Assessing Dynamic Hedge Effectiveness and Statistical Arbitrage with Conventional and Deep Learning Models in Equity Futures

Komal Batool
Department of Mathematics, NED University of Engineering and Technology, Karachi 75270, Pakistan

Zaara Asim Mirza
Department of Mathematics, NED University of Engineering and Technology, Karachi 75270, Pakistan

Ammar Sheikh
Department of Mathematics, NED University of Engineering and Technology, Karachi 75270, Pakistan

Hanadi Sabir
Department of Mathematics, NED University of Engineering and Technology, Karachi 75270, Pakistan

Hussain Aftab
Department of Mathematics, NED University of Engineering and Technology, Karachi 75270, Pakistan

ABSTRACT
This study advances the frontier of financial risk management by rigorously comparing conventional econometric models with modern deep learning approaches for dynamic hedging in equity futures. Using weekly data from October 2019 to June 2024 on the KSE-30 Index and its futures, the authors of this study examine whether established techniques such as DCC-GARCH and GARCH-Copula can match the adaptability and predictive strength of advanced architectures, including LSTM–CNN hybrids and the FT-Net Hybrid. Optimal hedge ratios are estimated on a dynamic basis, with performance assessed through variance reduction, RMSE, Sharpe ratios, hedge effectiveness, and directional accuracy along with 4 other metrics. Beyond risk mitigation, the study extends and applies the Fuzzy TOPSIS framework for optimal model selection and testing statistical arbitrage opportunities between the models. The results highlight the transformative potential of deep learning in capturing complex market dynamics that traditional models often overlook, offering actionable insights for traders, portfolio managers, and policymakers in emerging markets.

General Terms
Time series forecasting, mathematical models

Keywords
Volatility Modelling, Hybrid Deep Learning Models, Fourier Transform in Finance, Multivariate GARCH Models, Fuzzy TOPSIS, Quantitative Trading.

1. INTRODUCTION
The management of portfolio risk remains a central concern for investors and institutions, particularly in emerging markets such as Pakistan, where equity markets are shaped by volatility, liquidity constraints, and evolving regulation. While global literature has extensively analyzed futures markets for volatility transmission, risk management, and price discovery [1] [2], the Pakistani market suffers from scarce evidence on effective hedging strategies using futures contracts [3]. Traditional static models such as OLS [4] and ECM [5] provided the foundation for estimating hedge ratios, yet their inability to capture time-varying covariances, heteroskedasticity, and autocorrelation limited their effectiveness. This led to the adoption of ARCH/GARCH models proposed by [6] and [7], along with their variants which, despite their ability to model volatility clustering, rely on restrictive assumptions of linearity and normality that often fail in real-world applications.

Recent advances in computational finance have shifted the landscape toward deep learning, enabling models capable of capturing nonlinearities and long-term dependencies without imposing rigid assumptions. In particular, LSTM networks, hybrid LSTM–CNN models, and spectral architectures such as the FT-Net Hybrid have demonstrated strong performance in option hedging and risk forecasting [8]. While most research applies these tools to option markets, futures contracts, especially in underexplored markets like Pakistan, offer an equally critical application space. Furthermore, new propositions suggest that discrepancies between classical econometric and deep learning–based hedge ratios may themselves generate exploitable "statistical arbitrage between models" [9] [10] thereby expanding the scope of arbitrage beyond traditional price misalignments.

1.1 Significance and Motivation
This study is motivated by three central gaps. First, despite the growing role of derivatives in emerging markets, hedging research on the Pakistan Stock Exchange remains underdeveloped. Second, econometric frameworks, though they are sophisticated, they still struggle to capture nonlinearities and regime shifts. whereas deep learning models flexibly approximate these dynamics. Yet comparative evidence between the two paradigms remains sparse. Finally, inspired by the statistical arbitrage perspective of François [9], this research extends the inquiry beyond hedge effectiveness into the possibility that model-based discrepancies themselves may form systematic, zero-cost trading opportunities.

1.2 Aims and Objectives
The aims are threefold: (i) to design dynamic hedging strategies for the KSE-30 Index using its futures contracts, employing both econometric (DCC-GARCH, Copula-GARCH) and deep learning models (LSTM–CNN, FT-Net Hybrid); (ii) to evaluate and contrast their hedging effectiveness across multiple quantitative criteria, including variance reduction, RMSE, hedged return, Sharpe ratio, VaR/CVaR reduction, and directional accuracy; and (iii) to investigate whether model discrepancies give rise to statistically significant arbitrage opportunities.

1.3 Methodology
To capture time-varying hedge ratios, the authors of this study

first employ the DCC-GARCH framework with ARMA–eGARCH margins, later enhanced with a Student-t copula to account for symmetric tail dependence under market stress [11]. Parallelly, they apply two advanced deep learning models: the LSTM–CNN hybrid, where convolutional layers extract local temporal features before LSTM layers capture long-term dependencies; and the FT-Net Hybrid, which incorporates Fourier spectral modules to detect cyclical patterns and non-stationarities, followed by convolutional and recurrent/transformer blocks for dynamic learning. Model performance is rigorously evaluated using nine hedging effectiveness metrics, with Fuzzy TOPSIS employed to aggregate results across criteria. To complement effectiveness analysis, hedged P&L series are examined for statistical arbitrage opportunities, providing insights into market equilibrium and model-driven trading prospects.

1.4 Contributions of the Study

This research contributes in four distinct ways. First, it pioneers the application of cutting-edge deep learning—LSTM–CNN and FT-Net Hybrid—alongside Copula-GARCH models directly to futures hedging in Pakistan, bridging a significant regional literature gap. Second, it extends statistical arbitrage beyond price misalignments by empirically testing the arbitrage potential between model-implied hedge ratios. Third, it introduces Fourier-transform–based spectral analysis into deep hedging pipelines, enhancing the detection of cyclical and non-stationary structures in volatile emerging markets. Finally, through the integration of a multi-criteria decision-making framework, this study offers a transparent, practitioner-relevant comparison of econometric and machine learning models. Beneficiaries include institutional investors seeking adaptive risk management tools and academics pursuing methodological innovation in underexplored markets.

2. LITERATURE REVIEW

Hedging risk exposure in equity markets through derivatives—particularly stock index futures—has been one of the central topics in financial risk management. The effectiveness of a hedge depends on accurately estimating hedge ratios, and research has evolved from static econometric methods toward dynamic and machine learning–based approaches. This chapter reviews the literature a systematic review that highlights methodological progress, limitations, and emerging trends. Through this, the authors of this research paper establish the foundation for their study while identifying research gaps that motivate their contribution.

2.1 Traditional and Econometric Approaches

Early studies such as Ederington [4] employed OLS regression to estimate minimum variance hedge ratios (MVHRs). While simple, OLS assumes constant relationships between spot and futures returns and often leads to under-hedging. Error Correction Models by Ghosh [5] introduced cointegration to capture short- and long-term linkages, yet they too assumed time-invariant relationships. Subsequent research established that hedge ratios are inherently dynamic [12] [2]. This led to GARCH-type models, which captured volatility clustering and time-varying conditional correlations. Key extensions include BEKK-GARCH [13], CCC-GARCH [7] [14], and DCC-GARCH [15]. More recently, copula-based GARCH models [11] [16] were developed to account for nonlinear and asymmetric dependencies, particularly in tail events. Despite these advances, GARCH-type frameworks are often criticized for overestimating volatility persistence and failing to adapt to sudden regime shifts [17].

2.2 Emergence of Deep Hedging

The limitations of econometric models opened the door for deep learning approaches, which rely less on rigid statistical assumptions and more on data-driven pattern recognition. The "Deep Hedging" framework proposed by [18] demonstrated that neural networks can learn optimal hedges directly by minimizing risk objectives without relying on Greeks or distributional assumptions. Subsequent works in studies [19] [20] extended this to commodity and equity derivatives, consistently reporting superior performance relative to GARCH-based hedges. LSTM architectures, in particular, have shown strong ability to capture nonlinear dependencies and long-term dynamics in financial time series [21].

2.3 Hybrid and Advanced Architectures

Hybrid models combining LSTM with CNN layers have been widely adopted due to their ability to jointly capture local short-term price movements and long-term dependencies. Applications in futures markets, such as EU Emissions Trading Scheme contracts, demonstrated up to a 43% reduction in MAPE compared with single-model benchmarks [22]. Incorporating attention mechanisms further improved adaptability in volatile and illiquid markets. Building on these successes, FT-Net Hybrid architectures emerged in 2024, integrating Fourier-transform spectral modules with temporal convolutional blocks [23]. Preliminary evidence suggests FT-Net models reduce hedged return variance by an additional 5–8% compared to CNN–LSTM hybrids, particularly effective in environments with cyclical and non-stationary dynamics [24].

2.4 Statistical Arbitrage Between Models

A novel strand of literature extends hedging beyond variance reduction by exploring statistical arbitrage opportunities between models. [9] and [10] argue that in complete markets, discrepancies between hedge ratios from traditional replicating portfolios and deep learning models may generate zero-cost arbitrage strategies. This reconceptualization shifts arbitrage from temporary mispricing to systematic differences in model assumptions and learning dynamics. Such a perspective is particularly relevant for emerging markets, where market inefficiencies amplify the potential profitability of model-based arbitrage.

3. DATA AND PROCESSING

3.1 Data Description

This study examines hedging effectiveness for a spot equity portfolio of the KSE-30 index using both conventional econometric and deep learning frameworks. The analysis relies on daily data from January 2019 to June 2024, sourced from the Pakistan Stock Exchange (PSX) Data Portal. The dataset includes:

1. Index cash price: Spot value of the KSE-30 index.
2. Interest/financing rate: Derived from KIBOR-based cost of carry, crucial for futures valuation.
3. Index dividend yield: Expected annual yield for adjustment of fair value.
4. Days to expiry: Contract maturity horizon.
5. Indicative fair value: Benchmark settlement price for all outstanding futures contracts.

Following the PSX rulebook, indicative fair value is computed as:

$$Indicative\ fair\ value = Underlying\ Index \times \left(1 + r\left(\frac{x}{365}\right)\right) - d \quad (1)$$

where r is the financing rate, x denotes days to expiry, and d represents dividend accruals.

For empirical work, the authors of this study focus on the nearest-to-expiry three-month KSE-30 futures contract, rolling positions monthly to maintain continuity.

3.2 Handling Missing Data

A central challenge was the presence of missing values, particularly for the indicative fair value and financing rates. While missing fair values could be recomputed via the PSX formula, missing interest rates required a more sophisticated imputation strategy.

To address this, 1-month KIBOR was utilized from the State Bank of Pakistan as an explanatory factor and implemented a Random Forest Regressor to interpolate missing financing rates [24]. This ensemble model captures non-linear dependencies and avoids parametric restrictions typical in time-series econometrics. Figure 1 below shows the architecture and pipeline used.

Training was performed on observed financing rate entries, with evaluation based on R^2, MSE, and Pearson correlation. Across 1000 seeds, the best model achieved:

- $R^2 = 0.9092$
- $MSE = 0.0003$

These values confirm that over 90% of the variance was explained, ensuring high-fidelity imputations. The computed missing interest rates were then incorporated back into Equation (1) to yield a complete and internally consistent futures dataset, suitable for econometric and machine learning analysis.

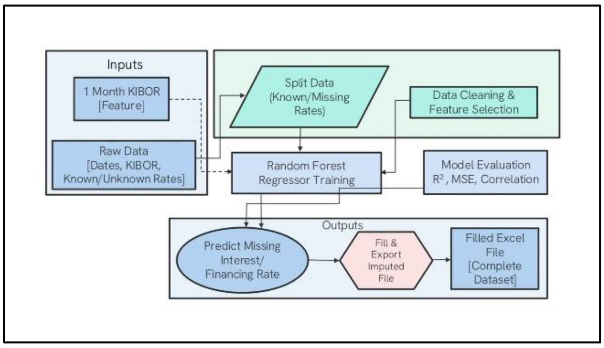

Figure 1. Architecture and Pipeline of the Random Forest Regressor

3.3 Futures Contract Adjustments

Rolling futures contracts introduces artificial "price jumps" at expiry due to the divergence between the expiring contract (converging to spot) and the newly listed contract (reflecting cost-of-carry and risk premia). These jumps can distort volatility estimates, hedge ratios, and model performance [25].

To eliminate such distortions, mean-back adjustment method was applied, which smooths rollovers by shifting expiring contracts against the new series, inspired from [26]. Specifically:

1. Mean Roll-Over Price:

$$M_{roll} = \frac{P_{t_{roll}}^{(\text{exp})} + P_{t_{roll}}^{(\text{new})}}{2} \quad (2)$$

2. Roll-Over Difference:

$$D_{roll} = P_{t_{roll}}^{(\text{exp})} - M_{roll} \quad (3)$$

3. Adjusted Historical Prices:

$$P_{t_{roll}}^{(\text{exp,adj})} = P_{t}^{(\text{exp})} \pm D_{roll} \quad (4)$$

where $P_{t_{roll}}^{(\text{exp})}$ is the price of the expiring contract at time t within its active month. The adjustment was applied uniformly across each contract's life, ensuring seamless transitions and preserving intra-month volatility dynamics.

3.4 Dataset Integrity

The combined data pipeline, Random Forest–based imputation for missing rates and mean-back adjustment for contract rollovers, produced a continuous, high-quality time series of futures prices. This ensures that subsequent econometric estimations and deep learning models operate on a dataset free from artificial breaks, preserving the statistical and economic validity of hedge effectiveness tests.

4. RESEARCH METHODOLOGY

This chapter outlines the methodological framework adopted to construct, estimate, and evaluate hedge ratios for the KSE-30 index using futures contracts. The methodology is divided into three key stages:

1. Pre-estimation diagnostics: Statistical testing of the time series properties to ensure modeling assumptions are satisfied.
2. Modeling approaches: Application of econometric and machine learning methods to estimate time-varying hedge ratios.
3. Validation and robustness checks: Post-estimation diagnostics, performance metrics, and comparative evaluation.

Through combining rigorous econometric testing with advanced deep learning models, this study establishes a robust foundation for dynamic hedging in emerging equity markets.

4.1 Pre-Estimation Diagnostics

4.1.1 Stationarity Testing: Augmented Dickey-Fuller (ADF) Test

Time series stationarity is a fundamental requirement for volatility modeling and hedge ratio estimation. The authors of this study applied the Augmented Dickey-Fuller (ADF) test [27] to both spot and futures return series. The test evaluates the null hypothesis of a unit root (non-stationarity) using the regression specification with a constant and linear trend.

The lag length k, was selected following the ARMA upper bound rule:

$$k = trunc((length(x) - 1)^{1/3} \quad (5)$$

This yielded a lag value of 6 for the data used. The results strongly rejected the null of a unit root at the 1% significance level across both return series (futures and spot), confirming stationarity.

4.1.2 Heteroscedasticity Testing: Engle's ARCH Test

Volatility clustering—periods of high and low variance—is a hallmark of financial returns. To confirm its presence, Engle's ARCH test was applied from the study in [15], which regresses squared residuals on their lags. The test statistic is:

$$LM = T \times R^2 \quad (6)$$

where T is sample size and R^2 is the coefficient of determination from the auxiliary regression. Under the null hypothesis (no ARCH effects), LM follows a χ^2 distribution.

Using the ARCHTest() function in R (FinTS package), significant ARCH effects were found in both the spot and futures return series, consistent with volatility clustering. Figure 2 and Figure 3 illustrate the ACF plots of residuals and squared residuals, confirming persistence in volatility.

