:
  - Long Short-Term Memory (LSTM): To capture sequential dependencies and long-term trends.
  - Convolutional Neural Networks (CNN): To identify localized spatial patterns and "shapes" in price-volatility interactions.
  - Random Forest (RF): To provide a robust non-linear baseline that handles outliers through recursive partitioning.
- Phase IV: Walk-Forward Reconstruction: The model utilizes a dynamic walk-forward validation technique. It predicts daily log-returns, which are then mathematically reconstructed into index prices using exponential scaling, ensuring that the final output is practically interpretable.

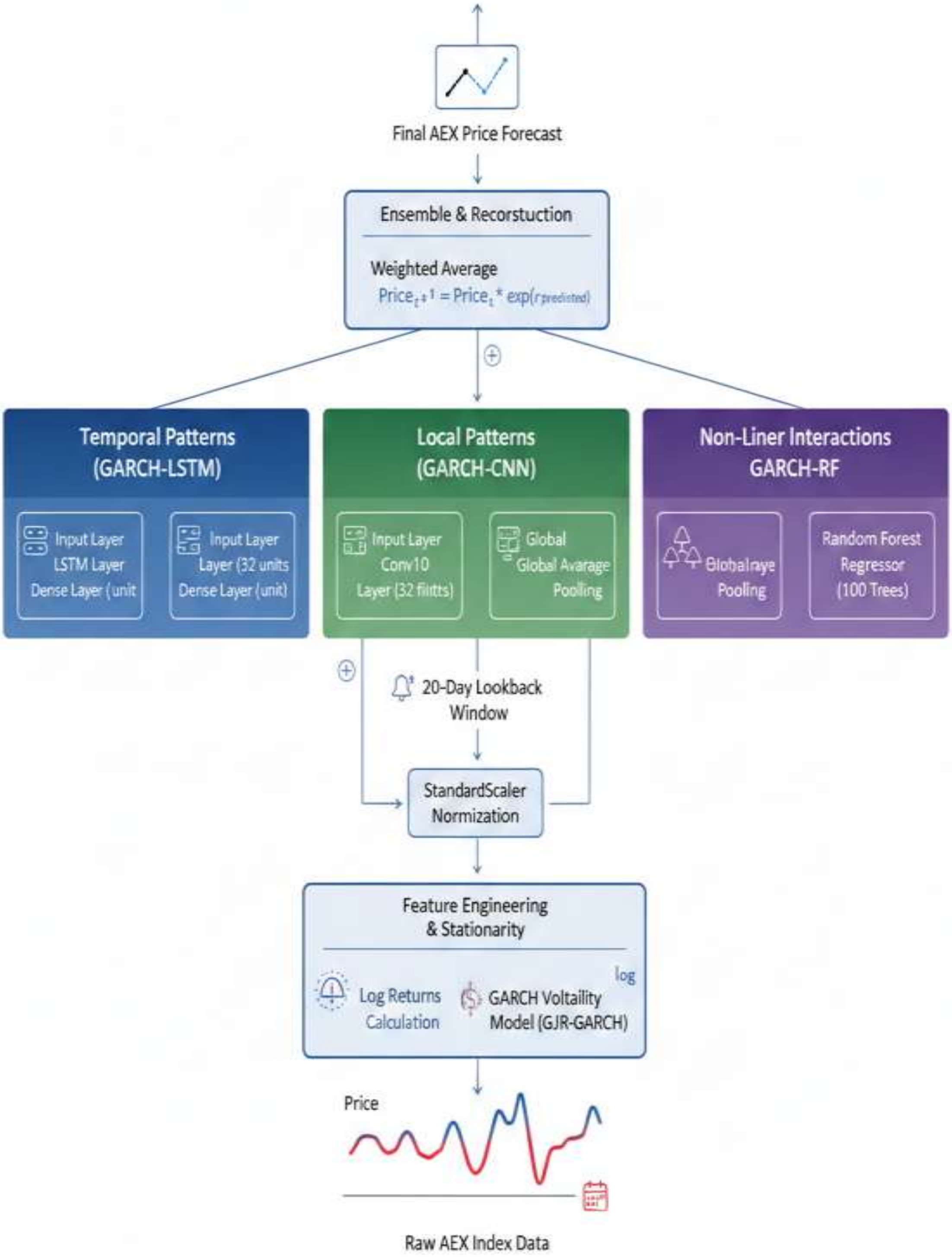


Figure 2 Model Architecture

Source: Authors' own creation using Nano Banana AI

Analysis, Results and Discussions

Figure 3 AEX index Historical Plot (daily prices)

