ACADEMIC PUBLISHER

INTERNATIONAL JOURNAL OF DATA SCIENCE AND MACHINE LEARNING (ISSN: 2692-5141)

Volume 05, Issue 02, 2025, pages 258-268
Published Date: - 26-10-2025

A Hybrid Volatility-Driven Statistical Arbitrage Framework: Integrating Advanced Time-Series Econometrics and Machine Learning for Enhanced Stock Market Trend Prediction

Piotr S. Volkov
Faculty of Computational Finance, Moscow State University, Moscow, Russia

Aisha N. Mensah
Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Abstract

Purpose: This study introduces the Hybrid Volatility-Driven Statistical Arbitrage (HVSA) framework, an integrated quantitative strategy designed to enhance stock market trend prediction and exploit mean-reversion opportunities. The methodology synergistically combines advanced time-series econometrics with machine learning to generate robust and non-spurious trading signals, specifically focusing on mid-range volatility assets.
Methods: The HVSA framework employs a multi-stage approach. First, the Gaussian Mixture Model is utilized to cluster a large universe of assets based on their drift-independent realized volatility profiles, isolating those in optimal, tradable volatility regimes. Second, the Granger Causality Test is applied to the resultant clusters to rigorously identify predictive, causal linkages between asset pairs, moving beyond simple co-integration. The extracted causal features, along with advanced Volatility-of-Volatility metrics, are then fed into a Deep Neural Network classifier, which is trained using an adaptive resampling protocol to predict the directional trend of the arbitrage spread. The entire strategy is validated using a stringent forwardtesting protocol, which accounts for realistic market constraints and transaction costs.
Results: Empirical evaluation demonstrates that the HVSA framework achieves superior risk-adjusted returns compared to traditional benchmarks. The strategy's predictive power, rooted in verified causal volatility dependencies, resulted in an Annualized Sharpe Ratio of 2.31 and a low Maximum Drawdown of 4.2% during the forwardtesting period. The Deep Neural Network, continuously retrained via an adaptive rolling-window scheme, proved highly effective in capturing the non-linear patterns of mean reversion, a task where simpler linear models often fail.
Conclusion: The HVSA framework provides compelling evidence that the integration of statistically-rigorous volatility analysis with advanced, adaptively trained machine learning classification is crucial for developing robust statistical arbitrage strategies. This hybrid methodology successfully enhances the predictability of stock market trends and offers a viable pathway for generating alpha in contemporary financial markets.

Keywords

Statistical Arbitrage, Volatility, Machine Learning, Deep Neural Networks, Time-Series Econometrics, Granger Causality, Forwardtesting.

INTRODUCTION

The dynamics of global financial markets are intrinsically linked to volatility, a measure of the dispersion of returns that serves as a critical proxy for risk and uncertainty [2]. In contemporary quantitative finance, the pursuit of alpha—returns in excess of a market benchmark—increasingly revolves around the systematic exploitation of these volatility dynamics. Statistical Arbitrage (StatArb), a class of strategies built on the principle of mean-reversion, seeks to capitalize on transient price deviations between statistically related assets, anticipating their convergence to an equilibrium price [7]. This approach inherently relies on the stable co-movement of assets, a relationship that is frequently disrupted by market microstructure noise, unpredictable external shocks, and shifts in investor sentiment [3].

The traditional foundation of StatArb, primarily pairs trading, has historically utilized simpler econometric tools such as co-integration or correlation measures to establish asset relationships [9]. However, the ever-increasing efficiency and complexity of modern markets have exposed the limitations of these linear models. They often prove insufficient to capture the intricate, non-linear dependencies that govern high-frequency price movements, nor can they effectively adapt to the sudden shifts in volatility regimes that signal changes in market structure [2]. The quest for more robust and adaptive models has naturally led researchers and practitioners toward the integration of computational intelligence.

1.2. The Role of Computational Intelligence in Quant Trading

The past decade has witnessed a pivotal shift with the adoption of Machine Learning (ML) and Deep Learning (DL) in quantitative trading [1, 10]. These techniques offer a powerful means to model the highly complex, non-linear relationships and high-dimensional features inherent in financial time series. DL, particularly with architectures like Deep Neural Networks (DNNs) and Multi-Layer Perceptrons (MLPs), possesses a remarkable capacity to map input features—such as historical prices, technical indicators, and volatility metrics—to future directional outcomes without the restrictive distributional assumptions of traditional econometrics [4].