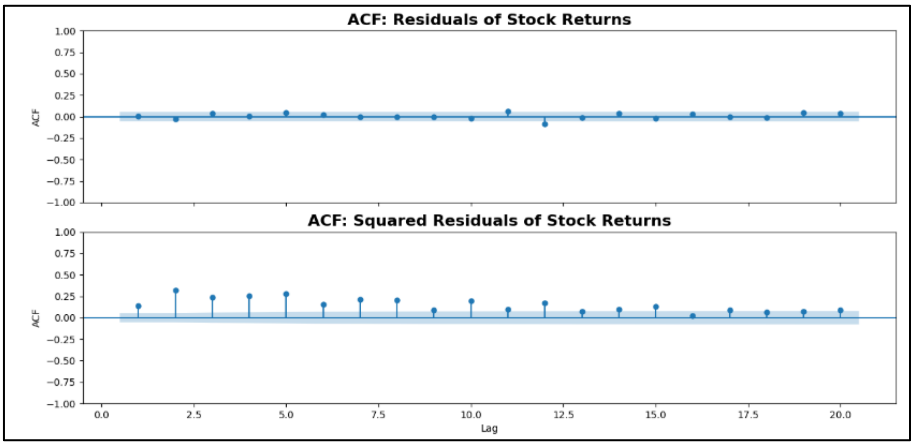

Figure 2. ACF plots of residuals and squared residuals of KSE 30 Index Return.

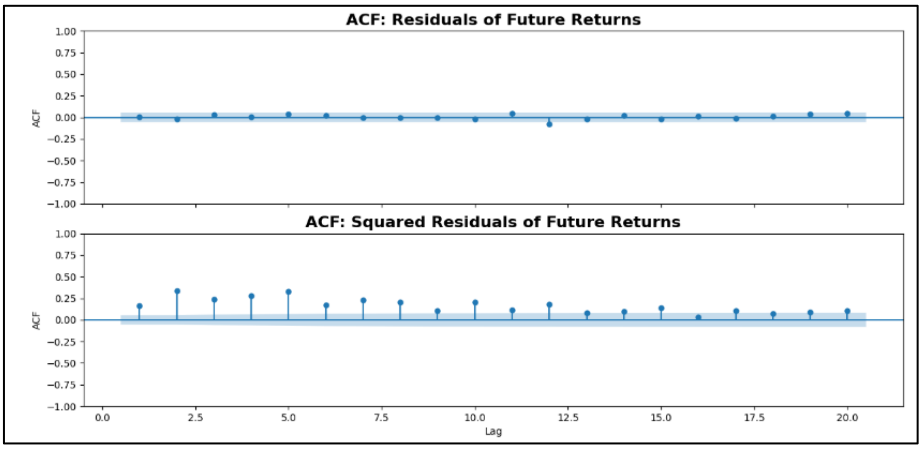

Figure 3. ACF plots of residuals and squared residuals of KSE 30 Index Futures Return.

4.2 Minimum Variance Hedge Ratio Estimation Framework

The hedging objective is to minimize portfolio variance by optimally combining spot and futures positions. Let R_s denote spot returns and R_f future returns. A hedged portfolio return is:

$$R_v = R_s - NR_f \quad (7)$$

The portfolio variance is given by:

$$\sigma_v^2 = \sigma_s^2 + N^2\sigma_f^2 - 2N\sigma_{s,f} \quad (8)$$

Minimizing variance requires differentiating Equation (8) with respect to N:

$$\frac{\partial \sigma_v^2}{\partial_N} = -2N\sigma_f^2 + 2\sigma_{s,f} \quad (9)$$

Setting this to zero yields the optimal hedge ratio:

$$N^* = \frac{\sigma_{sf}}{\sigma_f^2} \quad (10)$$

This framework, known as the Minimum Variance Hedge Ratio (MVHR), is the benchmark for all subsequent modeling.

4.3 Econometric Models

To capture time-varying dynamics, advanced GARCH-based models are employed:

DCC-eGARCH: extends the study of [15] Dynamic

Conditional Correlation framework, allowing for asymmetric responses to shocks. It captures both volatility spillovers and correlation dynamics between spot and futures returns.

Copula-DCC GARCH: Incorporates non-linear dependence structures via copulas, enabling more accurate modeling of tail dependencies, particularly during market stress.

Both models are specifically designed to handle volatility clustering and conditional correlation, making them highly suitable for hedge ratio estimation.

4.4 Deep Learning Models

Traditional econometric models, while powerful, often struggle with capturing non-linear, long-memory effects in high-frequency financial data. To address this, deep learning is integrated by:

1. LSTM–CNN Hybrid: Combines Long Short-Term Memory (LSTM) networks' ability to model sequential dependencies with Convolutional Neural Networks (CNNs), which extract local temporal features. This hybrid excels at capturing both short-term volatility shocks and long-term structural dependencies.
2. Fourier Transform Network (FT-Net): Applies Fourier decomposition to extract frequency-domain features from the return series before feeding them into a neural network. This enhances the model's ability to capture cyclical dynamics, a key feature of financial time series.

The deep learning models are benchmarked against econometric models to test whether data-driven, non-parametric approaches can outperform classical frameworks in hedging accuracy.

5. MODEL SPECIFICATION AND IMPLEMENTATION

5.1 Dynamic Conditional Correlation (DCC) GARCH

Inspired from [28], which highlighted that for enhanced volatility modelling the Exponentially Weighted Moving Average (EWMA) can be combined with the GARCH model. The EWMA models volatility while assigning greater weight to recent events and the GARCH models account for volatility clustering therefore, combined they can portray a more accurate representation of market dynamics. So, before fitting the DCC GARCH model on the training data the returns of the KSE 30 index and its future contracts are standardized. These standardized returns were calculated using the equation 11 below.

$$Z_{i,t} = \frac{R_{i,t} - \mu}{\sigma_t} \qquad (11)$$

Where, $R_{i,t}$ is the raw log return of the asset i, and μ is the mean of the returns. The standard deviation of the returns was estimated using the EWMA method. The EWMA method is a powerful technique for modelling conditional volatility because it captures volatility clustering without requiring a full GARCH specifications. The EWMA variance is estimated using the following equation 12:

$$\sigma_{i,t}^2 = \lambda\sigma_{i,t-1}^2 + (1-\lambda)R_{i,t-1}^2 \qquad (12)$$

Here $\sigma_{i,t}^2$ is the variance of the raw log returns of asset i. Since the raw returns of the training data are converted into standardized returns while forecasting the volatilities of the testing data, so they are converted to standardized forms as well using standard deviations estimated in the training data, as seen in figure 4.

The DCC GARCH model is a multivariate GARCH model proposed by [15] which is a generalization of the constant conditional correlation (CCC) GARCH estimators [14]. The DCC framework is a two-step methodology as follows:

5.1.1 Univariate modelling using GARCH

The DCC model proposed by [15] requires the use of univariate GARCH models to estimate the conditional variance of individual assets' returns, specifically the KSE 30 index returns and KSE 30 futures contracts returns. To model the conditional variances, the Exponential GARCH (eGARCH) model introduced by [29] was used, which, unlike standard GARCH models, allows for asymmetry in the impact of positive and negative shocks on volatility. This is an essential feature in financial time series data, such as the one used in this research, where negative news trends tend to increase volatility more than positive news of the same magnitude.

This research defines the return process using an ARMA (1,1) model represented by equation 13 below.

$$Z_{i,t} = \mu_i + \emptyset_1 Z_{i,t-1} + \theta_1\epsilon_{i,t-1} + \epsilon_{i,t} \qquad (13)$$

where $Z_{i,t}$ is the standardized return of asset i, μ_s is the constant mean and $\epsilon_{i,t}$ is a white noise error term with zero mean and constant variance. Then the eGARCH (2,1) is used in equation 14 from [29] to estimate the volatilities of KSE 30 index and KSE 30 index future contracts returns.

$$\begin{aligned}\log h_{i,t} = \omega_i + \beta_{i,1}log h_{i,t-1} \\ +\beta_{i,2}log h_{i,t-2} + \propto_i\left(\frac{\epsilon_{i,t-1}}{\sqrt{h_{i,t-1}}}\right) \\ +\gamma_i\left(\left|\frac{\epsilon_{i,t-1}}{\sqrt{h_{i,t-1}}}\right| - E\left|\frac{\epsilon_{i,t-1}}{\sqrt{h_{i,t-1}}}\right|\right)\end{aligned} \qquad (14)$$

Where $h_{i,t}$, is the variance of asset i. The model parameters were estimated using Maximum Likelihood Estimation (MLE) in R via the ugarchspec() and ugarchfit() functions from the rugarch package.

5.1.2 Dynamic Conditional Correlation Estimation

After estimating conditional volatilities, the second step was to estimate the dynamic conditional correlation matrices. To do so, the methodology presented in [15] where let $\epsilon_t = H_t^{1/2} z_t$ be the vector of residuals and $z_t \sim N(0,1)$ and H_t is the conditional covariance matrix as presented in below.

$$H_t = D_t R_t D_t \qquad (15)$$

Where D_t is a diagonal matrix of time-varying standard deviation from univariate eGARCH models:

$$D_t = diag(\sqrt{h_{1,t}}, \sqrt{h_{2,t}}, \dots, \sqrt{h_{3,t}} \quad (16)$$

The DCC model assumes that the standardized residuals $\eta_t = D_t^{-1}\epsilon_t$ have a time-varying correlation structure governed by a GARCH-type equation:

$$Q_t = (1 - a - b)S + a\eta_{t-1}\eta_{t-1} + bQ_{t-1} \quad (17)$$

Where Q_t is the time-varying covariance matrix of standardized residuals, S is the unconditional variance of η_t. The dynamic correlation matrix is then obtained by:

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \quad (18)$$

The DCC model was estimated using the dccspec() and dccfit() functions from the rmgarch package in R, with univariate eGARCH models used for the marginal distributions.

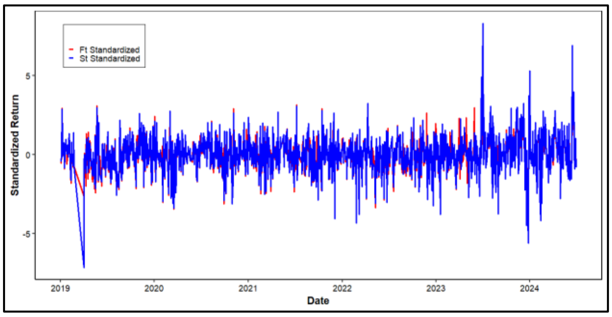

Figure 4. Standardized Returns of KSE 30 Index and its Future Contracts

5.2 Dynamic Copula DCC GARCH Model

Copula-GARCH models combine [15] DCC-GARCH with copula theory, developed further by [30], [31] and [16]. Despite the fact that the DCC-GARCH model presents a comprehensive approach for modelling time-varying correlation between assets, it relies on the assumption that the joint distribution of the standardized residuals is multivariate normal. However, financial time series often exhibit non-linear dependencies, especially in the tails of the distribution (tail dependence), which the normal distribution may not adequately capture. To address this limitation, a Copula-based DCC-GARCH approach is employed that allows us to capture complex dependencies and nonlinear co-movements, by focusing on the collective and individual behavior of the assets returns. Therefore, to optimize the hedge ratios and reduce the variance, ARMA-eGARCH margins combined with DCC correlation structure and a student-t Copula adequately captures changing volatilities and tail dependence. [31] [32]

5.2.1 Pre-Copula Fitting Procedure

As mentioned in the DCC-GARCH methodology; to remove the noise and stabilize volatility, an Exponentially Weighted Moving Average (EWMA) filter is applied for standardized returns in training and testing data. After this, the first step in the Copula-DCC GARCH framework involves modeling the marginal distributions of each return series to account for volatility clustering and asymmetry. Similar to the earlier section titled "Univariate modelling using GARCH", the KSE 30 spot index and futures returns are modelled using Exponential GARCH (eGARCH) processes. The residuals $\varepsilon_{i,t}$ from the eGARCH models are standardized to obtain standardized residuals:

$$z_{i,t} = \frac{\varepsilon_{i,t}}{\sqrt{h_{i,t}}} \quad (19)$$

Assuming normality of standardized residuals, the authors transform them to uniform margins using the cumulative distribution function (CDF) of the standard normal distribution, represented in equation 20 below.

$$u_{i,t} = \Phi(z_{i,t}) \quad (20)$$

This transformation ensures the marginal uniformity required for copula estimation. The resulting $u_{i,t} \in (0, 1)$ are then used to model the joint distribution of returns. The output of this step serves as the input for both the DCC estimation and the copula transformation steps that follow.

The estimation of the dynamic conditional correlation (DCC) model follows the methodology already outlined in the preceding section, based on the framework introduced in [15]. To avoid repetition, the authors of this study refer the reader to the earlier section titled "Dynamic Conditional Correlation Estimation", where the equations governing the DCC model (Equations 15-18) are presented in detail. These equations define how the conditional correlation matrix R_t evolves over time using past standardized residuals and a weighted moving average structure.

5.2.2 Copula Estimation

Once the residuals are transformed into uniform margins, a copula function is used to model the dependence structure. To flexibly model the joint distribution between the standardized residuals of spot and futures returns beyond linear correlation, a bivariate Student-t copula is fitted to the standardized residuals. The student's t copula was selected for its ability to capture tail dependence, which is crucial for modeling co-movements during extreme market conditions (Demarta & McNeil, 2005).

Given the fitted marginal models, let equation 20 define the probability integral transformation of each standardized residuals into uniform variables, then the student-t copula can be defined by its correlation coefficient ρ and degrees of freedom v as:

$$C(u_1, u_2;\ \rho, v) = t_{v,\ p(t_v^{-1}(u_1), t_v^{-1}(u_2))} \quad (21)$$

Where, $C(u_1, u_2;\ \rho, v)$ is the copula function, $t_{v,\ p}$ is the CDF of the bivariate Student's t-distribution with v degrees of freedom and correlation ρ, t_v^{-1} is the quantile function of the univariate Student's t-distribution with v degrees of freedom [33]. These parameters are estimated using Maximum Likelihood Estimation via the fitCopula() function in R.

According to Demarta & McNeil [32], the density of t-copula for Maximum Likelihood Estimation, can be estimated as:

$$C_{v,P}^t = \frac{f_{v,P}\,(t_v^{-1}(u_1),\dots\dots,t_v^{-1}(u_d))}{\prod_{i=1}^d f_v\ (t_v^{-1}(u_i))},\ u \in (0,1)^d \quad (22)$$

Where, $f_{v,P}$ is the joint density of a multivariate t-distributed random vector and f_v is the density of the univariate standard t-distribution with v degrees of freedom.

To simulate the t-copula, firstly a multivariate t-distributed random vector X is generated, which means a random value that follows a t-distribution, with degrees of freedom v, a constant mean everywhere, and a correlation matrix P. A normal mixture technique explained by Demarta & McNeil [32] is used to generate the vector X. Then, the standard t-distribution's Cumulative Distribution Function (CDF) to each element of X is applied, so that the values fall between 0 and 1. This way, the required sample from the t-copula is obtained. Secondly, to estimate the density, the authors of this study first map the sample values back into the original t-distribution space, using inverse CDF, then they apply the multivariate t-distribution's density function at those points, and divided by the product of the marginal t-densities. Using the formula at equation 21, the density of the t-copula, which is useful for estimation is obtained. It is important to note that the t-copula remains invariant under strictly increasing transformations of the marginals, ensuring that the dependence structure is purely modelled by the copula, independent of the marginals themselves.

5.3 Long-Short Term Memory – Convolutional Neural Network (LSTM–CNN) Hybrid Model

Financial time series data often defy classical econometric assumptions, exhibiting nonlinearities, volatility clustering, regime shifts, and structural breaks. To navigate these complexities, hybrid deep learning model is employed that unites Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks [34]. CNNs excel at extracting short-term features such as bursts of volatility or abrupt spot–futures co-movements, while LSTMs are adept at modeling long-term dependencies through sophisticated gating mechanisms [35]. This synergy allows the model to capture both fleeting anomalies and persistent market behaviors, providing a dynamically adaptive, data-driven alternative to static or linear hedging strategies. The inclusion of CNN-LSTM hybrids in financial time series forecasting is increasingly common; numerous studies highlight that such models consistently outperform standalone CNN or LSTM architectures in predicting intricate patterns like cryptocurrency trends, stock movements, or foreign exchange rates [36].