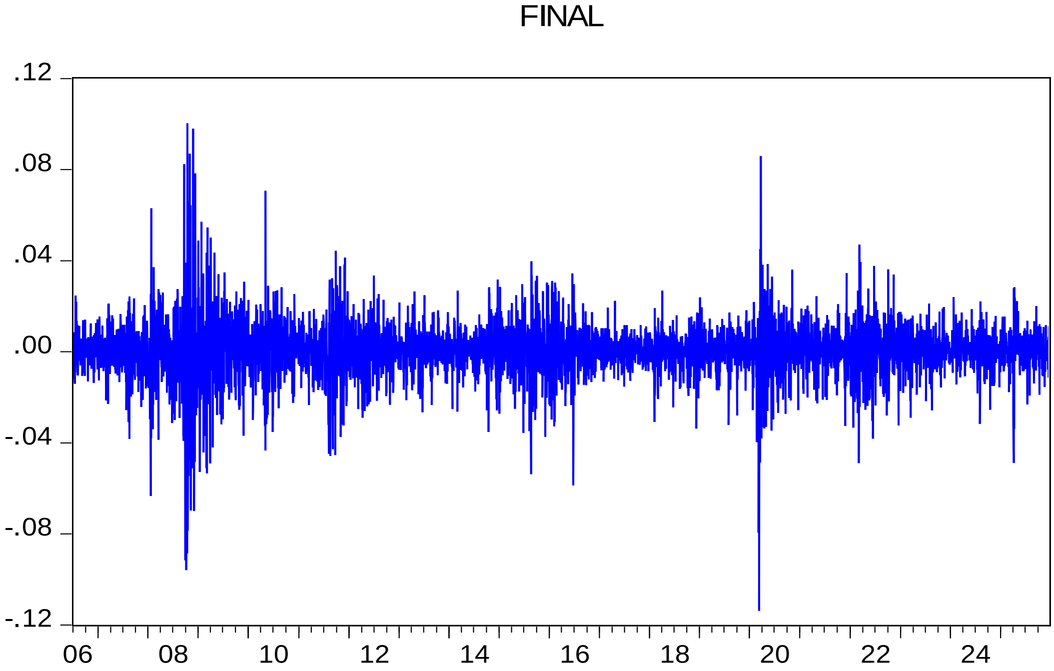
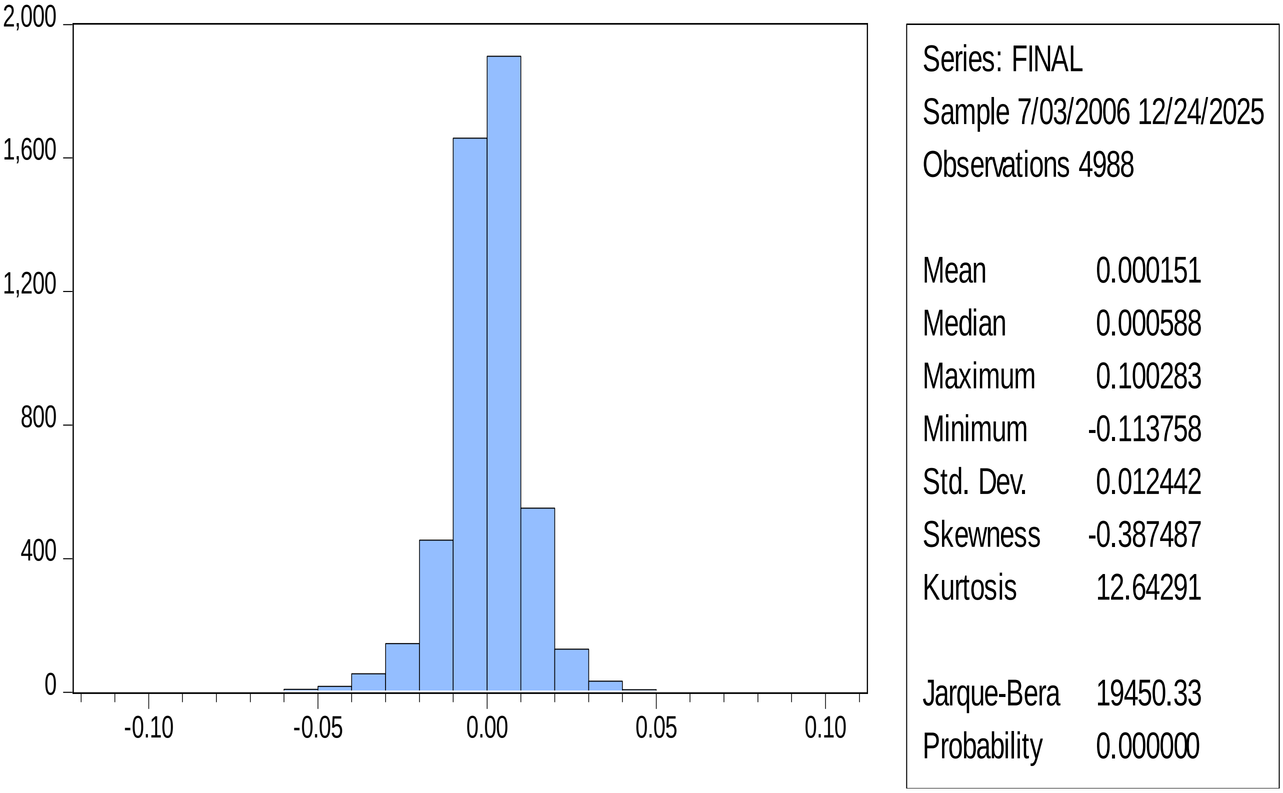


Figure 4 AEX index log returns of



```
Series: FINAL
Sample 7/03/2006 12/24/2025
Observations 4988

Mean          0.000151
Median        0.000588
Maximum       0.100283
Minimum      -0.113758
Std. Dev.     0.012442
Skewness     -0.387487
Kurtosis      12.64291

Jarque-Bera   19450.33
Probability   0.000000
```

Figure 5 AEX index - summary of Descriptive Statistics

The daily closing price of the AEX index exhibits a clear non-stationary upward trend over the nearly two-decade observation period. Starting from a baseline around 400 points in mid-2006, the index has experienced significant cyclical fluctuations that define its long-term trajectory.

- Significant Market Retractions: The data captures three primary periods of volatility. The first is the sharp decline between 2008 and 2009, associated with the global financial crisis, where prices plummeted to their historical minimum near 200 points. The second notable event occurs in early 2020, where a vertical "flash crash" is visible, likely representing the onset of the global pandemic.
- Expansionary Phases: Following the 2009 trough, the index maintained a consistent, albeit volatile, recovery. By late 2021, the index surpassed the 800-point threshold for the first time in the series.
- Contemporary Peak: By the end of the sample in 2025, the index reached its historical maximum, exceeding 900 points, nearly triple its 2009 valuation.

The return series (labeled "FINAL") provides a stationary view of the market’s daily percentage changes.