The application of computational intelligence is not merely a matter of improved curve-fitting; it fundamentally changes the nature of feature engineering. Sophisticated algorithms can be deployed to automatically discern intricate market patterns, for example, identifying cross-asset dependencies through graph-based representations [6]. Furthermore, in finance, where profitable trading opportunities are often fleeting and data is inherently imbalanced, ML provides the tools for robust data-centric solutions, including advanced resampling and optimization techniques designed to mitigate dataset bias and enhance classifier performance [12, 13].

1.3. Problem Statement and Research Gap

Despite the clear benefits of applying advanced computational techniques, a critical gap persists in the StatArb literature. Existing models often treat volatility and predictive signal generation as separate, sequential steps. They primarily use simple price-based co-movement (e.g., a z-score of the spread) as the signal, which may be susceptible to regime changes in the underlying asset volatility. The central challenge is thus to construct a StatArb framework that is: (1) Volatility-Informed—meaning the trading signal is explicitly linked to and conditional on the prevailing volatility regime; and (2) Causally-Validated—ensuring that the identified price discrepancies and predictive relationships are statistically robust and less likely to be spurious [14, 15].

A robust StatArb strategy must not only identify when a price divergence occurs but also correctly predict when and if mean-reversion is probable, while strictly managing risk during periods of high, unpredictable volatility [5]. The existing research often fails to integrate a rigorous causal inference mechanism into the feature selection stage, leading to strategies that, while profitable in backtesting, are fragile in a live-trading environment. This study proposes to address this deficiency through a new, integrated framework.

1.4. Key Contributions and Article Structure

This paper introduces the Hybrid Volatility-Driven Statistical Arbitrage (HVSA) framework, a novel methodology for enhanced stock market trend prediction. The core contribution lies in a multi-stage, interdisciplinary pipeline that explicitly links asset clustering based on volatility, causal inference using time-series econometrics, and non-linear classification via deep learning.

The HVSA pipeline is structured as follows:

1. Volatility Clustering: Identifying stable groups of assets sharing analogous volatility characteristics.
2. Causal Feature Selection: Rigorous application of Granger Causality to filter for non-spurious, predictive relationships within these clusters.
3. Non-linear Trend Prediction: Employing a Deep Neural Network to classify the spread's future directional movement based on the causally-validated features.
4. Forwardtesting Validation: Utilizing a strict, out-of-sample forwardtesting protocol to validate the strategy’s efficacy under realistic market conditions [1, 8].

The subsequent sections detail the methodology, present empirical results from the framework's application, discuss the implications of integrating causal inference with machine learning, and conclude with a summary of the findings and potential avenues for future research.

2. Methods: The Hybrid Volatility-Driven Statistical Arbitrage (HVSA) Framework

The HVSA framework is a systematic approach designed to move beyond the limitations of simple price-based co-integration in statistical arbitrage. The objective is to construct a market-neutral portfolio whose excess returns are driven by mean-reversion signals filtered through a rigorous volatility and causality lens.

2.1. Data Acquisition and Preprocessing

The analysis utilizes minute-level high-frequency Open-High-Low-Close (OHLC) price data and volume for a large universe of constituent stocks from the S&P 500 index. A multi-year historical dataset is partitioned into an in-sample training set, a validation set, and a final, untouched out-of-sample forwardtesting set.

Data preprocessing is critical in high-frequency finance. Log-returns are calculated to induce stationarity in the time series, a prerequisite for most econometric tests, including Granger Causality [9]. The raw OHLC data is also utilized to compute high-efficiency realized volatility estimators, which are less susceptible to market microstructure noise than simple daily returns. All time series are standardized via z-score normalization to ensure that the scale differences between assets do not unduly influence the clustering or machine learning stages.

2.2. Volatility Estimation and Asset Clustering

2.2.1. Realized Volatility Estimation

The classical method of estimating volatility via the standard deviation of close-to-close returns is known to be biased due to the presence of market noise and infrequent trading effects. To mitigate this, we employ the Yang-Zhang (YZ) estimator [11], an extension of the Garman-Klass estimator, which incorporates the open price in addition to the high, low, and close prices. The YZ estimator is drift-independent and is considered one of the most efficient for daily volatility estimation, providing a more robust measure of the true, unobservable underlying volatility of the asset. The realized volatility \sigma_t^2 is computed over a rolling window of N periods, with the choice of N calibrated to capture mid-range volatility regimes suitable for StatArb.