The models used directly integrates hedging objectives into its architecture and loss function, enabling it to predict Minimum Variance Hedge Ratios (MVHR) that actively aim to minimize portfolio variance. This approach is especially relevant for equity futures markets, where optimal hedge ratios shift over time in response to evolving market regimes. The hybrid structure not only enhances predictive accuracy but aligns closely with the practical goal of real-world hedging effectiveness, making it an especially compelling solution for dynamic portfolio risk management.

5.3.1 Theoretical Foundations of the LSTM–CNN Hybrid Model

The LSTM–CNN hybrid exploits the complementary strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to handle the nonlinearities, volatility clustering, and regime shifts common in financial time series [37]. CNNs, though developed for spatial data, are highly effective in temporal contexts by applying learnable filters to sequential returns. For a one-dimensional series x_t and kernel w_k of size K, the convolution is:

$$(x * w)(t) = \sum_{k=0}^{K-1} x_{t+k} w_k \tag{23}$$

In hedging, this operation detects localized patterns like price jumps, volatility bursts, or short-lived spot–futures correlations, while causal padding preserves time order and avoids future leakage.

LSTMs extend the model's reach by capturing long-term dependencies through gated memory cells [38]. At time t, the cell state c_t and hidden state h_t evolve as:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{24}$$

$$h_t = o_t \odot tanh(c_t) \tag{25}$$

This gating selectively retains or discards information, enabling accurate modeling of persistent volatility and regime effects. Together, CNNs extract short-term anomalies and LSTMs capture structural trends, producing a unified, data-driven framework for robust, real-world dynamic hedge ratio estimation.

5.3.2 Model Architecture and Implementation

The LSTM–CNN hybrid pipeline, implemented in TensorFlow Keras, with design choices guided by both statistical diagnostics and domain expertise in financial time series. The dataset comprises daily spot and futures prices for the KSE-30 index, which are transformed into continuously compounded logarithmic returns:

$$R_{S,t} = \ln\left(\frac{P_{S,t}}{P_{S,t-1}}\right), \quad R_{F,t} = \ln\left(\frac{P_{F,t}}{P_{F,t-1}}\right) \tag{26}$$

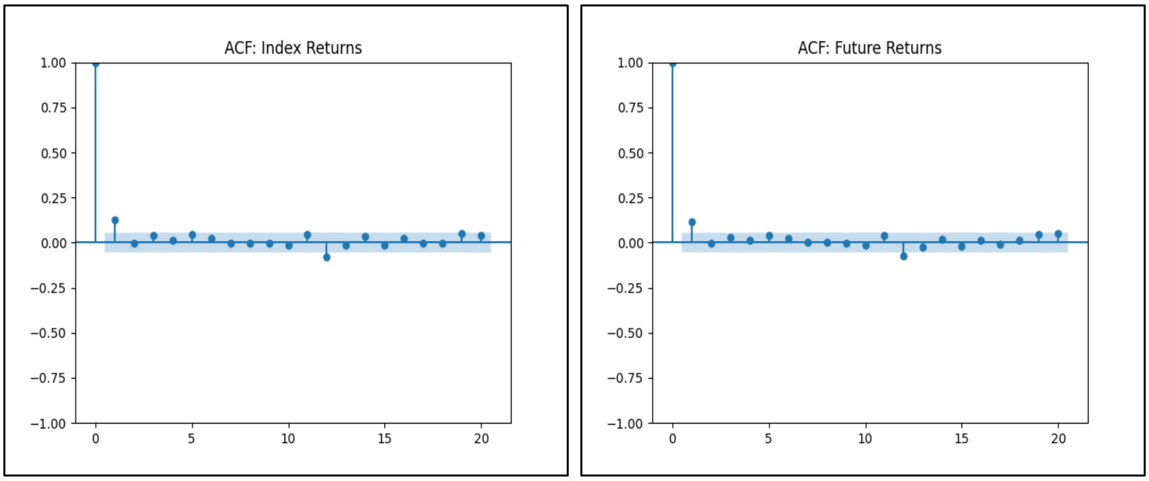

Figure 5. ACF of Index and Future Returns.

Autocorrelation analysis in figure 5, reveals a pronounced lag-1 spike for both series, indicating short-term dependence, followed by rapid decay to statistical insignificance. This justifies including both contemporaneous and one-period lagged returns as model inputs:

$$x_t = [R_{S,t}, R_{F,t}, R_{S,t-1}, R_{F,t-1}] \quad (27)$$

These four features are arranged into rolling sequences of length $p = 20$, forming tensors $X_t \in \mathbb{R}^{p\times 4}$ that balance sufficient market history with noise control. Targets are the next-day spot and futures returns.

The network begins with an input layer feeding a one-dimensional convolutional layer (32 filters, kernel size 3, causal padding) that extracts localized temporal patterns without future leakage. ReLU activation introduces nonlinearity while preserving computational efficiency. The CNN output is passed to an LSTM layer with 32 units, which learns long-term dependencies and compresses the sequence into a fixed-length embedding. A fully connected dense layer maps this embedding to the predicted hedge ratio $\hat{\beta}_t$, which is directly integrated into the hedging objective through a custom Lambda layer:

$$\hat{R}_{hedged,t} = R_{S,t} - \hat{\beta}_t R_{F,t} \quad (28)$$

This end-to-end design ensures the model learns hedge ratios that minimize the variance of the hedged portfolio, aligning statistical performance with practical risk-management objectives. The model architecture pipeline can be seen in figure 6 below.

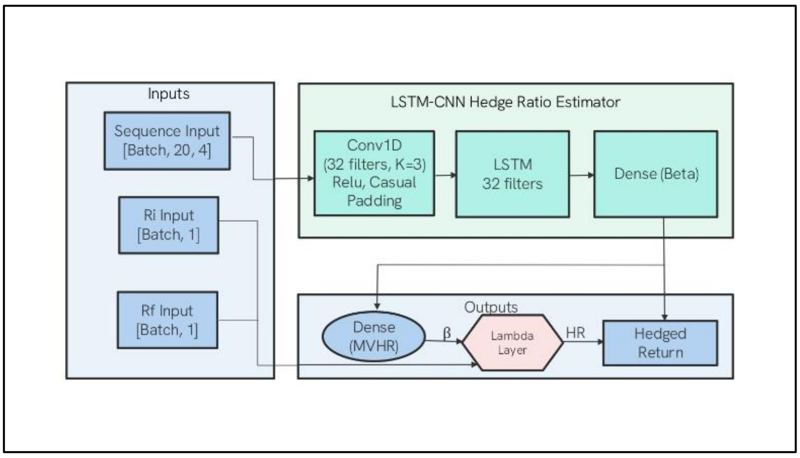

Figure 6. Architecture and Pipeline of LSTM-CNN Hybrid Model.

5.3.3 Model Compilation, Training, and Evaluation Framework

The CNN–LSTM model was compiled in TensorFlow–Keras using the Adam optimizer (learning rate 10^{-3}), chosen for its stability and rapid convergence in training deep architectures on noisy financial series [40]. The learning objective directly targets portfolio risk minimization by reducing the mean squared error between the hedged return and zero, embodying the ideal of a perfectly neutralized position:

$$\min_{\theta} \frac{1}{n}\sum_{i=1}^{n}(R_{S,t} - \hat{\beta}_t R_{F,t})^2 \quad (29)$$

where θ denotes all trainable parameters. Chronological integrity of the data was preserved with an 80/20 train–test split [40]. Training ran for 50 epochs with a batch size of 64, incorporating early stopping and adaptive learning rate scheduling to safeguard against overfitting.

The input tensor $X_t \in \mathbb{R}^{p\times 4}$ encapsulates 20 days of contemporaneous and lagged spot–futures returns. These pass through a causal 1D convolution layer with 32 filters (kernel size $K = 3$) that extract short-term temporal features:

$$h_c^{(t)} = f\left(\sum_{k=0}^{K-1} w_c^{(k)} \cdot x_{t-k} + b^{(c)}\right) \quad (30)$$

The convolutional output is then processed by an LSTM with 32 hidden units to capture nonlinear long-range dependencies. The final hidden state h_p is mapped to the hedge ratio $\hat{\beta}_t = W_\beta h_p + b_\beta$, which in turn yields the residual hedged return $\hat{R}_{hedged,t} = R_{S,t} - \hat{\beta}_t R_{F,t}$.

Performance evaluation on the test set applied the KPSS and Ljung–Box tests to confirm stationarity and absence of autocorrelation, alongside ARCH LM tests to check for volatility clustering. Hedging effectiveness was quantified via variance reduction, RMSE, Sharpe ratio, directional accuracy, mean absolute deviation, and tail risk metrics such as VaR and CVaR under extreme value theory. This architecture's design—causal convolutions for temporal integrity, LSTM memory for complex dependencies, and a loss function aligned with economic objectives which ensures both statistical rigor and financial relevance, marking it as a robust tool for risk mitigation in dynamic markets.

5.4 Fourier Transform Network (FT-Net) Hybrid Model

Contemporary financial markets are shaped by an interplay of irregular shocks, cyclical forces, and persistent non-stationarities, producing dynamics that cannot be fully resolved by purely time-domain models. Classical econometric frameworks and deep learning architectures such as LSTMs and CNNs capture temporal dependencies well yet remain inherently blind to spectral structures like dominant frequencies, seasonal patterns, business cycles, volatility regimes, that often drive return dynamics. The omission of such frequency-domain information can lead to systematically biased hedge ratios and overlooked arbitrage windows.

The FT-Net architecture addresses this gap by integrating explicit Fourier Transform–based modules with advanced temporal modeling blocks, ranging from CNNs for localized feature extraction to recurrent layers such as LSTM or GRU for sequential dependency capture. Through jointly operating in the frequency and time domains, FT-Net can decompose price series into their spectral components while tracking evolving temporal interactions [41]. It enables the model to detect cyclical risks and regime shifts in equity futures before they materialize in the raw time series.

This dual-domain approach produces hedge ratios and arbitrage signals that are both spectrally aware and temporally responsive, surpassing the limitations of conventional deep learning and econometric models. In high-frequency equity futures markets, where cycles may span from intraday periodicities to multi-month business regimes, FT-Net's capacity to jointly learn across domains offers a path toward more robust dynamic hedging and statistically significant arbitrage detection. The result is a framework aligned with modern trading demands such as anticipatory, adaptive, and grounded in both the structural and stochastic realities of market behavior.

5.4.1 Theoretical Foundations of the FT-Net Hybrid Model

The FT-Net architecture is grounded in spectral analysis, leveraging the Discrete Fourier Transform (DFT) to uncover hidden periodicities and regime dynamics in financial time series. As mentioned by Brigola [9], the DFT decomposes a sequence of returns $\{x_n\}_{n=0}^{N-1}$ into sinusoidal components of varying frequencies, amplitudes, and phases:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-2\pi ikn/N}, k = 0,1,2 \dots \mathrm{N} - 1 \qquad (31)$$

Here, X_k is the complex coefficient for the k-th frequency bin, $|x_k|$represents the strength of that frequency, and $\arg(x_k)$ encodes its phase. In market data, these spectral signatures often correspond to cyclical forces, macroeconomic expansions and contractions, seasonal trading patterns, or volatility regimes, that may remain obscured in purely time-domain analysis. Crucially, spectral shifts can precede observable changes in the time series, offering early indicators of structural market transitions [42].

Empirical research has repeatedly shown that financial returns exhibit distinct frequency-domain features, from multi-year business cycles to intraday periodicities, and that volatility clustering and contagion often leave identifiable spectral fingerprints [43]. Yet, most deep learning applications in finance remain confined to the time domain, failing to exploit this rich structure. Through embedding DFT-based modules directly into its architecture, FT-Net fuses temporal dependency modeling with spectral decomposition, enabling simultaneous learning of time–frequency interactions [44]. This joint-domain approach equips the model to detect emerging risks and arbitrage opportunities with greater timeliness and precision than either domain alone.

5.4.2 Model Architecture and Implementation

The FT-Net Hybrid is a modular deep learning architecture that fuses spectral analysis with temporal modeling to address the dual-domain complexity of financial time series [45]. Its design rests on the premise that asset returns carry intertwined signatures in both time and frequency domains, and only their joint modeling can fully reveal cyclical structures, transient shocks, and regime shifts critical for hedging decisions. The pipeline comprises four principal stages: Fourier-based spectral decomposition, temporal convolution, feature fusion, and sequence modeling, culminating in the estimation of dynamic hedge ratios.

Given a rolling input window $x_t \in \mathbb{R}^{w \times d}$of past returns, lags, and optional statistical features, the Fourier branch transforms each feature channel $x_{t,j}$ to the frequency domain via the Discrete Fourier Transform:

$$F_k^{(j)} = \sum_{n=0}^{w-1} x_{t-w+1+n,j} \cdot e^{-\frac{2\pi ikn}{w}}, \qquad k = 0,1,2, \dots, w-1 \qquad (32)$$

Real and imaginary components (or magnitude and phase) are concatenated to form the spectral vector $S_t \in \mathbb{R}^{2wd}$, with optional attention weighting α_k to emphasize dominant predictive frequencies. Parallel to this, the temporal branch applies one-dimensional convolutions:

$$C_{t,s}^{(m)} = \sum_{l=0}^{f-1} \sum_{j=1}^{d} W_{l,j,s}^{(m)} \cdot x_{t-w+1+l,j} + b_s^{(m)} \qquad (33)$$

Followed by non-linear activations, enabling the detection of localized events such as volatility bursts or short-lived arbitrage opportunities.

Outputs from both branches are fused into a joint feature vector $Z_t = [Flatten(C_t), S_t]^\top$, which is then processed by a recurrent layer, called LSTM , to model long-range dependencies and adapt to non-stationary market regimes. The final dense readout produces hedge ratios under appropriate activation constraints (sigmoid for [0,1], tanh for bounded leverage, or linear for unrestricted cases). This architecture's multi-resolution representation and end-to-end trainability make FT-Net a powerful framework for exploiting the full informational richness of index–futures dynamics. The model architecture pipeline can be seen in figure 7 below.

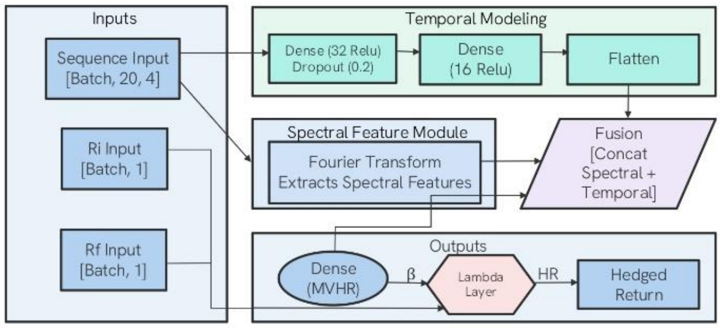

Figure 7. Architecture and Pipeline of FT-Net Hybrid Model.

5.4.3 Model Compilation, Training, and Evaluation Framework

The FT-Net Hybrid produces the one-period-ahead dynamic hedge ratio β_t directly from its final hidden state h_t through a fully connected mapping,

$$\beta_t = \phi(w_{out}^{\top} h_t + b_{out}) \tag{34}$$

where $\phi(\cdot)$ enforces the desired range of leverage or exposure. The predicted hedged return is then constructed in the manner of a minimum-variance portfolio, adjusting exposure each period in response to evolving market conditions. Training is formulated as the minimization of the variance of the hedged return series, operationalized as a mean squared error against zero,

$$\mathcal{L}_{var} = \frac{1}{N}\sum_{i=1}^{n}(\hat{R}_{hedged,i})^2 \tag{35}$$

with regularization via L2 weight decay, dropout, and batch normalization to improve generalization. Optimization employs adaptive stochastic gradient descent (Adam), enabling efficient parameter tuning across the model's spectral and temporal pathways.

Model validation spans both statistical and financial diagnostics to ensure risk minimization and exploitation of arbitrage opportunities. Stationarity is evaluated via the KPSS test, while the Ljung–Box and ARCH–LM tests assess residual autocorrelation and volatility clustering. Performance is quantified through MAE, RMSE, variance reduction, Theil's U, and Sharpe ratio, complemented by directional accuracy metrics. Tail risk is probed using extreme value theory, computing VaR and CVaR reductions relative to benchmarks. Visual diagnostics—time-varying β_t plots, cumulative return curves—reveal how the model adjusts under volatility bursts, regime shifts, and seasonal cycles, offering transparency into its adaptive mechanics.