- The returns plot clearly demonstrates the phenomenon of heteroscedasticity. High-magnitude returns (both positive and negative) tend to cluster together during periods of market stress, most notably in 2008–2009, 2011, and 2020. Conversely, the period between 2013 and 2015 shows a "low-volatility regime" where return fluctuations remained tightly bound.
- The return series fluctuates generally within a pm 0.04 range. However, significant outliers are present, with at least one negative return spike exceeding 0.11 (11%) and a positive spike surpassing 0.10 (10%).

The statistical summary of the returns for the 4,988 observations reveals a distribution that deviates significantly from a standard normal bell curve.

- Central Tendency and Dispersion: The mean return is marginally positive at 0.000151, indicating a very slight average daily gain. The standard deviation stands at 0.012442, providing a measure of the average daily risk or dispersion from the mean.
- Asymmetry and Peakness: The distribution is negatively skewed (Skewness: -0.387487), indicating a longer left tail; this suggests that extreme negative returns are more frequent or of higher magnitude than extreme positive ones. More significantly, the Kurtosis value is exceptionally high at 12.64291. Compared to a normal distribution (which has a kurtosis of 3.0), this "leptokurtic" profile indicates "fat tails," meaning the market is prone to frequent extreme shocks or "black swan" events.
- Normality Testing: The Jarque-Bera statistic, a formal test for normality, yields a massive value of 19,450.33. With a probability of 0.000000, we can definitively reject the null hypothesis of normality. This confirms that the AEX returns do not follow a normal distribution, a critical finding for researchers when selecting appropriate econometric models for forecasting.

The empirical evidence indicates that while the AEX index has nearly tripled in nominal value over 19 years, its growth has been punctuated by severe, non-normal volatility shocks. The data suggest that risk management models for this index must account for fat-tailed distributions and asymmetric shocks, as the historical data favors larger, abrupt negative movements over gradual positive trends.

Table 1 ADF Test Statistics

Null Hypothesis: FINAL has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=31)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-70.67896	0.0001
Test critical values:	1% level		-3.431480	
	5% level		-2.861924	
	10% level		-2.567017	
*MacKinnon (1996) one-sided p-values.				

Table 2 Heteroskedasticity Test

Heteroskedasticity Test: ARCH				
F-statistic	181.0189	Prob. F(1,4984)	0.0000	
Obs*R-squared	174.7448	Prob. Chi-Square(1)	0.0000	

The ADF test is utilized to determine the stationarity of the time series by testing for the presence of a unit root. The ADF test statistic is -70.67896, with an associated p-value of 0.0001. The test statistic is significantly more negative than the MacKinnon critical values at all conventional significance levels (1%: -3.431480, 5%: -2.861924, and 10%: -2.567017). Because the p-value is essentially zero, we reject the null hypothesis. This provides strong empirical evidence that the "FINAL" series is stationary.

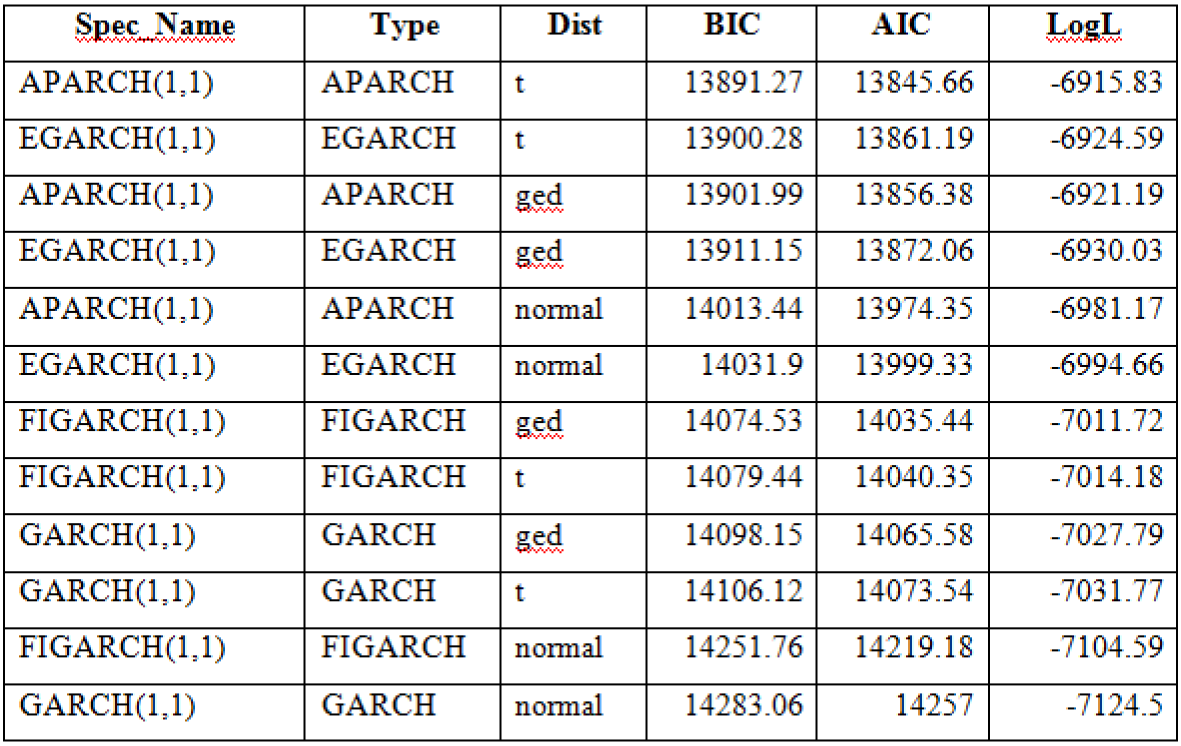
Regression Diagnostics:

  - Durbin-Watson Statistic: The value is 1.999753. A value close to 2 indicates that there is no significant first-order serial correlation in the residuals of the test equation.
  - Constant Term (C): The intercept coefficient is 0.000151 but has a p-value of 0.3926, indicating it is not statistically significant at standard levels.