2.2.2. Volatility Regime Clustering

The HVSA framework postulates that the stability and predictability of arbitrage spreads are conditional on the assets belonging to similar, non-extreme volatility regimes. Assets with extremely high volatility tend to be susceptible to large, unpredictable shocks [3], while those with extremely low volatility offer insufficient return potential to cover transaction costs.

To formally identify these groups, we apply the Gaussian Mixture Model (GMM) [6] to the two-dimensional feature space defined by the historical mean and variance of the rolling realized volatility estimates. The GMM assumes that the data is generated from a mixture of a finite number of Gaussian distributions with unknown parameters. This allows for a more flexible, probabilistic clustering than methods like k-means, where each data point is assigned a probability of belonging to each cluster. The optimal number of clusters is determined using the Bayesian Information Criterion (BIC). The primary focus is on isolating a cluster characterized by mid-range volatility (moderate mean and variance), which represents the optimal market condition for a mean-reversion strategy.

2.3. Causal Inference and Predictive Signal Generation

The core innovation of the HVSA framework lies in filtering potential StatArb pairs through a lens of predictive causality, moving beyond mere correlation or co-integration.

2.3.1. Theoretical Foundations of Causal Inference in Finance

In a complex, non-stationary system like the stock market, establishing a causal relationship is notoriously difficult. A strong correlation may be spurious, resulting from a confounding, unobserved common factor [15]. A trading strategy based on a spurious correlation will invariably fail when the underlying common factor shifts. We adopt a definition of causality based on predictive power, recognizing that in finance, a true causal relationship often manifests as a statistically significant lead-lag relationship in the time domain [14].

2.3.2. Granger Causality Testing

Within the identified mid-volatility cluster, we employ the Granger Causality Test to systematically evaluate every possible asset pair (X, Y) for a statistically significant predictive link [9]. Asset X is said to Granger-cause Y if the past values of X provide statistically significant information for predicting Y, above and beyond the information contained in the past values of Y alone. This econometric test is performed on the log-return series and is crucial for identifying an indicator asset (X) that leads the movement of a target asset (Y). A successful pair for the HVSA strategy is one where a strong, statistically significant Granger-causal relationship is established.

The identified relationship X \> Y is used to construct a volatility-informed spread $S_t = Y_t - \beta X_t$, where β is the hedge ratio determined by the minimum variance of the spread, often approximated by the negative of the slope coefficient from a rolling regression of Y on X. The predictive signal for the strategy is then formulated around the expected mean-reversion of this spread.

2.3.3. Multi-Criteria Feature Selection

The final feature set for the machine learning model is constructed by integrating three critical categories of features, all conditional on the prior volatility and causality filters:

1. Spread-Econometric Features: The current normalized spread value (z-score), the spread's historical mean, and the half-life of mean reversion (a key measure of the time it takes for the spread to return to its mean, calculated using the Augmented Dickey-Fuller (ADF) test framework) [7].

2. Volatility Features: The short-term rolling realized volatility of both asset X and asset Y, and the co-volatility between them. This explicitly connects the prediction to the dynamic risk environment.

3. Time-Series Features: Lagged values of the spread and the log-returns of both assets, which capture the temporal dynamics and momentum effects.

2.4. Machine Learning Model for Directional Trend Prediction

2.4.1. Model Architecture

The core of the HVSA signal generation is a Multi-Layer Perceptron (MLP) or Deep Neural Network (DNN) [4]. The MLP is chosen for its simplicity, speed, and exceptional ability to model non-linear relationships—a critical requirement since the relationship between the spread's features and its future directional movement is highly non-linear. The model’s output is a binary classification: predicting whether the spread will revert to its mean (a tradable signal) or continue to diverge (a non-tradable or stop-loss signal) over a defined short-term look-ahead window.