5.5 Post Model Performance Evaluation

5.5.1 Statistical Tests Post-Model Implementation

Post-implementation diagnostics were conducted to validate model reliability, assess residual behavior, and benchmark hedging quality. Theil U statistics measured forecasting performance against a naïve 1:1 hedge, while ARCH-LM, Ljung-Box, and KPSS tests on training and testing residuals ensured the absence of volatility clustering, serial correlation, and non-stationarity. Econometric models used log-differenced spot and futures returns of the KSE-30, while deep learning models employed hedged return series derived from dynamically estimated hedge ratios, enabling adaptive risk management without overfitting.

5.5.2 ARCH LM Test

The ARCH-LM test, initially applied pre-model to confirm conditional heteroscedasticity, was re-applied post-model to assess volatility absorption. DCC-GARCH p-values (training: 0.0645/0.0586; testing: 0.1427/0.1687) indicated only marginal heteroscedasticity, with Copula DCC-GARCH showing p-values > 0.1 throughout—evidence of effective tail dependence modeling. Deep learning models exhibited p-values ≈ 1.000, suggesting complete removal of ARCH effects and validating their suitability for hedging comparison.

5.5.3 Ljung Box Q Test

The Ljung-Box test confirmed that all models produced uncorrelated residuals, with p-values > 0.1 for econometric models and > 0.5 for deep learning models, indicating no systematic temporal dependencies. This reinforces correct lag specification and absence of residual autocorrelation, even in the presence of non-linear financial time series structures.

5.5.4 KPSS Test

Stationarity, critical for valid volatility estimation, was confirmed across all models, with p-values > 0.1 for both training and testing sets. This stability ensures that volatility and hedge ratio estimates remain reliable and mean-reverting, aligning with recent findings that deep learning can capture non-linear stationarity structures that traditional models also accommodate.

5.5.5 Theil U Statistic

Against the naïve hedge, Copula DCC-GARCH (U = 0.8528) outperformed DCC-GARCH (U = 0.91), demonstrating superior forecast accuracy. LSTM-CNN (0.9179) and FT-Net (0.9128) trailed slightly yet still beat the naïve benchmark (U < 1), affirming their viability in practical hedging applications.

5.6 Predicted Minimum Variance Hedge Ratios (MVHRs)

The MVHRs, estimated by Copula DCC-GARCH and DCC-

GARCH models, shown in figure 8 and 9, highlight the in-sample period in light blue series and out-of-sample period in orange series. The MVHRs in the training sample for both models fluctuate approximately around 0.8 and 1.12, indicating stable hedge ratios (β) for the 2019-22 region. However, despite the EWMA filter to smooth out the noise and stabilize volatility, the hedge ratios in both models show spike of approximately 1.2 and 1.1 followed by a dip of approximately 0.75 and 0.8 due to the onset of the COVID-19 pandemic, reflecting unpredictable change in covariance's. Proceeding to the testing period, increased fluctuations are observed with a dip of 0.75 and peak of approximately 1.30 for both models in the 2023-24 region. This hedging pressure reflects the impact shown by the DCC-GARCH and Copula DCC-GARCH model to previous stress episodes. While comparing the Copula DCC-GARCH and DCC-GARCH models, it can be observed that Copula DCC-GARCH demonstrates short lived adjustments due to a slightly lower average MVHRs, combined with amplified oscillations in hedge ratios for the period of extreme shocks reflecting t-copulas ability to capture tail dependence and asymmetric co-movements, while DCC-GARCH has marginally higher average hedge ratios.

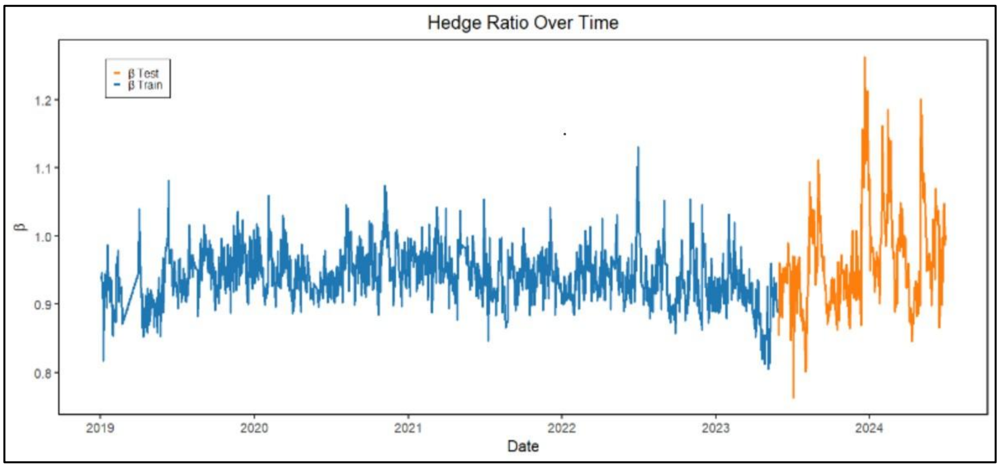

Figure 8. Dynamic MVHRs obtained using DCC GARCH

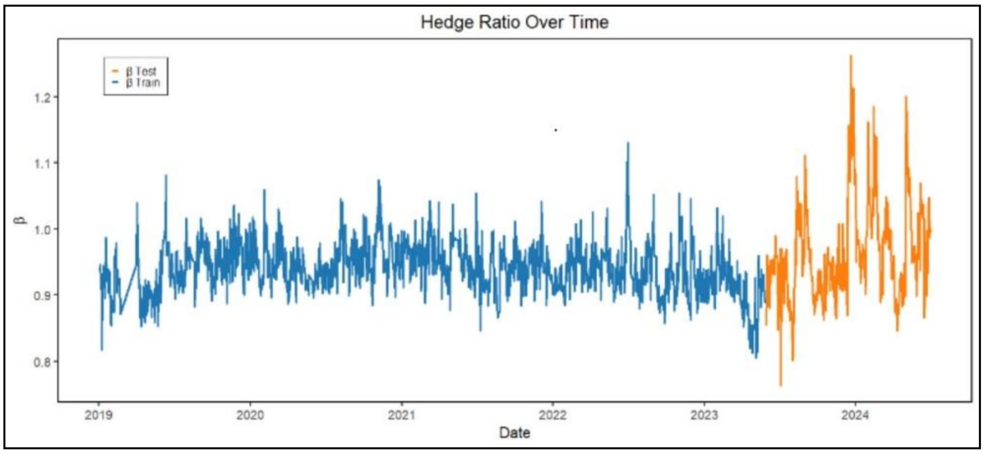

Figure 8. Dynamic MVHRs obtained using Copula DCC GARCH.

Moreover, the MVHRs estimated by LSTM-CNN and FT-Net hybrid models, shown in figure 10 and 11, highlight the in-sample period in dark blue series and out-of-sample period in orange series. The hedge ratios (β) in LSTM-CNN model with its peak of 1.05 in the training period has a subdued response to extreme events, when compared to FT-Net, reflecting the model's ability to avoid overstating and understating hedge ratios, suitable for volatile` Pakistani market. The convolutional layer and temporal dependency of the LSTM layer of the LSTM-CNN model fluctuates around 0.94-0.98 range, showing reduced short noise yet lower and non-reactive hedging performance when compared to FT-Net. On the other hand, FT-Net shows hedge ratios (β) mainly close to 1, suggesting better hedge performance. In both deep learning models, a sudden spike i.e. 1.10 in FT-Net and 1.05 in LSTM-CNN, is observed in the early 2020, due to the onset of COVID-19 pandemic, reflecting uncertainty similar to the mathematical model. However, FT-Net highlighting its mean-reverting behavior stabilizes the hedge ratios quickly compared to the other models. The testing period fluctuating around 0.96 and 0.99 shows reliable out of sample performance and the model's ability to quickly capture non-linear complexities. Comparing the hedge ratios of mathematical and deep learning models, it is observed that although all the models exhibit mean reverting behavior, yet FT-Net has higher suitability in dynamic conditions, as it shows greater responsiveness to the volatile Pakistani market.

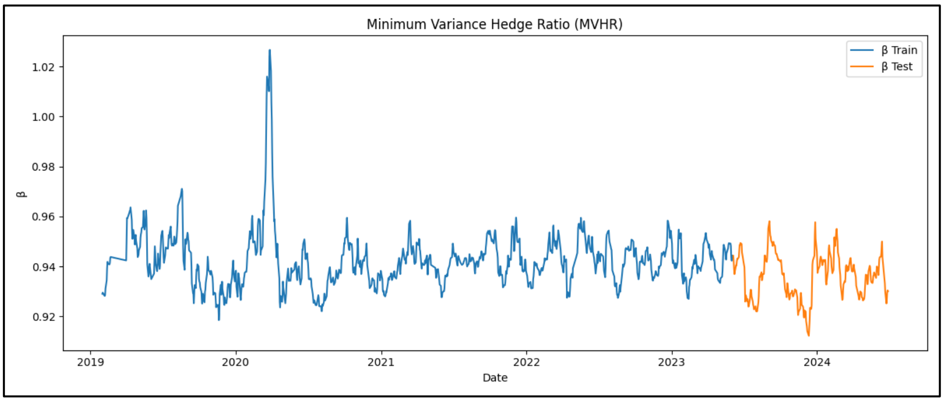

Figure 9. Dynamic MVHRs obtained using LSTM-CNN

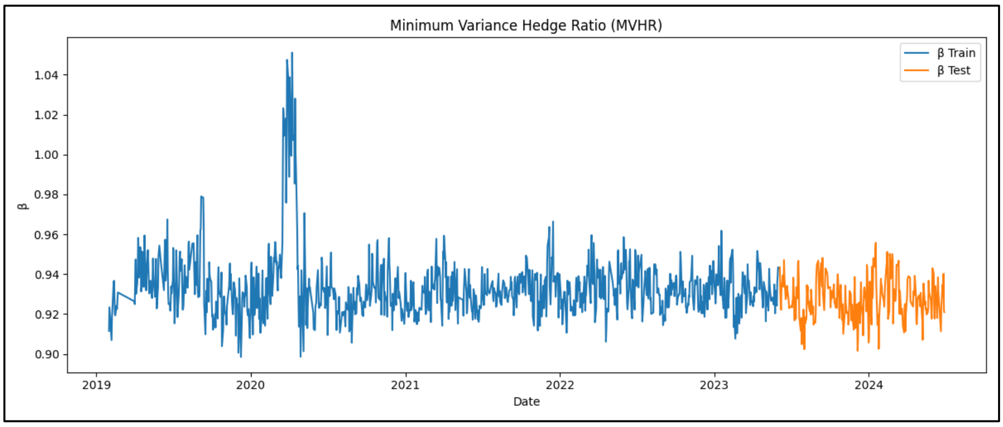

Figure 10. Dynamic MVHRs obtained using FT-Net Hybrid

5.7 Hedged Returns Estimation

After the hedging ratios in the training and testing data, the hedged returns are computed for both DCC-GARCH and Copula DCC-GARCH. For qualitative inspection, the time series of predicted hedge ratios $\hat{\beta}_t$ is plotted on training and test dates, as well as the cumulative unhedged versus filtered hedged returns. Cumulative hedged and unhedged returns are calculated using formulae below.

$$CumUnhedged(t) = \sum R_{S,i} \ , \quad (36)$$

$$CumHedged(t) = \sum_{i \le t} R_{hedged,i}^{final}$$

These figures 12 and 13 below vividly illustrate the risk-reduction benefits of the dynamic hedging procedure used in this research paper. For DCC-GARCH, the hedged returns (in orange line) show smoother returns as compared to the unhedged returns, which show clear fluctuations especially during the COVID-19 pandemic period. Whereas Copula DCC-GARCH depicts slightly higher returns as compared to DCC-GARCH, due to effective hedging and risk reduction.

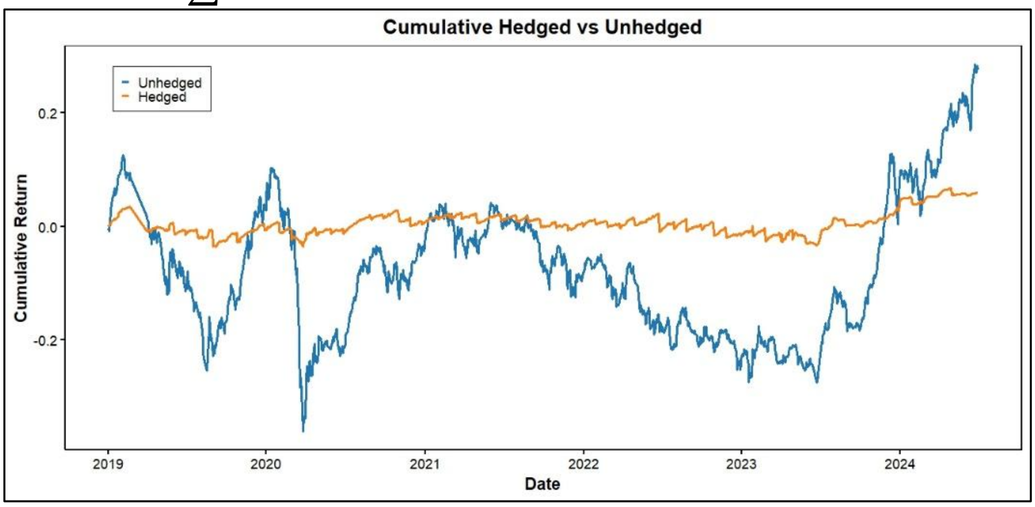

Figure 11. Cumulative Hedged and Unhedged Returns – DCC GARCH.

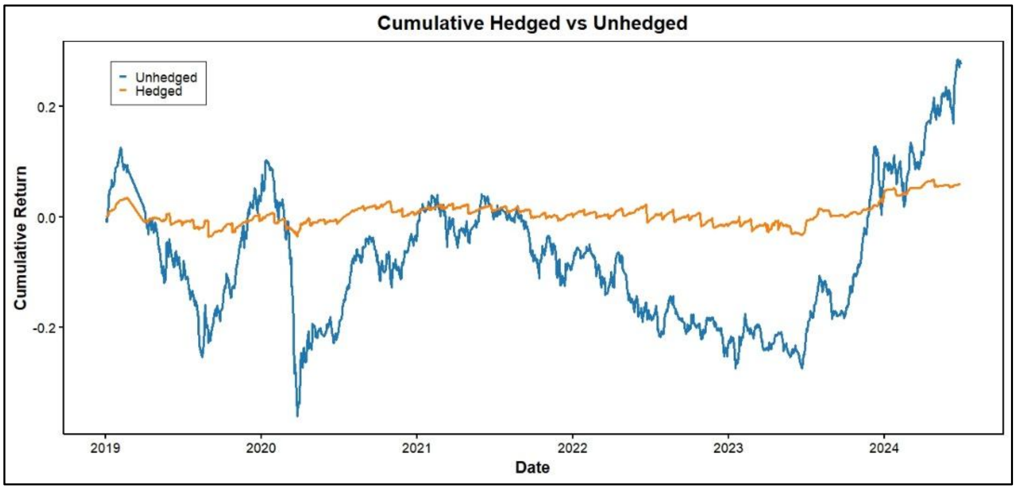

Figure 12. Cumulative Hedged and Unhedged Returns – Copula DCC GARCH.

In the LSTM–CNN model, $\hat{\beta}_t$ was estimated for both training and test windows, and hedged returns were computed. The resulting $\hat{\beta}_t$ series from both sets was concatenated to produce a full-period, unfiltered hedged return series. To address systematic biases, residuals $e_t = \hat{R}_{hedged,t}$ were regressed against their recent lags and contemporaneous spot/futures returns. After removing initial NaNs, the data was split chronologically and used to train three gradient boosting regressors—XGBoost, LightGBM, and CatBoost—each with 200 trees. The ensemble average of their predictions yielded e_t, which was subtracted from the original hedged returns, producing a variance-reduced, filtered series.