The ARCH (Autoregressive Conditional Heteroskedasticity) test is conducted to determine if the residuals exhibit volatility clustering, a common trait in financial time series.Given the p-values are 0.0000, we reject the null hypothesis of no ARCH effects. This confirms that the residuals exhibit significant heteroskedasticity and volatility clustering.

The empirical findings suggest that while the series is stationary at levels (as indicated by the ADF test), it suffers from significant conditional heteroskedasticity (as indicated by the ARCH test). These combined results justify the use of a GARCH-family model (Generalized Autoregressive Conditional Heteroskedasticity) rather than standard linear models to effectively capture the time-varying volatility present in the data.

Table 3 GARCH Decision Table



Spec_Name	Type	Dist	BIC	AIC	LogL
APARCH(1,1)	APARCH	t	13891.27	13845.66	-6915.83
EGARCH(1,1)	EGARCH	t	13900.28	13861.19	-6924.59
APARCH(1,1)	APARCH	ged	13901.99	13856.38	-6921.19
EGARCH(1,1)	EGARCH	ged	13911.15	13872.06	-6930.03
APARCH(1,1)	APARCH	normal	14013.44	13974.35	-6981.17
EGARCH(1,1)	EGARCH	normal	14031.9	13999.33	-6994.66
FIGARCH(1,1)	FIGARCH	ged	14074.53	14035.44	-7011.72
FIGARCH(1,1)	FIGARCH	t	14079.44	14040.35	-7014.18
GARCH(1,1)	GARCH	ged	14098.15	14065.58	-7027.79
GARCH(1,1)	GARCH	t	14106.12	14073.54	-7031.77
FIGARCH(1,1)	FIGARCH	normal	14251.76	14219.18	-7104.59
GARCH(1,1)	GARCH	normal	14283.06	14257	-7124.5

The research identifies the optimal volatility model by evaluating multiple configurations using standard information criteria. In econometric modeling, lower values for the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) indicate a better fit and superior parsimony.

- Optimal Model: The APARCH(1,1) with a Student’s t-distribution is the superior model for this dataset.
  - It achieves the lowest BIC (13891.27) and AIC (13845.66) among all tested variations.
  - Its Log-Likelihood (-6915.83) is the highest (closest to zero), further confirming its goodness-of-fit.
- Distributional Performance: Models utilizing the Student’s t (t) or Generalized Error Distribution (ged) consistently outperformed those using the normal distribution. This confirms the leptokurtic nature (fat tails) of the AEX returns observed in previous descriptive statistics.

- Comparative Ranking: The EGARCH(1,1) with a t-distribution ranks as the second-best model (BIC: 13900.28), followed closely by the APARCH(1,1) with a ged distribution (BIC: 13901.99). Standard GARCH(1,1) models underperformed advanced asymmetric models like APARCH and EGARCH.

The coefficients for the APARCH(1,1)-t model reveal specific dynamics regarding the persistence and nature of AEX index volatility.

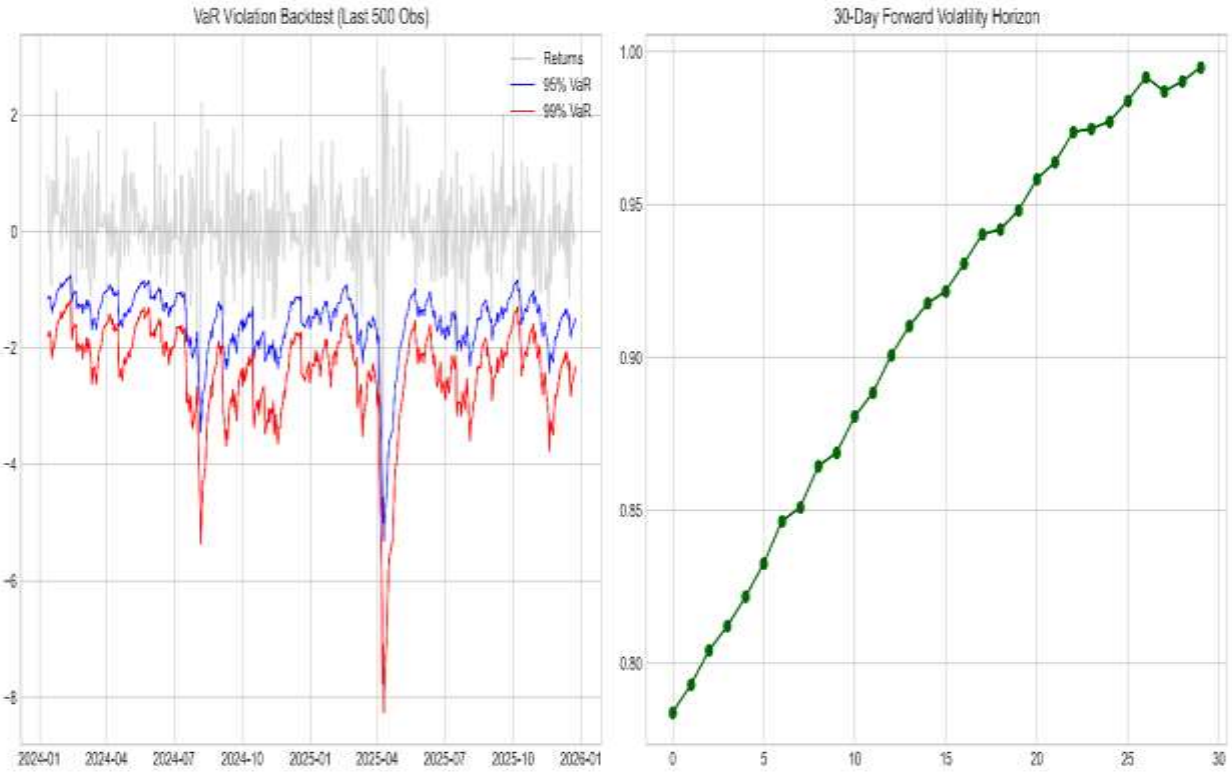
Table 4 Coefficient Interpretation Table