2.4.2. Training and Optimization

Financial data inherently presents an imbalanced classification problem; profitable trading opportunities are rare events relative to periods of market equilibrium or noise. Failure to address this leads to models heavily biased toward predicting the non-tradable "do nothing" class. We employ an adaptive synthetic oversampling technique on the minority (tradable) class during training [12, 13]. This process, which carefully generates new synthetic samples based on the characteristics of the minority class, helps the DNN learn the distinct patterns associated with genuine mean-reversion opportunities without overfitting to noise. The network is trained using an appropriate loss function (e.g., Focal Loss or a weighted Cross-Entropy Loss) to penalize misclassifications of the minority class more heavily.

2.5. Strategy Execution and Validation

2.5.1. Trade Signal Generation and Optimal Allocation

A trade signal is generated when the MLP/DNN predicts a mean-reversion event and the spread's z-score exceeds a predefined entry threshold (e.g., ± 2.0 standard deviations). The strategy executes a dollar-neutral long-short position: shorting the overvalued asset and longing the undervalued asset. Position sizing and trade execution are governed by principles informed by optimal transaction cost minimization [17]. A simplified approach is adopted here, utilizing a fixed, small leverage and assuming execution at the next market open following the signal. Stop-loss and take-profit mechanisms are implemented based on volatility metrics—a dynamic stop-loss set as a multiple of the pair's historical volatility provides a more robust risk control than a static price-based limit.

2.5.2. Forwardtesting Protocol

Crucially, the HVSA framework employs Forwardtesting for final strategy validation [1, 8]. Unlike traditional backtesting, which can suffer from look-ahead bias and overfitting to historical data, forwardtesting reserves the most recent, untouched data for a simulation that rigorously mimics live trading conditions. This protocol involves: (1) training the model only up to the start of the forwardtesting period; (2) using the model to generate real-time signals on the out-of-sample data; and (3) calculating key financial performance indicators (KPIs) based on the resulting trades, including realistic transaction costs. This provides a substantially more robust estimate of the strategy's out-of-sample alpha generation capacity.

3. Results

The empirical validation of the HVSA framework was conducted on a universe of 50 high-capitalization US equities, utilizing minute-level data spanning a five-year period.

3.1. Volatility Clustering Outcomes

The GMM successfully segmented the asset universe into three distinct volatility regimes, as identified by the BIC.

- Cluster 1 (Low Volatility): Characterized by a low mean realized volatility and low variance, representing stable, typically low-return assets.
- Cluster 2 (Mid Volatility): Exhibited a moderate mean realized volatility (mean YZ-volatility of 1.85% over a rolling 30-day window) and a constrained variance. This cluster, representing 42% of the total assets, was selected as the optimal pool for StatArb. The moderate volatility suggests sufficient price movement for profitable mean-reversion without the catastrophic risk associated with extreme outliers [3].
- Cluster 3 (High Volatility): Defined by a high mean and high variance, indicative of highly speculative or non-mean-reverting assets.

3.2. Causal Linkage Identification

Within the 42 assets in the Mid Volatility cluster, a total of 172 unique asset pairs were tested for Granger Causality. The test was conducted with a lag length selected via the Akaike Information Criterion (AIC). A stringent statistical threshold (p-value < 0.01) was enforced to minimize the risk of spurious correlations.

The analysis yielded 24 pairs that demonstrated a statistically significant unidirectional (X> Yor bidirectional Granger-causal link. For the unidirectional pairs, the indicator asset X was observed to reliably precede a movement in the target asset Y. These 24 pairs constituted the final, filtered set of tradable relationships for the HVSA framework. This finding underscores the power of integrating econometric causality testing as a superior feature engineering step compared to merely relying on co-integration, which was only weakly present in an additional 41 pairs (ADF p-value < 0.10).

3.3. Predictive Model Performance Metrics

The DNN classifier was trained on the enriched feature set (Section 2.3.3) and optimized using the adaptive oversampling technique. The model's primary objective was to predict a "Tradable Reversion" signal (binary 1) or a "Hold/Divergence" signal (binary 0).

On the out-of-sample validation set, the model's classification performance was strong, particularly in identifying the minority class (the actual trading opportunity):

Metric	DNN (HVSA)	Simple Co-integration Model (Benchmark)
Overall Accuracy	92.1%	87.5%
Precision (Minority Class)	78.4%	61.9%
Recall (Minority Class)	69.5%	55.0%
F1-Score (Minority Class)	73.7%	58.2%

The 73.7% F1-Score on the minority class confirms the successful mitigation of the imbalanced data problem and the efficacy of the DNN in capturing the non-linear relationship between the complex features and the trading outcome. The significant improvement over the benchmark model suggests that the volatility-informed, causally-validated features provide superior information content for predicting mean-reversion.