For the FT-Net model, $\hat{\beta}_t$ was similarly generated for both train and test sets, but without explicit variance–covariance computation, as the network's spectral–temporal architecture implicitly learned the mapping to optimal hedge ratios. Residual correction followed the same ensemble boosting procedure as in LSTM–CNN, producing a "double-filtered" hedged return series with even lower volatility.

Cumulative return analysis highlights the practical edge of FT-Net over LSTM–CNN. Both models maintained exceptionally flat, low-volatility hedged profiles through mid-2023, but FT-Net achieved a very slightly higher cumulative hedged return peak and exhibited slower drawdown decay. This suggests that FT-Net's frequency–temporal layers adapt marginally faster to trending conditions, capturing gains more decisively, whereas LSTM–CNN's convolution–recurrent design delivers comparable smoothness but slightly lower total hedge payoff.

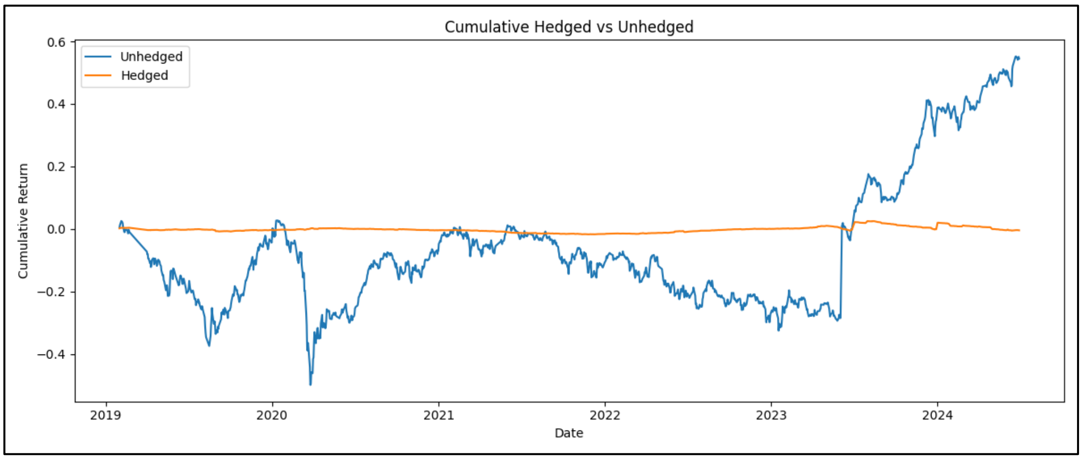

Figure 13. Cumulative Hedged and Unhedged Returns – LSTM-CNN.

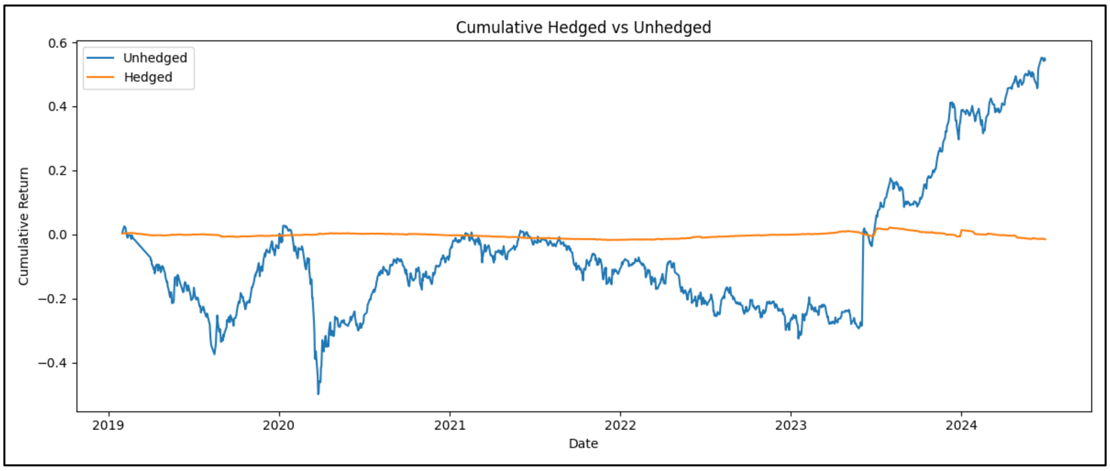

Figure 14. Cumulative Hedged and Unhedged Returns – FT-NET Hybrid.

6. COMPARATIVE ANALYSIS USING FUZZY TOPSIS AND OTHER DECISION-MAKING CRITERIA

Hedging strategies for a KSE 30 index portfolio is formulated using the KSE 30 futures contract through both sophisticated econometric models, and cutting-edge deep learning architectures, to dynamically minimize risk in the equity futures market. The primary objective of this chapter is to rigorously assess the hedging effectiveness of the model by using a comprehensive suite of quantitative performance metrics and rank them with respect to their hedging performance. The authors of this study have employed nine performance criteria to gauge the hedge effectiveness, which include the variance reduction percentages, Root Mean Square Errors (RMSE), average hedged returns, Sharpe ratios, directional accuracy, Value at Risk (VaR) Reduction, Conditional Value at Risk (CVaR) Reduction, mean absolute deviation, and time complexity. These metrics effectively capture the risk minimization capability and the operational efficiency of the models. However, the challenges arise from the multicriteria analysis, as no single model unequivocally dominates all criteria, emphasizing the conflicts in the Multi-Criteria Decision-Making (MCDM) context. This predicament created the need for the adoption of a systematic methodology to assess the performance of each model and rank them accordingly.

To address this problem, the authors of this study incorporated the Fuzzy Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method, a multicriteria decision-making technique [46]. The Fuzzy TOPSIS method ranks multiple alternatives based on their closeness to an ideal solution while maximizing benefits and minimizing costs. The following section provides an in-depth discussion of the Fuzzy TOPSIS methodology used in this research.

6.1 Implementation of the Fuzzy TOPSIS methodology

6.1.1 Define the decision-making problem

The first step of the TOPSIS methodology is to define the decision-making problem. To obtain fair and multidimensional assessment results, the authors of this study utilize nine key performance metrics that gauge the risk minimization capability, predictive accuracy, profitability, and operational efficiency of the models.

i. Variance reduction helps us assess the extent to which a model reduces the variance of the hedged portfolio compared to the variance of the unhedged portfolio.
ii. The RMSE evaluates the model's ability to forecast the hedge ratios; lower RMSEs indicate more precise predictions.
iii. The average hedged return is an indicator of the profitability of the models, as it reflects the model's ability to maintain the portfolio returns while minimizing the risk.
iv. Sharpe ratio is a measure of risk-adjusted return as it reflects the ability of a model to achieve higher returns given the risk minimization criteria.
v. Directional accuracy is also an indicator of the predictive accuracy of the models; however, it focuses on the frequency with which the model correctly predicts the direction of market movements.
vi. Value at Risk (VaR) reduction is a measure of how effectively a model reduces potential losses at a specified confidence level.
vii. Conditional Value at Risk (CVaR) reduction further extends the analysis of downside risk by assessing the expected average loss beyond the VaR threshold.
viii. Mean absolute deviation (MAD) is an additional measure of predictive accuracy.
ix. Finally, time complexity measures the computational resources and time required for each model to execute. It is an essential metric to gauge the operational efficiency of the models.

Tables 1 and 2 below show the criteria for all models on the training and testing datasets.

Table 1. ***Performance Metrics of Hedging Models – Training.***

METRICS	COPULA GARCH	DCC GARCH	LSTM-CNN HYBRID	FT-NET HYBRID
Variance reduction	92.54	92.71	99.9	99.9
RMSE	0.01241	0.01234	0.00041	0.00042
Average Return	-0.000024	-0.000046	0.000004	0.000004
Sharpe Ratio	-0.0067	-0.013	0.0098	0.0102
Directional Accuracy	66.63	70.09	50.81	52.64
VaR Reduction	83.58	83.97	97.66	97.55
CVaR Reduction	86.63	87.02	97.31	97.23
MAD	0.00886	0.0088	0.000239	0.000238
Time Complexity (s)	6.66	4.48	10.79	9.63

Table 2. Performance Metrics of Hedging Models – Testing

METRIC	COPULA GARCH	DCC GARCH	LSTM-CNN HYBRID	FT-NET HYBRID
Variance Reduction	95.09	95.05	96.96	96.56
RMSE	0.01118	0.01121	0.00206	0.00219
Average Return	0.00029	0.00034	-0.00009	-0.00005
Sharpe Ratio	0.1113	0.129	-0.0438	-0.0227
Directional Accuracy	62.07	67.82	51.34	51.34
VaR Reduction	89.39	88.54	94.65	94.38
CVaR Reduction	90.97	90.1	91.74	91.62
MAD	0.00824	0.00817	0.000611	0.000619
Time Complexity (s)	4.48	6.39	10.79	9.63

6.1.2 Determination of AHP weights

After identifying all decision criteria, the fuzzy TOPSIS multicriteria decision-making methodology requires weights to be assigned to each of them. Therefore, to assign weights to every criterion, the authors of this study use the Analytic Hierarchy Process (AHP), which converts a complex problem into a hierarchy of sub-problems [47]. They begin by constructing a pairwise comparison matrix, where each element represents the relative importance of one metric over the other. A 9x9 pairwise comparison matrix is created (Table 3), where each entry (i, j) in this matrix expresses how much more important criterion i is compared to criterion j. In the table below, each metric is represented as follows: variance reduction (C1), RMSE (C2), average hedged return (C3), sharpe ratio (C4), directional accuracy (C5), VaR reduction (C6), CvaR reduction (C7), mean absolute deviation (C8), time complexity (C9).

Table 3. Computation and Fuzzification of AHP Pairwise Comparison Matrix.

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	1	5	5	5	7	7	7	9	9
C2	1/5	1	3	3	5	5	5	7	7
C3	1/5	1/3	1	2	4	4	4	6	6
C4	1/5	1/3	1/2	1	3	3	3	5	5
C5	1/7	1/5	1/4	1/3	1	2	2	4	4
C6	1/7	1/5	1/4	1/3	1/2	1	2	3	3
C7	1/7	1/5	1/4	1/3	1/2	1/2	1	3	3
C8	1/9	1/7	1/6	1/5	1/4	1/3	1/3	1	2
C9	1/9	1/7	1/6	1/5	1/4	1/3	1/3	1/2	1

Once the initial pairwise matrix is constructed, it is normalized to ensure consistency and to allow for proper weighting. The matrix is normalized by dividing each entry in a given column by the sum of that column, transforming the original elements into relative proportions. The normalized matrix reflects the proportional importance of each criterion relative to every other criterion in the set.

Then the weight of each criterion is extracted by computing the mean value of each row in the normalized matrix. This average quantifies the overall relative importance of each criterion, aggregating its normalized influence across all pairwise comparisons. The vector of these row means represents the initial set of crisp (precise) weights, which together sum to one, ensuring a proper probability distribution over the criteria.

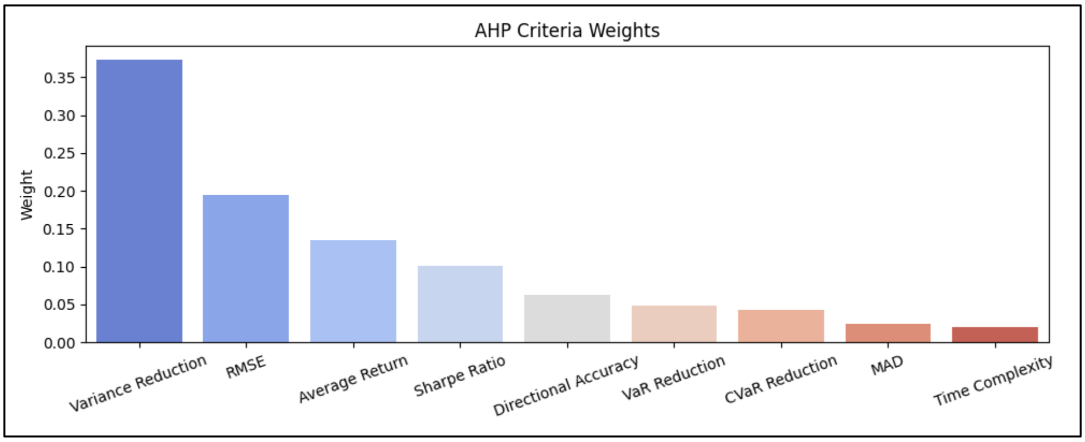

Figure 15. Fuzzified AHP Criteria Weights.

Instead of considering each crisp weight, they are converted to fuzzified values by constructing a Triangular Fuzzy Number (TFN) centered on the crisp value but allowing for a range of uncertainty. Specifically, for each weight w, a TFN is defined as (0.9w, w, 1.1w), thereby incorporating a ±10% spread around the nominal value. This fuzzification step captures the ambiguity and lack of perfect precision, making the criteria weighting process more realistic and defensible.

The fuzzified AHP weights are presented in the figure 16 above from highest to lowest. The highest weights were assigned to variance reduction and RMSE, while the lowest weights were assigned to the mean absolute deviation and time complexity.

6.1.3 *Fuzzification of decision criteria*

After defining and calculating all criteria and their fuzzified weights, the next step was to convert the criteria to fuzzy numbers, using Triangular Fuzzy Number (TFN), denoted by three points: lower bound l, modal value m, and upper bound u [48]. If a criterion has an observed value v, fuzzification is defined as:

$$(l, m, u) = (v \times (1 - \delta), v, v \times (1 + \delta)) \quad (37)$$

where δ is the fuzziness factor capturing measurement ambiguity. Here, l reflects a pessimistic scenario, m the most likely estimate, and u an optimistic evaluation. By constructing the decision matrix with TFNs, the analysis formally incorporates uncertainty into subsequent steps.

Next, the fuzzy decision matrix is normalized to eliminate scale effects. For benefit-type criteria (e.g., Sharpe ratio, variance reduction), normalization ensures that larger values map closer to unity, while for cost-type criteria (e.g., RMSE, MAD, time complexity), values are inverted such that smaller values are preferable. This guarantees comparability across heterogeneous metrics.

Following normalization, the weighted normalized fuzzy matrix is constructed. Each normalized TFN is multiplied by its fuzzified weight, ensuring that more critical criteria exert proportionally stronger influence. The weighted decision matrix thus integrates both the intrinsic fuzziness of data and the subjective prioritization of criteria.

The next step identifies Positive Ideal Solutions (PIS) and Negative Ideal Solutions (NIS). For each criterion, the PIS represents the most favorable TFN (highest benefit or lowest cost), while the NIS captures the least favorable scenario [49]. Each model's desirability is then determined by computing its Euclidean distance to the PIS and NIS. For TFNs, the distance between two fuzzy numbers (l_1, m_1, u_1) and (l_2, m_2, u_2) is expressed as:

$$d = \sqrt{\frac{1}{3}[(l_1 - l_2)^2 + (m - m_2)^2 + (u_1 - u_2)^2]} \quad (38)$$

Finally, the closeness coefficient (CC) is derived as:

$$CC_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (39)$$

where D_i^+ is the distance of model i from the PIS and D_i^- its distance from the NIS. The closeness coefficient, bounded between 0 and 1, serves as a scalar indicator of how closely each model approximates the ideal benchmark [50].