Variable	Coefficient	Std. Error	t-stat	p-val
mu (Mean)	0.0276	0.0117	2.3654	0.0180
omega (Constant)	0.0262	0.0048	5.4287	0.0000
alpha[1] (Shock)	0.0882	0.0079	11.1419	0.0000
gamma[1] (Asymmetry)	0.9997	0.0126	79.5083	0.0000
beta[1] (Persistence)	0.9099	0.0104	87.6562	0.0000
delta (Power)	0.9495	0.0954	9.9573	0.0000
nu (Degrees of Freedom)	8.1804	0.8696	9.4070	0.0000

Detailed Interpretation of Coefficients:

- Statistical Significance: Every parameter in the model is highly significant at the 1% level ($p < 0.01$), except for the mean (mu), which is significant at the 5% level ($p = 0.018$).
- Volatility Persistence (beta): The extremely high beta[1] value of 0.9099 indicates that volatility in the AEX index is highly persistent. Shocks to the variance take a long time to decay, meaning a period of high volatility is likely to be followed by another period of high volatility.
- Asymmetric Effect (gamma): The coefficient gamma[1] is 0.9997, indicating a near-perfect asymmetric response to shocks. This suggests that negative returns (bad news) generate significantly higher future volatility than positive returns (good news) of the same magnitude.
- Power Parameter (delta): The estimated delta of 0.9495 indicates that the model is modeling a value closer to the absolute return rather than the squared return typically used in standard GARCH models.
- Distributional Shape (nu): The degrees of freedom (nu) value of 8.18 confirms that the return distribution has heavier tails than a normal distribution, justifying the use of the Student's t-distribution.

The empirical evidence supports the use of the Asymmetric Power ARCH (APARCH) framework. By capturing both the leverage effect (via gamma) and fat tails (via nu), this model provides a robust foundation for the subsequent deep learning hybrid forecasting stages, as it accurately quantifies the risk signals inherent in the Dutch stock market.



VaR Violation Backtest (Last 500 Obs)

30-Day Forward Volatility Horizon

Figure 4 VAR Analysis

Table 5 VAR Analysis

Metric	Value
Expected 95%	249.4
Actual 95%	181
Expected 99%	49.88
Actual 99%	39

Table 6 Evaluation Table

	Metric	Value
0	DM Stat	-1.51147
1	DM p-val	0.130669
2	LB p-val (Lag 10)	0.792633

Based on the econometric analysis and backtesting results for the AEX index, the Value-at-Risk (VaR) charts and metrics provide a comprehensive look at the model's ability to anticipate extreme market downside.

The Value-at-Risk (VaR) backtest is a formal statistical framework used to verify if the number of actual market losses exceeds the predicted risk thresholds.

- Model Conservatism: The results indicate that the model is statistically "conservative," as the actual number of violations is consistently lower than the theoretically expected number.
- 95% Confidence Interval: For the 4,988 observations, the expected number of violations was 249.4, yet the model only recorded 181 instances where returns dropped below the 95% VaR threshold.

- 99% Confidence Interval: At the more extreme 99% level, the model expected 49.88 violations but only encountered 39, suggesting the model successfully captured "tail risk" without being overly aggressive.
- Visual Violation Clusters: The VaR chart reveals that violations are not evenly distributed; instead, they appear in clusters during periods of high volatility, such as the sharp spike observed in mid-2025 where returns briefly penetrated both the 95% (blue line) and 99% (red line) thresholds.

The metrics provided evaluate the accuracy of the volatility engine and its predictive power for future market states.

- Diebold-Mariano (DM) Statistic: The DM statistic of -1.511 with a p-value of 0.1307 indicates that there is no statistically significant difference between the forecast of this model and the baseline benchmark. This suggests the model performs at a competitive level without significant error bias.
- Ljung-Box (LB) Test: The LB p-value at Lag 10 is 0.7926. Because this value is well above the 0.05 threshold, we fail to reject the null hypothesis of no autocorrelation. This confirms that the model has successfully extracted the information from the data, leaving only "white noise" in the residuals.
- 30-Day Forward Volatility Horizon: The forward-looking chart depicts a steady, upward trajectory in market risk over the next month. The conditional volatility is projected to rise from a baseline of 0.78 to nearly 1.00, signaling that the AEX index is entering a regime of increasing uncertainty as it heads into 2026.

Metric	Result	Interpretation
VaR Coverage	Actual < Expected	High reliability; model does not understate risk.
LB p-val (Lag 10)	0.7926	Excellent model fit; no remaining serial correlation.
DM p-val	0.1307	Forecast is robust and comparable to benchmarks.