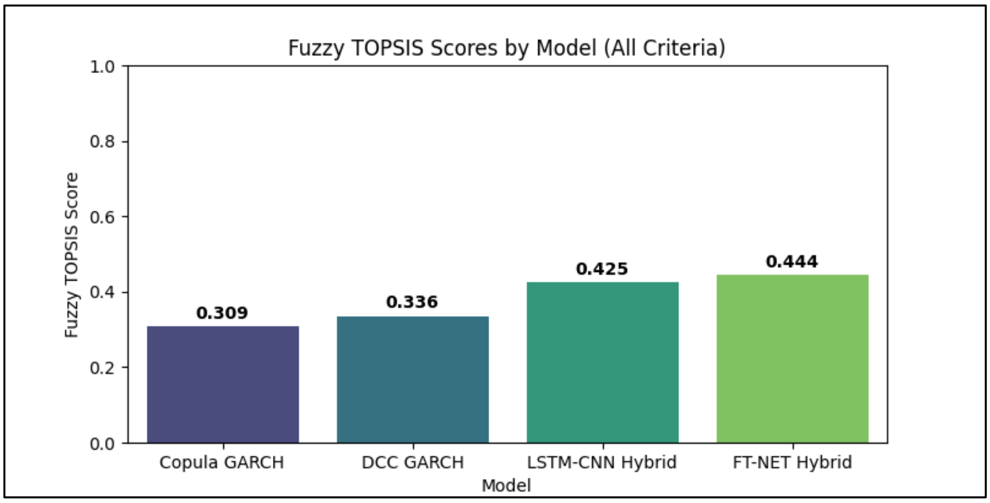

Figure 16. Model Ranking by Fuzzy TOPSIS.

Figure 17 presents the resulting ranking. The FT-Net Hybrid model achieves the highest coefficient (0.444), narrowly outperforming the LSTM–CNN Hybrid (0.425). In contrast, econometric benchmarks lag significantly, with DCC-GARCH (0.336) and Copula GARCH (0.309). These ranking underscores two insights: (i) deep learning hybrids better capture nonlinear dynamics and structural breaks in financial time series, and (ii) integrating temporal and frequency-domain features (as in FT-Net) provides a marginal but meaningful edge over purely sequential architectures.

The results demonstrate that fuzzy TOPSIS offers a transparent, multidimensional, and uncertainty-aware evaluation of hedging performance. Through joint consideration of risk minimization, predictive accuracy, profitability, tail risk management, and computational cost, while embedding fuzziness into every step, the methodology ensures robustness to both statistical noise and subjective judgment.

6.2 The Composite Efficiency Index (CEI)

This paper further extends the comparative analysis by incorporating the Composite Efficiency Index (CEI), a metric historically employed in efficiency-based assessments [51]. Building upon the fuzzy TOPSIS methodology, the CEI is derived from the closeness coefficients, providing a complementary perspective on model performance. While the fuzzy TOPSIS coefficient measures each model's absolute proximity to the ideal solution, the CEI translates these values into relative efficiency shares, thereby revealing the proportional contribution of each model to the overall decision-making efficiency. By normalizing performance across all models, the CEI not only enhances interpretability but also facilitates a more intuitive and quantitative comparison of multicriteria outcomes.

6.2.1 Implementation

To synthesize the multi-criteria evaluation results into a single interpretable metric, the CEI was constructed based on the fuzzy TOPSIS closeness coefficients. The CEI quantifies the relative efficiency of each model as a normalized share of the total system performance, ensuring both comparability and scale independence. Formally, the CEI for model $i$is defined in equation 40 as:

$$CEI_i = \frac{CC_i}{\sum_{j=1}^{n} CC_j} \tag{40}$$

Where CC_i denotes the fuzzy TOPSIS closeness coefficient of model i, and n represents the total number of candidate models.

The resulting CEI values (figure 18) indicate that the *FT-Net Hybrid* model achieves the highest composite efficiency (0.293), followed closely by the *LSTM–CNN Hybrid* (0.281). Together, these deep learning hybrid models account for approximately 57% of the total efficiency, significantly outperforming the econometric baselines: *DCC-GARCH* (0.222) and *Copula-GARCH* (0.204). These findings confirm that hybrid architectures, which integrate temporal and frequency-domain representations, deliver more robust and adaptive hedging performance in volatile equity markets.

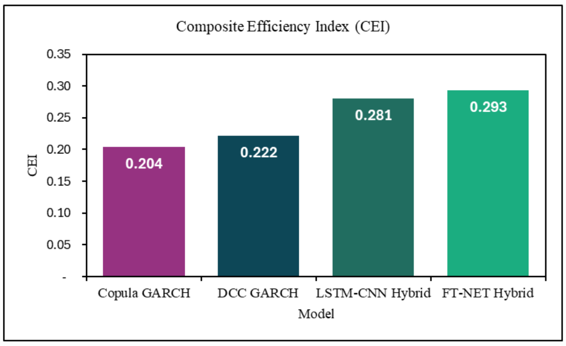

Figure 18. Model Ranking by CEI.

7. STATISTICAL ARBITRAGE ANALYSIS

Statistical arbitrage refers to a class of trading strategies that use statistical, mathematical and computational methods to exploit pricing inefficiencies between financial instruments [52]. Statistical arbitrage is widely understood as a high-volume, short-term trading strategy. Unlike traditional arbitrage, which exploits mispricing between identical or related securities for risk-free profits, statistical arbitrage distinguishes itself by assuming such mispricing is subtle, short-lived, and revealed only by complex analysis. Statistical arbitrage involves taking advantage of mutual relationships, for example, mean reversion, cointegration or other statistical properties. These relationships must last long enough to be traded systematically, regardless of the noise of the market or frictions [53] [54].

Statistical arbitrage is based on the idea that the prices of related financial assets are not always perfectly in sync or adjust instantaneously, thus giving rise to a temporary deviation from their equilibrium relation. These deviations arise due to various reasons like liquidity shock, investor reaction and information lag and so on. They open up arbitrage opportunities if they can be detected and acted on before prices revert to normal co-movement. It's worthwhile noting that signals based on statistical patterns can affect prices themselves. A similar point can be made regarding time-series based strategies. Statistical arbitrageurs use quantitative techniques for creating trading signals, managing risk, and executing trades either at high frequency or in a disciplined, repeatable manner [55]. They rely on the law of large numbers, and diversification across many trades to secure consistent returns.

In the context of equity futures market, the use of statistical arbitrage strategies has become relevant due to the high liquidity, transparency and standardization of contracts making systematic trading possible. Equity futures like the ones on broad-based indices exhibit strong statistical relationships with their underlying cash indices as they are arbitraged to link the spot and futures price. Nevertheless, due to market microstructure effects, varying levels of liquidity, rolling contract dynamics, and short-term supply-demand imbalances, temporary mispricing between the futures and their respective underlying indices do occur. These transitory departures from theoretical pricing models, such as the cost-of-carry model, create an environment in which statistical arbitrage strategies can thrive.

The practical execution of statistical arbitrage strategies makes a powerful analytical technique that converts the raw data provided by the market into powerful trading signals. In this research, statistical arbitrage is grounded on the dynamic modelling of hedge ratios, calculation of hedged returns, generation of signals using rolling statistics and statistical properties like stationarity along with others are usefully checked to make inference. Each stage in this framework is designed to maximize both the interpretability and the effectiveness of the arbitrage signals, ensuring that the trading opportunities identified are both statistically valid and economically meaningful.

7.1 Construction of Hedged Return Series

At its core statistical arbitrage methodology is the construction of the hedged return series, which forms the primary spread to be analyzed and traded. In financial markets, the concept of a "spread" refers to the difference between two related financial quantities—most commonly, the prices or returns of assets that are expected to move together due to fundamental or statistical relationships [56]. In the context of this research, the spread is specifically defined as the return on a dynamically hedged portfolio comprising a position in the KSE 30 index (the spot asset) and an offsetting position in the corresponding futures contract. The central idea is that, by carefully calibrating the exposure to the futures contract, it is possible to reduce or neutralize the risk arising from movements in the spot market.

The formula used to calculate the hedged return at each time step t is mentioned in equation 7. The hedge ratio $\hat{\beta}_t$ quantifies the degree to which the futures position offsets the risk of the spot position. The value of $\hat{\beta}_t$ is not static but is recalculated at each time point using sophisticated models, such as DCC-GARCH, Copula-GARCH, LSTM-CNN, or FT-Net Hybrid, so as to capture changing market conditions and correlations.

The resulting hedged return series, $\hat{R}_{hedged,t}$, can be interpreted as the "residual risk" or "spread" that remains after applying the hedge. Ideally, if the hedge were perfect and all market movements were fully anticipated by the model, the hedged return would be close to zero [6]. However, in reality, due to estimation errors, market frictions, and the inherent unpredictability of financial markets, the spread will exhibit variability and, importantly for statistical arbitrage, may display patterns of mean reversion or deviation from equilibrium. By focusing on this spread, the statistical arbitrage methodology is able to detect and systematically exploit these temporary dislocations for profit.

7.2 Rolling Statistics and Z-Score Normalization

Next steps after computing the hedged return series, or spread, are to track how this spread behaves over time to get actionable trading signals. The primary tool used for this analysis is the calculation of rolling statistics, which ensure that abnormal deviations are detected with respect to most of the recent history rather than a fixed mean or standard deviation. This is essential in financial markets where statistical properties may evolve over time due to regime changes, volatility clustering, or shifts in investor sentiment.

The rolling mean and rolling standard deviation are defined mathematically as follows (equation 41) for a window of size w:

$$\mu_t = \frac{1}{w} \sum_{i=t-w+1}^{t} \hat{R}_{hedged,t} \tag{41}$$

$$\sigma = \sqrt{\frac{1}{w-1} \sum_{i=t-w+1}^{t} (\hat{R}_{hedged,t} - \mu_t)} \tag{42}$$

where μ_t represents the rolling mean of the hedged return at time t, and σ denotes the rolling standard deviation over the same window. By recalculating these statistics at each time step, the method adapts to evolving market conditions and ensures that the detection of anomalies is contextually relevant.

To further standardize the detection of trading opportunities, the spread is normalized into a z-score (see equation 43), which expresses the current hedged return in terms of its distance from the rolling mean, scaled by the rolling standard deviation:

$$z_t = \frac{\hat{R}_{hedged,t} - \mu_t}{\sigma_t} \tag{43}$$

The z-score, z_t, provides a dimensionless measure of extremity: values close to zero indicate that the spread is near its recent average, while large positive or negative values signify abnormal deviations. In the context of statistical arbitrage, these z-score thresholds form the basis for generating entry and exit signals. For example, when the z-score exceeds a certain positive threshold, the spread is considered "overbought" and likely to mean-revert downwards, triggering a short position. Conversely, when the z-score falls below a negative threshold, the spread is "oversold" and expected to revert upwards, prompting a long position. This approach leverages the inherent tendency of mean-reverting series to oscillate around a stable equilibrium, thus enabling the systematic exploitation of temporary mispricing.

7.3 Stationarity Testing and Justification

The effectiveness and validity of any statistical arbitrage strategy is largely dependent on whether the hedged return series is stationary. Statistical features that enable time series analysis are often gathered at regular intervals, making the time-series analysis of stationary data important. Stationarity implies that history matters for what may happen in the future. This is exactly the assumption behind repeated arbitrage opportunities. If the spread is not constant over time, then any movement away from the average could be permanent, which would render any trading strategy based on the idea of moving back towards the average too erratic to implement or too dangerous to implement.

To formally test for stationarity, the Augmented Dickey-Fuller (ADF) test is employed [57]. The ADF test is a statistical hypothesis test in which the null hypothesis is that a unit root is present in the time series, indicating non-stationarity. The test is based on estimating the regression shown in equation 44:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta \Delta y_{t-i} + \varepsilon_t \qquad (44)$$

where y_t is the value of the spread (here, $\hat{R}_{hedged,t}$), Δy_t denotes the first difference of y_t, t is a time trend, p is the number of lagged differences included to account for autocorrelation, and ε_t is the error term. The key parameter of interest is γ is significantly less than zero, the null hypothesis of a unit root is rejected, indicating stationarity.

The outcome of the ADF test includes the test statistic, critical values at standard significance levels, and the p-value. A sufficiently negative test statistic, or a p-value below the chosen significance threshold (typically 0.05), provides statistical evidence that the series is stationary. In the context of this research, the application of the ADF test to the hedged return series ensures that the mean-reversion signals generated by the z-score normalization are grounded in sound statistical properties, rather than being artifacts of a trending or random-walk process. This will not only make the arbitrage signals more reliable but also protect against model failure or false positive signal in the live trading environment.

7.4 Mean-Reversion Signal Design

The detection of deviations forms the basis for generating actionable trading signals. In implementation, the strategy uses the previously calculated z-score. This z-score tells us how many standard deviations the current value of the spread is from its rolling mean [58].

Hedged return also called a spread at time taken equal to the rolling mean over rolling standard deviation over a window equal μ_t and σ_t respectively. A z-score converts the spread into a standardized measure, which allows comparing offsets today directly, regardless of changes in its volatility or average level over time.

Trading signals are produced by determining an entry threshold (z_{entry}) and an exit threshold (z_{exit}). The thresholds are statistical cutoffs that determine what deviations away from the mean are significant and what level of reversion will be sufficient to close a trade. The entry and exit conditions for trading is mathematically defined as follows:

1. Long Entry (Buy Spread): When the z-score falls below the negative of the entry threshold, i.e., $z_t < -z_{entry}$, this signals that the spread is abnormally low and likely to mean-revert upwards. A long position is initiated, betting on an increase in the spread.
2. Short Entry (Sell Spread): Conversely, when the z-score rises above the positive entry threshold, i.e., $z_t > z_{entry}$, the spread is deemed abnormally high and expected to revert downwards. A short position is initiated.
3. Exit Signal (Close Position): Regardless of the initial direction, when the absolute value of the z-score returns below the exit threshold, i.e., $|z_t| < z_{exit}$, it indicates that the spread has normalized, and the trade should be closed to realize profits and mitigate the risk of reversal.

These rules can be mathematically summarized as in equation 45:

$$\Delta Position = \begin{cases} 1 & if z_t < -z_{entry} \\ -1 & if z_t > z_{entry} \\ 0 & if\ |z_t| < z_{exit} \end{cases} \qquad (45)$$

This systematic approach increases the likelihood of making profitable trades or closing out losing trades. They are opened once confirmed deviations and closed when the deviation is gone. Using z-scores rather than absolute values, it provides an adaptive strategy that responds to market conditions and volatility regimes.

7.5 Trade Position Management

The management of trading positions in a statistical arbitrage strategy is governed by the signals described above, with explicit rules dictating when to enter, hold, or exit trades. At any given time, the strategy can be in one of three possible states: long, short, or flat (no position).

A long position is taken when the spread is judged to be excessively low, based on the z-score dropping below the negative entry threshold. In practical terms, this involves buying the spot index and selling the corresponding number of futures contracts as specified by the hedge ratio, with the expectation that the spread will rise. Conversely, a short position is initiated when the z-score exceeds the positive entry threshold, indicating the spread is excessively high; this entails selling the spot index and buying futures contracts, profiting from a decline in the spread.

Trade transitions are managed through a straightforward process. When the z-score signal triggers a new position (long or short) and the strategy is currently flat, the position is opened accordingly. If a position is already open and the z-score crosses the exit threshold in the direction of normalization, the position is closed, returning the strategy to a flat state. Importantly, the system is designed to avoid simultaneous long and short positions; at most, only one direction is active at any time. If the signal reverses before the exit threshold is hit (for example, from long entry to short entry without normalization), the previous position is first closed before the new position is opened, ensuring clear transitions and accurate accounting of profits and losses.

This position management logic ensures discipline, prevents overtrading, and allows for clean measurement of individual trade performance. The sequence of position changes is directly mapped to the time series of z-score signals, making the strategy transparent and auditable.

7.6 Parameter Optimization

The performance of statistical arbitrage hinges significantly on the selection and tuning of key parameters; namely, the rolling window size used for computing statistics, and the entry and exit z-score threshold values. How often and how profitably a trading strategy might trade is determined by these parameters, as is how robust the strategy is to various situations.

The size of the rolling window (w) affects the mean and standard deviation of the spread. When a window is smaller in size. The strategy will respond to the latest change effectively. But, it can miss on big trends. On the other hand, if you have a bigger window, you will smooth short-term variations and perhaps prevent overfitting, but you can be slow to react to real regime changes. Choosing a window size is, therefore, a trade-off between sensitivity and stability. Using out-of-sample back testing and validation for empirical testing helps find the optimal window to balance these competing concerns specifically for the KSE 30 futures market.