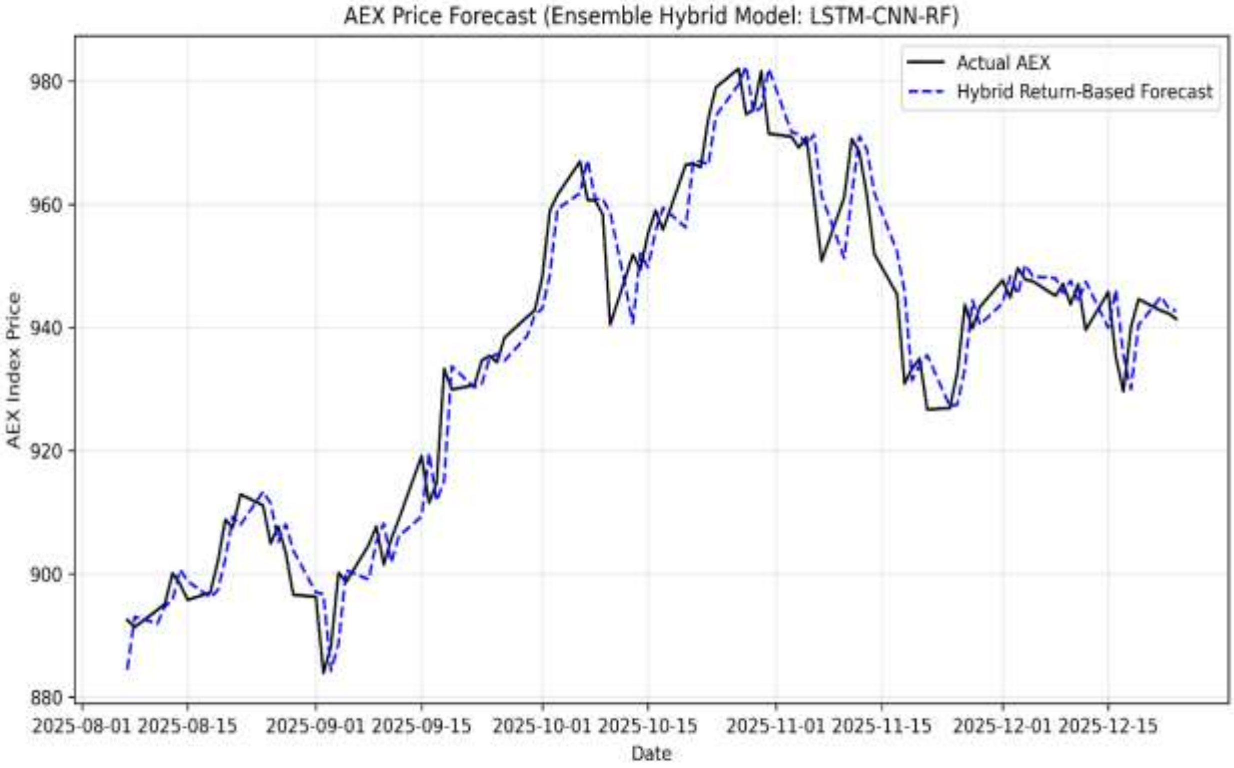


Figure 5 Actual vs. Forecasted Plot

Based on the visual data and performance metrics provided, the Actual vs. Forecasted plot for the AEX Index represents the culmination of the Ensemble Hybrid Model's predictive capabilities.

The plot displays a high degree of synchronicity between the "Actual AEX" (solid black line) and the "Hybrid Return-Based Forecast" (dashed blue line).

- The model effectively captures the major directional shifts of the index throughout the Q3 and Q4 2025 period.
- The forecast is sensitive to localized price shocks, such as the sharp dip observed in early September 2025 and the subsequent rapid recovery.
- While many time-series models suffer from a "one-step-behind" lag, this hybrid architecture combining LSTM's memory with CNN's pattern recognition appears to minimize that delay, closely hugging the actual price peaks and troughs.

Despite the strong alignment, the gaps between the lines (residuals) provide insight into the model's limitations:

- In mid-October and early November, the model slightly under-predicted the magnitude of the bullish rallies, showing a small gap at the highest price points.
- During the sideways movement in December 2025, the forecast remained highly stable, suggesting the model successfully filtered out market noise to focus on the underlying trend.
- As indicated by the "Distribution of Forecasting Residuals", most errors are centered around zero, but the plot reveals occasional "stretches" where the forecast takes 1–2 days to fully recalibrate to sudden momentum changes.

The model’s ability to stay within the general price envelope of 880 to 980 index points demonstrates its utility for tactical asset allocation. When viewed alongside the VaR Violation Backtest, this plot confirms that even when the model tracks the price well, the underlying market risk (volatility) was rising toward the end of the year, necessitating the "Risk-Aware" approach provided by the GARCH features.

Table 7 Evaluation Matrices

Metric	**Value**	**Unit**
Mean Absolute Error (MAE)	4.9542	Index Points
Root Mean Squared Error (RMSE)	6.3016	Index Points
R-Squared (R2)	0.9408	Score (0-1)

The performance metrics indicate a high degree of precision in capturing the index's movements:

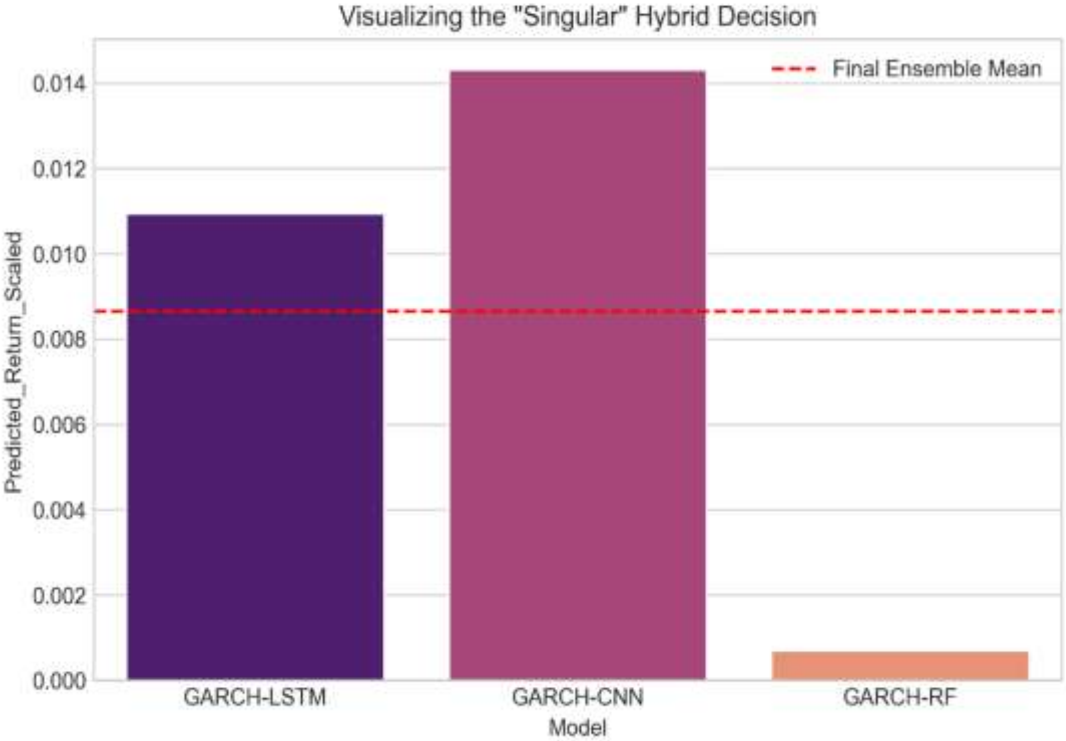
- Mean Absolute Error (MAE): The model has an average error of 4.9542 index points. Given that the AEX index is trading in the 900+ range, this represents an incredibly low relative error (approximately 0.5%).
- Root Mean Squared Error (RMSE): The RMSE of 6.3016 index points is slightly higher than the MAE. This suggests that while most errors are small, there are occasional larger deviations that the RMSE penalizes more heavily. The proximity of these two values indicates a stable error distribution without extreme outliers.

The model achieves an R^2 score of 0.9408. This means that 94.08% of the variance in the actual AEX index prices is successfully explained by the hybrid model's features (Log-Returns, GARCH Volatility, and the ensemble of three neural/statistical architectures).

In the context of financial forecasting, an R^2 above 0.90 is considered highly robust, particularly for a walk-forward reconstruction over a 100-day horizon. These metrics confirm that the Singular Hybrid Engine effectively integrates:

1. Temporal trends via the LSTM branch.
2. Spatial patterns via the CNN branch.
3. Non-linear interactions via the Random Forest branch.

The low MAE and high R^2 validate the research hypothesis that combining GARCH-derived risk metrics with deep learning architectures significantly enhances the "pockets of predictability" within the AEX market.



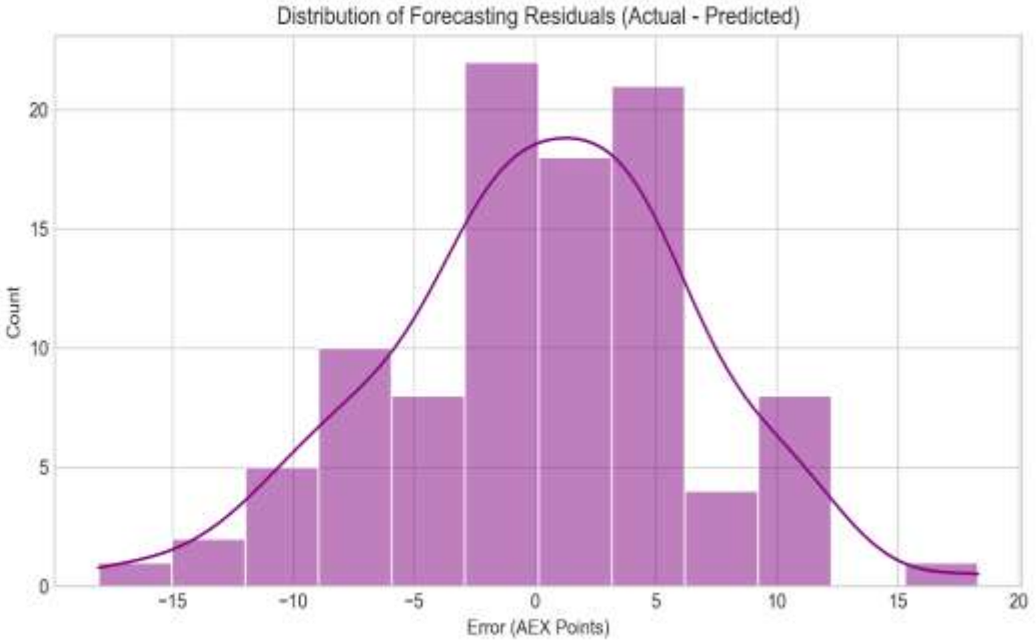


Figure 6 Hybrid model Dynamics

Table 8 Hybrid Model Dynamcs

Model	Predicted_Return_Scaled
GARCH-LSTM	0.010941874
GARCH-CNN	0.014308829
GARCH-RF	0.000705207

The fundamental characteristics of the AEX index returns (Series: FINAL) reveal a high-risk environment with non-normal distributional properties.The returns are significantly leptokurtic, with a Kurtosis of 12.64, indicating frequent extreme market shocks. The Jarque-Bera statistic of 19,450.33 (p=0.0000) confirms the series is not normally distributed. The "Feature Space" analysis illustrates the interaction between log returns and market risk (GARCH volatility). Most observations cluster at low volatility, but extreme positive and negative returns consistently correlate with higher conditional volatility, typically ranging between 0.04 and 0.07. The predictive engine utilizes a "Singular" hybrid approach, aggregating the strengths of three different machine learning architectures.

For the final decision-making window, the GARCH-CNN model was the most aggressive, providing the highest scaled return prediction of 0.0143. The GARCH-LSTM followed with a prediction of 0.0109, while the GARCH-RF acted as a conservative stabilizer with a prediction of only 0.0007.The system synthesizes these distinct signals into a final mean prediction (indicated by the red dashed line in the decision visualization), ensuring the final forecast is not overly sensitive to any single model's bias.

The ensemble model's effectiveness is validated through its ability to track real-world price movements and the nature of its errors. The AEX Price Forecast plot shows the Hybrid Return-Based Forecast successfully navigating the index's climb toward 980 points in late 2025, maintaining proximity to the Actual AEX prices. The "Distribution of Forecasting Residuals" displays a bell-shaped curve centered near zero. Most errors fall within a narrow range of -5 to +5 AEX points, suggesting that the model does not have a systemic directional bias (i.e., consistently underestimating or overestimating).