Changes to the entry z-score threshold (z_{entry}) and exit z-score threshold (z_{exit}) will affect the responsiveness of the strategy. When entry thresholds are lower, traders trade frequently as signals are triggered with even small deviations from the mean. Although this may lead to greater opportunities, it also increases the chances of a false positive and transaction costs. When the thresholds are high, the strategy will only activate when conditions are extraordinary. The idea is that the quality of trades will be superior on average. However, it will also activate far less often. Thus, total profitability might be lowered due to fewer returns and trades. The exit threshold controls how tightly anyone control trades. A tighter exit locks in the profit quickly and limits the drawdown. A looser exit will allow us to capture more profit but increases the odds of a reversal.

Usually, the optimization of parameters is done through in-depth back testing of a grid of parameters. Then the raw return, risk-adjusted return (Sharpe ratio), drawdown, and number of trades are measured. The choice of final parameters is done to balance profitability and risk and operational suitability. The optimized strategy should not rely heavily on any specific period or historical regime. Thus, a sensitivity analysis must be conducted to test the results across different market conditions.

7.7 Empirical Results and Visualization

7.7.1 Model-by-Model Strategy Performance

Upon executing the statistical arbitrage strategy testing interface, seen in figure 19, the analysis was conducted with a rolling window of 60 days, an entry z-score threshold of 2.0, and an exit z-score threshold of 0.5. The empirical evaluation covers four distinct models for hedge ratio estimation: Copula-GARCH, DCC-GARCH, LSTM–CNN Hybrid, and FT-Net Hybrid. Each model file encapsulates daily returns data for the KSE 30 index and its futures contract, along with the dynamically computed hedge ratio and the resulting hedged return series. Figure 18 also shows the GUI developed to assess the statistical arbitrage effectiveness. A summary of the performance metrics for each model is provided in Table 4 below. These metrics include total return, annualized return, annualized volatility, Sharpe ratio, maximum drawdown, number of trades executed, and results of the Augmented Dickey-Fuller (ADF) stationarity test. The Sharpe ratio, which measures risk-adjusted performance, is used as the principal comparative criterion for determining the best performing strategy.

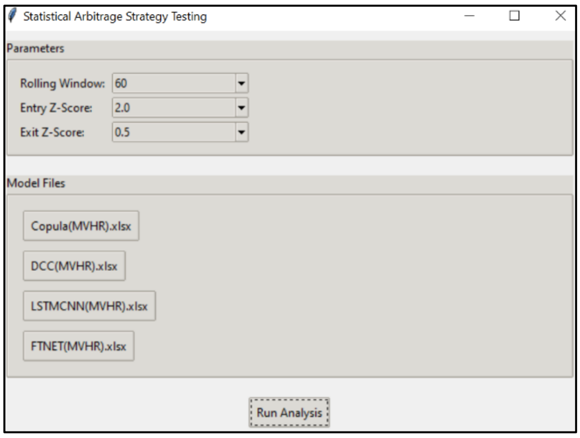

Figure 19. Graphical User Interface (GUI) for Statistical Arbitrage.

Table 4. Statistical Arbitrage Performance Metrics by Model.

Model	Total Return	Annual Return	Annual Volatility	Sharpe Ratio	Max Drawdown	Trades	ADF Statistic	ADF p-value	Crit 5%
Copula GARCH	0.056	0.011	0.012	0.864	-0.015	54	-13.003	0.000	-2.864
DCC GARCH	0.059	0.011	0.012	0.951	-0.014	54	-13.020	0.000	-2.864
LSTM-CNN Hybrid	0.066	0.013	0.012	1.074	-0.015	52	-39.123	0.000	-2.864
FT-NET Hybrid	0.066	0.013	0.012	1.077	-0.015	52	-38.939	0.000	-2.864

The FT-Net Hybrid model marginally outperformed all other models in terms of Sharpe ratio, closely followed by the LSTM–CNN Hybrid. Both advanced deep learning models produced higher total and annualized returns than the econometric Copula-GARCH and DCC-GARCH models, while maintaining comparable volatility and drawdown characteristics. The strong negative ADF statistics and zero p-values confirm robust stationarity in the hedged return series for all models, substantiating the statistical basis for mean-reversion-based arbitrage.

7.7.2 Visualization of Hedged Spread Dynamics

The dynamic behavior of the hedged return (spread) for each model is visualized in figures 20-23. Each graph plots the time series of hedged returns along with its rolling mean, entry bands (±2.0 standard deviations), and exit bands (±0.5 standard deviations) over the entire sample period from 2019 to 2024.

The hedged return series produced by the Copula-GARCH model oscillates around the rolling mean, with most values contained within the entry bands. Periodic spikes represent significant short-term dislocations, often coinciding with market stress or contract rollover periods. The rolling mean remains stable near zero, validating the effectiveness of the hedge. The width of the entry and exit bands adapts dynamically with volatility, expanding during turbulent periods and narrowing during tranquil market regimes.

The DCC-GARCH model's hedged return dynamics shows patterns broadly similar to the Copula-GARCH model. The frequency and magnitude of excursions beyond the entry bands are slightly higher during high volatility episodes, yet the spread consistently reverts to the mean. This mean-reversion property is essential for arbitrage, as it allows repeated entry and exit opportunities throughout the sample.

Next comes the LSTM–CNN Hybrid model's hedged return series. This model exhibits a slightly tighter clustering of returns around the mean, with fewer extreme outliers. The dynamic entry and exit bands provide a visual cue for when the arbitrage strategy is likely to activate trading signals. The persistently mean-reverting behavior is evident, reinforcing the statistical foundation of the arbitrage approach.

The FT-Net Hybrid model's spread demonstrates exceptional mean-reverting tendencies, with the vast majority of values remaining within the rolling bands. The spectral features incorporated by this model appear to enhance the stability and predictability of the spread, reducing the occurrence of unprofitable outliers. The visual compactness and symmetry of the hedged returns suggest robust risk control.

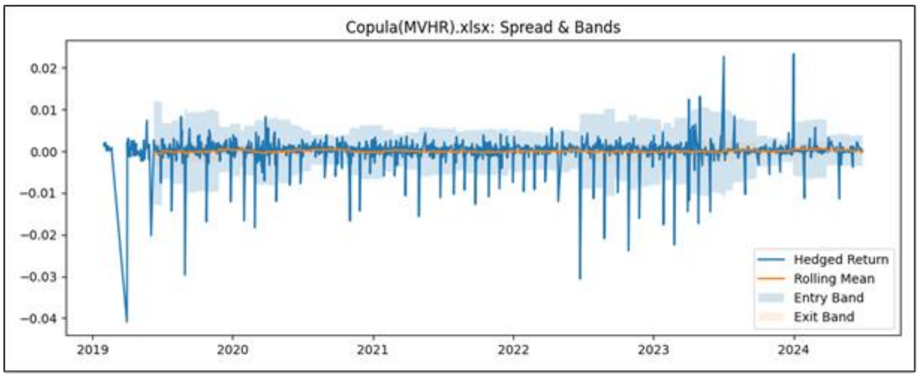

Figure 20. Spreads and Bands – Copula DCC GARCH.

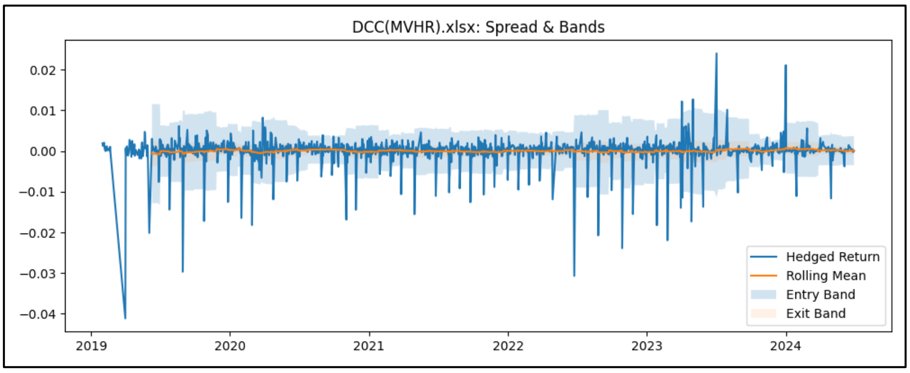

Figure 21. Spreads and Bands –DCC GARCH.

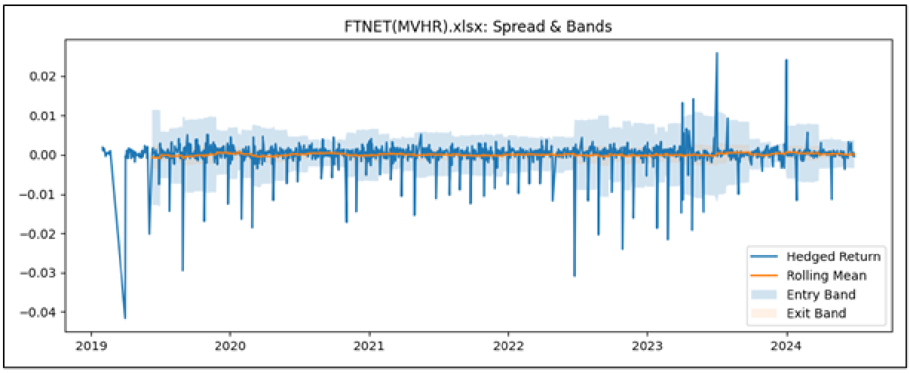

Figure 22. Spreads and Bands – FTNET Hybrid.

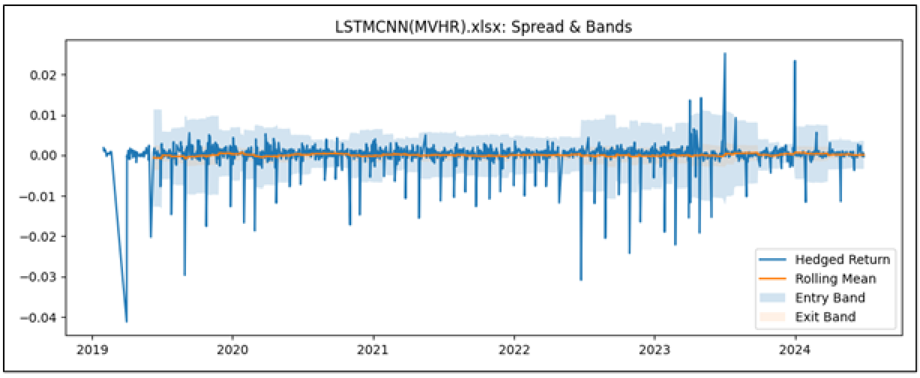

Figure 23. Spreads and Bands – LSTM-CNN

7.7.3 Equity Curve and Comparative Performance

To evaluate actual trading performance, Figures 24 – 27 plots the accumulated profit-and-loss (PnL) curves and entries of all the individual trades for the model. Accumulated PnL represents the total impact of each arbitrage trade, signed based on the s-z-score before its entry and exit decision.

The equity curve of the Copula-GARCH model demonstrates a stable rising tendency with a moderate number of drawdowns in the early sample. Each trade is marked to facilitate understanding, clearly demonstrating that participating in long and short trades both contribute to the increased profitability of both parties.

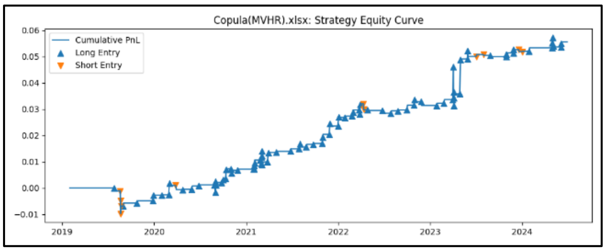

Figure 24. Strategy Equity Curve – Copula DCC GARCH.

The DCC-GARCH model also follows a comparable trend, albeit with a smoother evolution and lesser drawdown episodes. The gradual changes in the equity curve show that the model can deal with changing volatility conditions and achieve consistent performance.

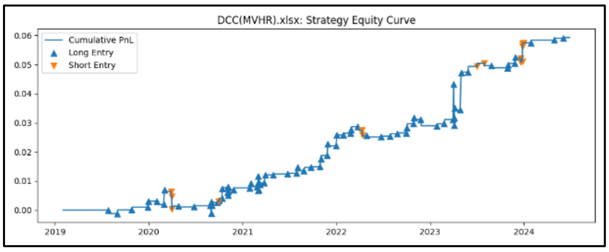

Figure 25. Strategy Equity Curve – DCC GARCH.

The LSTM-CNN Hybrid model has seen an increase in total PnL, with a longer rising equity curve and small drawdowns. The model's adaptive learning mechanism determines when to go long or short and helps the investor earn a higher compounded return than competing econometric models.

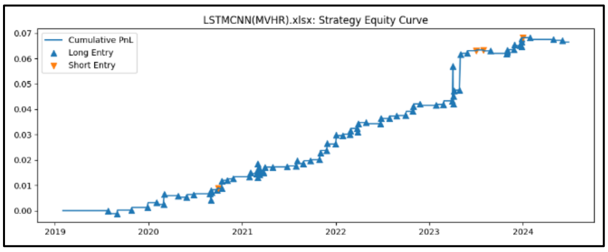

Figure 26. Strategy Equity Curve – LSTM-CNN

FT-Net Hybrid has the highest cumulative return among all models in terms of equity curve. The equity curve shows a generally upward direction, especially in the later sample years, with few reversals. The FT-Net Hybrid model is the number one model by Sharpe ratio. This is due to its robust and consistent execution of trade.

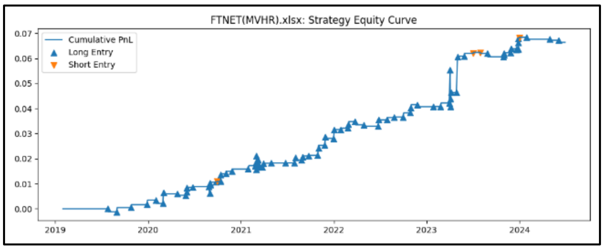

Figure 27. Strategy Equity Curve – FTNET Hybrid.

Finally, figure 27 overlays the cumulative PnL curves for all four models, providing a direct visual comparison of strategy performance over time. This comparative equity curve reveals the relative strengths and weaknesses of each model. While all models demonstrate positive growth and mean-reversion-driven profitability, the FT-Net Hybrid and LSTM–CNN Hybrid models consistently outperform the econometric alternatives, both in terms of total return and in the smoothness of the equity curve. The absence of large, persistent drawdowns further confirms the statistical robustness and practical viability of the deep learning approaches.

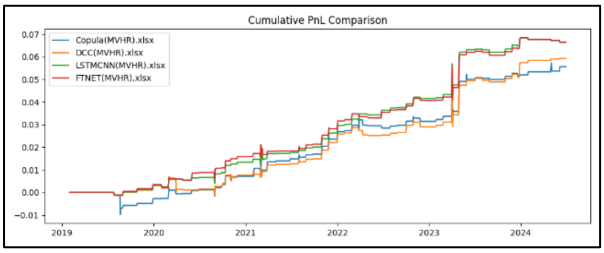

Figure 28. Cumulative PnL Comparison Across Models.

7.8 Interpretation

The comparative analysis of model performance is grounded in a comprehensive suite of quantitative metrics and visual diagnostics. Across all four tested models—Copula-GARCH, DCC-GARCH, LSTM–CNN Hybrid, and FT-Net Hybrid—the results consistently indicate that advanced deep learning architectures, specifically the FT-Net Hybrid and LSTM–CNN Hybrid, offer a superior edge in the detection and exploitation of statistical arbitrage opportunities.

The FT-Net Hybrid model emerges as the most effective, as evidenced by its highest Sharpe ratio of 1.07680 and a total return nearly identical to that of the LSTM–CNN Hybrid. Both models deliver not only higher absolute and risk-adjusted returns but also demonstrate more consistent equity curve progression, marked by persistent upward momentum and limited drawdowns throughout the out-of-sample evaluation period. The econometric models, while still profitable and statistically significant, underperform relative to their deep learning counterparts both in terms of total returns and risk-adjusted metrics.