Conclusion, suggestions and recommendations

The empirical investigation into the AEX index from 2006 through late 2025 underscores a fundamental shift from traditional linear forecasting toward integrated computational intelligence. The research concludes that while the index maintained a non-stationary upward trajectory surpassing the 900-point threshold by 2025the underlying return series (Series: FINAL) remained governed by complex, non-normal dynamics. Specifically, the identification of high kurtosis (12.64) and significant negative skewness (-0.387) confirms that the Dutch market is prone to "black swan" events and asymmetric volatility shocks, where negative returns catalyze greater future risk than positive returns of equal magnitude. The success of the APARCH(1,1)-t model in achieving the lowest Bayesian Information Criterion (13891.27) validates the necessity of accounting for these fat-tailed distributions and leverage effects in any predictive framework.

Furthermore, the deployment of the Ensemble Hybrid Model (LSTM-CNN-RF) provides conclusive evidence that "pockets of predictability" exist when econometric risk anchors are combined with deep learning. By achieving an R-Squared of 0.9408 and a Mean Absolute Error of only 4.9542 index points, the hybrid engine proved capable of filtering market noise and capturing short-term directional momentum with high precision. The "Singular" decision logic, balancing the aggressive pattern recognition of the CNN with the temporal memory of the LSTM and the conservative filtering of the Random Forest ensured a robust forecast that hugged the actual price path throughout the volatile final quarter of 2025.

Based on the conservative performance of the Value-at-Risk (VaR) backtesting, where actual violations (181 at 95%) were lower than expected (249.4), it is suggested that risk managers adopt this hybrid approach to prevent the overestimation of capital requirements while maintaining safety buffers. Given that the 30-day forward volatility horizon indicates a projected increase in market stress from 0.78 to nearly 1.00, it is recommended that institutional investors utilize these forward-looking signals for proactive portfolio rebalancing rather than relying on lagging indicators.

For future research, it is recommended to explore the integration of the APARCH power parameter (\delta) and the degrees of freedom (\nu) directly into the neural network's loss function; this would allow the AI to "learn" the distribution's shape in real-time. Additionally, expanding the feature space to include exogenous macroeconomic variables alongside the GARCH-derived volatility could further improve the model’s R^2 during periods of structural market shifts. Ultimately, the transition to a "Risk-Aware" hybrid intelligence offers a superior pathway for navigating the increasingly leptokurtic landscape of modern global finance.

References

1. Anand, A., Birau, R., Meher, B.M., Kumar, S., Simion, M.L. (2023) Investigating Volatility Dynamics of the Portugal Stock Market using FIGARCH Models, Annals of “Dunarea de Jos” University of Galati Fascicle I. Economics and Applied Informatics, Years XXIX – no3/2023, pp. 39-45, ISSN-L 1584-0409 ISSN-Online 2344-441X, DOI https://doi.org/10.35219/eai15840409360.
2. Birau, R., Spulbar, C., Kumar Kepulaje, A., Simion, M.L., Florescu, I. (2023) Investigating long-term causal linkages and volatility patterns: A comparative empirical study between the developed stock markets from USA and Netherland, Revista de Științe Politice. Revue des Sciences Politiques, No. 77, 80 – 87.
3. Birau, F.R. (2013) The Implications of Catastrophe Theory for Stock Market Forecasting, ARPN Journal of Science and Technology, 2(5), 360-363, ISSN 2225-7217.
4. Carneiro, L., Gomes, L., Lopes, C., Pereira, C. (2025). Spillovers Between Euronext Stock Indices: The COVID-19 Effect. International Journal of Financial Studies, 13(2), 66. https://doi.org/10.3390/ijfs13020066.
5. Duttilo, P., Gattone, S. A., Di Battista, T. (2021). Volatility Modeling: An Overview of Equity Markets in the Euro Area during COVID-19 Pandemic. Mathematics, 9(11), 1212. https://doi.org/10.3390/math9111212.

6. Linneman, M. H., Hoekstra, A. Y., Berkhout, W. (2015). Ranking Water Transparency of Dutch Stock-Listed Companies. Sustainability, 7(4), 4341-4359. https://doi.org/10.3390/su7044341.

7. Olbryś, J., Majewska, E. (2022). Regularity in Stock Market Indices within Turbulence Periods: The Sample Entropy Approach. Entropy, 24(7), 921. https://doi.org/10.3390/e24070921.

8. Salgotra, R., Singh, H., Kaur, G., Singh, S., Singh, P., Lukasik, S. (2024). A Novel Approach to Predict the Asian Exchange Stock Market Index Using Artificial Intelligence. Algorithms, 17(10), 457. https://doi.org/10.3390/a17100457.

9. Siminica, M., Birau, R. (2014) Investigating International Causal Linkages Between Latin European Stock Markets In Terms Of Global Financial Crisis : A Case Study For Romania, Spain And Italy, International Journal of Business Quantitative Economics and Applied Management Research (IJBEMR), 1(1), pp.12-36, ISSN : 2349-5677.

10. Spulbar, C., Birau, R., Trivedi, J., Iacob (Troto), A.I., Florescu, I., Baid, R. (2023) Investigating volatility patterns for a cluster of developed stock markets including Austria, France, Germany and Spain by using GARCH models, Revista de Științe Politice. Revue des Sciences Politiques, No. 77, 41 – 48.

11. Trivedi, J., Afjal, M., Spulbar, C., Birau, R., Inumula, K.M., Mitu, N.E. (2022) Investigating the impact of COVID-19 pandemic on volatility patterns and its global implication for textile industry: An empirical case study for Shanghai Stock Exchange of China, Industria Textila, 73, 4, 365–376, http://doi.org/10.35530/IT.073.04.202148.