Figures 24 - 27 depicting the cumulative profit and loss trajectory underscore this conclusion visually: the FT-Net Hybrid's equity curve is not only smoother and less volatile but also achieves the highest terminal value over the full testing horizon. The hedged return spread associated with this model is characterized by a strong mean-reverting tendency, fewer extreme outliers, and more predictable band crossings, all of which translate to high-quality arbitrage signals. Furthermore, the ADF test statistics for all models are strongly significant, but the FT-Net Hybrid and LSTM–CNN Hybrid models exhibit the most pronounced levels of stationarity in their spreads, further cementing the statistical reliability of the arbitrage process.

8. CONCLUSION

This study delivers a comprehensive examination of dynamic hedging in equity futures and statistical arbitrage detection, comparing traditional econometric frameworks with advanced deep learning architectures on the KSE-30 Index and its futures. Core contributions include implementing and empirically evaluating models under realistic market constraints, applying rigorous multi-criteria performance metrics, and integrating arbitrage-based trading analysis. A key preprocessing challenge, like contract rollover discontinuities, was resolved through a mean-reverting adjustment, aligning new contract prices with expiring series to eliminate artificial jumps and preserve hedge ratio integrity. Missing interest rate data was imputed using a Random Forest regressor, achieving strong R^2 and RMSE scores, ensuring a reliable dataset.

Econometric models, DCC-GARCH and its Student-t Copula extension, were statistically validated via ADF and ARCH-LM tests, confirming stationarity and conditional heteroscedasticity suitability. DCC-GARCH captured volatility clustering and leverage effects, while Copula DCC-GARCH modeled symmetric tail dependence, improving extreme co-movement representation. Parallelly, two hybrid deep learning models were developed. The LSTM–CNN combined convolutional feature extraction with recurrent sequence modeling, adapting to short- and long-term dependencies in noisy returns. The novel FT-Net hybrid incorporated Fourier spectral decomposition with temporal convolutional and recurrent layers, exploiting hidden cyclical patterns and structural shifts often missed in time-domain approaches. Post-model diagnostics confirmed that all models produced stationary, homoscedastic, and autocorrelation-free residuals, validating them for comparative hedging analysis.

Performance evaluation via Fuzzy TOPSIS ranked models across nine criteria, variance reduction, RMSE, Sharpe ratio, hedged return, directional accuracy, MAD, time complexity, and tail risk metrics (VaR, CVaR). The FT-Net hybrid consistently ranked first in both in-sample and out-of-sample testing, achieving the largest variance reduction, lowest forecast RMSE, highest directional accusracy, and superior Sharpe ratios, while mitigating extreme tail risks more effectively than all competitors. Copula DCC-GARCH showed resilience in stress periods but lacked adaptability under rapid volatility shifts, and LSTM–CNN, while robust, delivered slightly lower hedged returns than FT-Net. In statistical arbitrage testing, FT-Net's hedge ratio-derived spread signals displayed strong mean reversion, enabling profitable long-short strategies with economically significant returns.

Overall, the findings underscore the dominance of hybrid deep learning, particularly FT-Net Hybrid, in dynamically estimating hedge ratios, reducing portfolio risk, and uncovering actionable arbitrage opportunities. Its capacity to integrate spectral–temporal features with sequential modeling positions it as a scalable, high-performance alternative to traditional econometric models in modern financial markets.

8.1 Recommendations for Future Work

Although the current study contributes to the understanding of dynamic hedge effectiveness and statistical arbitrage based on econometric and deep learning models, there are still multiple roads to pursue in the future. Generalizing the framework to a multivariate asset context where a portfolio of hedge ratios is estimated as opposed to on a single index, in the authors view this would enable the future researchers to collect cross-asset correlations, sectoral connectivity, and non-linear interdependencies in a portfolio. This can be done by using multivariate DCC-GARCH, vine copulas or transformer-based architectures with attention mechanisms all of which have the potential to model inter-process dynamics more richly.

Other critical developments include using the higher frequency data. Although this work dealt with daily horizons on the basis of liquidity and availability considerations, more detailed analysis of volatility clustering, structural breaks and arbitrage opportunities could be done with intraday or tick-level modeling. Integrating these types of datasets with deep reinforcement learning or hybrid neural networks could help increase flexibility in high-frequency trading contexts, though it would demand ultra-low-latency data pipelines, strong noise filtering, and compute scaling.

Another area that should be looked into in the future is integration of macroeconomic indicators. GDP, inflation and policy rates may be added to this and also including wider domestic and global indicators would enrich the contextual interpretation of hedge ratio behaviors, e.g. industrial production, fiscal balances, foreign reserves, crude oil prices, U.S Treasury yields, and volatility indices. More advanced time-varying parameter models, Bayesian hierarchal models or interpretable models like Temporal Fusion Transformers might enhance regime detection and result in a more stable forecast over changing macro-financial conditions.

Lastly, it is also important to improve the transaction cost modeling. Instead of the fixed costs assumption, dynamic cost structures based on liquidity situations, bid ask spreads, slippage and execution lag would provide more realistic evaluation of strategy feasibility- especially in young or thin markets. Adding regulatory constraints, including restrictions on leverage (as in Dosdd-Frank), short-sale restrictions, and margin requirements would make results closer to those experienced in the real-world trading environments. Practically, the creation of modular, open-source APIs or code libraries around the hedging and arbitrage models, described here, would become an important benefit toward cross-market validation, model reproducibility, and ongoing model re-training via re-deployable, real-time, cloud-based architecture.

By making these additions, not only would one widen the field of this work, but one would also fill the gap between academic modeling and the deployable, adaptive risk management systems in the changing global markets.

9. ACKNOWLEDGMENTS

The authors of this study express their gratitude to NED University of Engineering and Technology, for their facilities and support, throughout the course of the research, which was instrumental in shaping the direction and successful completion of this research.

10. REFERENCES

[1] Khan , S. U. (2006). Role of the Futures Market on Volatility and Price Discovery of the Spot Market: Evidence from Pakistan's Stock Market. The Lahore Journal of Economics, 107-121.

[2] Malik, I. R., & Shah, A. (2017). The Impact of Single Stock Futures on Market Efficiency and Volatility: A Dynamic CAPM Approach. Emerging Markets Finance and Trade , 339-356.

[3] Khan, S. U., & Abbas, Z. (2013). Does Equity Derivatives Trading Affect the Systematic Risk of the Underlying Stocks in an Emerging Market: Evidence from Pakistan's Futures Market. The Lahore Journal of Economics, 63-80.

[4] Ederington, L. H. (1979). The hedging performance of the new futures markets. The Journal of Finance, 157-170.

[5] Ghosh, A. (1993). Hedging with Stock Index Futures: Estimation and Forecasting with Error Correction Model. Journal of Futures Markets, 743-743.

[6] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007.

[7] Bollerslev. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 307-327.

[8] Chen, W., Hussain, W., Cauteruccio, F., & Zhang, X. (2023). Deep Learning for Financial Time Series Prediction: A State-of-the-Art Review of Standalone and Hybrid Models. CMES - Computer Modeling in Engineering and Sciences, 187-224.

[9] François, P., Gauthier, G., Godin, F., & Mendoza, C. O. (2025). Is the difference between deep hedging and delta hedging a statistical arbitrage? Finance Research Letters.

[10] Horikawa, H., & Nakagawa, K. (2024). Relationship between deep hedging and delta hedging: Leveraging a statistical arbitrage strategy. Finance Research Letters.

[11] Lai, Y. S. (2018). Dynamic Hedging with Futures: A Copula-Based GARCH Model with High-Frequency Data. *Review of Derivatives Research*, 307-329.

[12] Baillie, R. T., & Myers, R. J. (1991). Bivariate GARCH estimation of the optimal commodity futures hedge. Journal of Applied Econometrics, 109-24.

[13] Baba, Y., Engle , R. F., Kraft, D., & Kroner, K. F. (1990). Multivariate simultaneous generalized ARCH. Department of Economics, Unuveristy of California, San Diego.

[14] Bollerslev, T. (1990). Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model . The Review of Economics and Statistics, 498-505.

[15] Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. Journal of Business & Economic Statistics, 339-350.

[16] Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. International economic review, 47(2), 527-556.

[17] Qu, H., Wang, T., Zhang, Y., & Sun, P. (2019). Dynamic hedging using the realized minimum variance hedge ratio approach - Examination of the CSI 300 index futures. Pacific-Basin Finance Journal.

[18] Buehler, H., Gonon, L., Teichmann, J., & Wood, B. (2018). Deep hedging. Quantitative Finance, 1271-1291.

[19] Brigola, R. (2025). Further Applications of the Fourier Transform. In Fourier Analysis and Distributions: A First Course with Applications (pp. 383-436). Cham: Springer Nature Switzerland.

[20] Kristjanpoller, W., & Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. Expert Systems with Applications, 233-241.

[21] Mahmoudi, A. (2025). Investigating LSTM-based time series prediction using dynamic systems measures. Evolving Systems.

[22] Ma, B. (2024). Meta-Learning Enhanced Trade Forecasting: A Neural Framework Leveraging Efficient Multicommodity STL Decomposition. International Journal of Intelligent Systems, 1-21.

[23] Yeo, J., & Papanicolaou, G. (2017). Risk control of mean-reversion time in statistical arbitrage. Risk and Decision Analysis, 263-290.

[24] Wang, H., Zhang, X., & Long. (2024). Wavelet-based Fourier-Enhanced Network Model Decomposition for Multivariate Long-Term Time-Series Forecasting. Neural Processing Letters.

[25] Carchano, Ó., & Pardo, Á. (2009). Rolling over stock index futures contracts. The Journal of Future Markets.

[26] Vojtko, R., & Padysak, M. (2020). Continuous Futures Contracts Methodology for Backtesting. SSRN.

[27] E, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, 427-431.

[28] Xu, R., & Li, X. (2017). Study About the Minimum Value at Risk of Stock Index Futures Hedging Applying Exponentially Weighted Moving Average - Generalized Autoregressive Conditional Heteroskedasticity . International Journal of Economics and Financial Issues, 104-110.

[29] Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 347-370.

[30] Jondeau, E., & Rockinger, M. (2006). The copula-garch model of conditional dependencies: An international stock market application. Journal of international money and finance, 25(5), 827-853.

[31] Hafner, C. M., & Manner , H. (2012). Dynamic stochastic copula models: Estimation, inference and applications. Journal of applied econometrics, 27(2), 269-295.

[32] Hsu, C.-C., Tseng, C.-P., & Wang, Y.-H. (2008). Dynamic Hedging with Futures: A Copula-Based GARCH Model. Journal of Futures Markets, Forthcoming, 34.

[33] Demarta, S., & McNeil, A. J. (2005). The t Copula and Related Copulas. International Statistical Review, 73(1), 111-129.

[34] Kumar, S., Rao, A., & Dhochak, M. (2025). Hybrid ML models for volatility prediction in financial risk management. International Review of Economics & Finance.

[35] Giantsidi, S., & Claudia, T. (2025). Deep Learning for Financial Forecasting: A Review of Recent Advancements. SSRN, 1-94.

[36] Widiputra, H., Mailangkay, A., & Gautama, E. (2021). Multivariate CNN-LSTM Model for Multiple Parallel Financial Time-Series Prediction. Complexity, 1-14.

[37] Balamohan, S., & Khanaa, V. (2024). Revolutionizing Stock Market Prediction: Harnessing the Power of LSTM and CNN Hybrid Approach. SN Computer Science.

[38] Salem, F. M. (2021). Gated RNN: The Long Short-Term Memory (LSTM) RNN. In F. M. Salem, Recurrent Neural Network (pp. 71-82). Springer.

[39] Wang, Y., Xiao, Z., & Cao, G. (2022). A convolutional neural network method based on Adam optimizer with power-exponential learning rate for bearing fault diagnosis. Journal of Vibroengineering, 666-678.

[40] Singh, P., Jha, M., Sharaf, M., El-Meligy, M. A., & Gadekallu, T. R. (2023). Harnessing a Hybrid CNN-LSTM Model for Portfolio Performance: A Case Study on Stock Selection and Optimization. IEEE Access, 104000-104015.

[41] Li, J., Yang, X., Qian, T., & Xie, Q. (2023). The adaptive Fourier decomposition for financial time series. Engineering Analysis with Boundary Elements, 139-153.

[42] Li, S., Wang, Z., Wang, X., Yin, Z., & Yao, M. (2025). Frequency-enhanced and decomposed transformer for multivariate time series anomaly detection. Applied Intelligence, 1-18.

[43] Prakash, A., James, N., Menzies, M., & Francis, G. (2021). Structural Clustering of Volatility Regimes for Dynamic Trading Strategies. Applied Mathematical Finance, 236-274.

[44] Yang, Z., Yin, M., Liao, J., Xie, F., Zheng, P., Li, J., & Hua, B. (2025). FFTNet: Fusing Frequency and Temporal Awareness in Long-Term Time Series Forecasting. Electronics, 1-33.

[45] Grabinski, M., & Klinkova, G. (2024). A new measure of risk using Fourier analysis. arXiv: Quantitative Finance, 14.

[46] Lotfi, F. H., Allahviranloo, T., Pedrycz, W., Shahriari, M., Sharafi, H., & GhalehJough, S. R. (2023). Technique for Order Preferences by Similarity to Ideal Solutions (TOPSIS) in Uncertainty Environment. In F. H. Lotfi, T. Allahviranloo, W. Pedrycz, M. Shahriari, H. Sharafi, & S. R. GhalehJough, Fuzzy Decision Analysis: Multi Attribute Decision Making Approach (pp. 141-178). Springer.

[47] Sood, K., Pathak, P., & Gupta, S. (2025). How do the determinants of investment decisions get prioritized? Peeking into the minds of investors. Kybernetes, 2175-2203.

[48] Ardil, C. (2021). Fighter Aircraft Evaluation and Selection Process Based on Triangular Fuzzy Numbers in Multiple Criteria Decision Making Analysis Using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). International Journal of Computer and Systems Engineering, 402-409.

[49] Manzoor, S., Mustafa, S., Gulzar, K., & Gulzar, A. (2024). MultiFuzzTOPS: A Fuzzy Multi-Criteria Decision-Making Model Using Type-2 Soft Sets and TOPSIS. Symmetry.

[50] Sureeyatanapas, P., & Sriwattananusart, K. (2018). Supplier selection towards uncertain and unavailable information: An extension of TOPSIS method. Operations Research Perspectives, 69-72.

[51] Mattera, R., Di Sciorio, F., & Trinidad-Segovia, J. E. (2022). A composite index for measuring stock market inefficiency. Complexity, 2022(1), 9838850.

[52] Krauss, C. (2016). Statistical Arbitrage Pairs Trading Strategies: Review And Outlook. The Journal of Economic Surveys, 513-545.

[53] Liu, P. (2023). Statistical Arbitrage with Hypothesis Testing. In In Quantitative Trading Strategies Using Python: Technical Analysis, Statistical Testing, and Machine Learning (pp. 225-255). Berkeley: CA: Apress.

[54] Ye, H., Chen, J., Gong, S., Jiang, F., Zhang, T., Chen, J., & Gao, X. (2025). Adaptive Time-Frequency Ensembled Network for Long-term Time Series Forecasting. ATF-Net.

[55] Chan, E. P. (2021). Quantitative Trading: How to Build Your Own Algorithmic Trading Business. John Wiley & Sons.

[56] Brandimate, P. (2017). An Introduction to Financial Markets: A Quantitative Approach. John Wiley & Sons.

[57] Paiva, D. d. (2021). Study of Tests For Trend In Time Series . Brazilian Journal of Biometrics.

[58] Li, X., Tripe, D., & Malone, C. (2017). Measuring bank risk: An exploration of z-score. 1-38.

[59] Figlewski, S. (1984). Hedging Performance and Basis Risk in Stock Index Futures. The Journal of Finance, 657-669.

[60] Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. Statistical Analysis and Data Mining.

[61] Alshenawy, F. Y. (2024). Using Tail Dependence on Copula-based Regression Models in Mixed Data. The Egyptian Statistical Journal , 68(2), 65-85.