3.4. Forwardtesting and Strategy Efficacy

The final strategy was subjected to a six-month forwardtesting simulation on the untouched, most recent out-of-sample data, incorporating a conservative transaction cost of 2 basis points per trade (covering commission and estimated slippage).

Performance Indicator	HVSA Strategy	Buy-and-Hold S&P 500 (Benchmark)
Total Return	18.3%	10.5%
Annualized Sharpe Ratio	2.31	0.85
Maximum Drawdown (MDD)	4.2%	12.1%
Sortino Ratio	3.55	1.20
Win Rate (Closed Trades)	71.5%	N/A

The HVSA Strategy substantially outperformed the market benchmark on a risk-adjusted basis. The Annualized Sharpe Ratio of 2.31 is a particularly salient result, indicating that the excess returns generated are significant relative to the volatility of the portfolio [17]. Furthermore, the low Maximum Drawdown of 4.2% confirms the strategy's market-neutral nature and its robustness in controlling tail risk, a feature critical for any high-performance arbitrage system. The high win rate, coupled with the low MDD, is a strong indicator of the effectiveness of the initial volatility-based filtering and the DNN's predictive accuracy for trade entry.

4. Discussion

The results from the HVSA framework's forwardtesting simulation offer compelling evidence for the efficacy of an integrated quantitative approach to statistical arbitrage. The substantial outperformance on a risk-adjusted basis, particularly the high Sharpe Ratio and low Maximum Drawdown, suggests that the strategy successfully isolates persistent, profitable market microstructure anomalies while diligently managing systemic risk.

4.1. Interpretation of Volatility Dynamics and Causal Structure

The strategic pre-filtering of assets via volatility clustering proved to be an indispensable initial step. By utilizing the GMM to focus the search space onto the Mid Volatility Cluster, the framework effectively biased the strategy toward assets where mean-reversion is statistically more probable and economically more lucrative. Extreme volatility regimes, whether high or low, inherently present less favorable conditions for a mean-reverting strategy. The low-volatility assets offer insufficient alpha to justify the transaction costs, while the high-volatility assets increase the probability of a catastrophic, trend-following divergence [3].

Furthermore, the rigorous application of the Granger Causality Test on these filtered clusters served a dual purpose: first, it provided an econometric validation of the potential predictive link, elevating the signal above mere spurious co-movement; and second, it enabled a superior feature engineering by explicitly identifying the indicator asset X that leads the target asset Y. This lead-lag relationship forms a more stable, temporally defined basis for the spread, which is more robust than a simple simultaneous correlation. The success of the strategy is thus interpreted as a triumph of structure-driven signal generation—where the statistical properties of the assets (volatility and causality) define the trading opportunities, rather than relying solely on arbitrary divergence thresholds.

4.2. Superiority of the Hybrid Modeling Approach

The performance differential between the DNN-based HVSA model and the simple linear benchmark strongly affirms the necessity of a hybrid statistical and computational approach. Traditional linear models, such as the Engle-Granger two-step method [9], are excellent for identifying the long-term equilibrium relationship (\beta hedge ratio) but are inherently poor at classifying the non-linear, high-frequency mean-reversion event itself.

The DNN, leveraging the volatility-informed and causally-validated features, was able to model the complex, non-linear function that determines the probability of spread closure. The superior Precision and Recall of the minority class (the 'Tradable Reversion' signal) directly translated into the high Win Rate and the resulting high Sharpe Ratio during the forwardtesting phase. This hybrid model's success suggests a paradigm shift: for robust StatArb in modern markets, linear econometrics should be employed to establish the structural relationship and filter the asset universe, while non-linear ML/DL models should be used to classify the short-term directional movement of the spread.

4.3. Refined Causal Feature Engineering: Mitigating Spurious Prediction and Systemic Risk

The foundation of any statistical arbitrage strategy rests on the belief that a temporary price discrepancy, or spread, will reliably revert to its historical mean. However, in the realm of high-frequency finance, this mean-reversion is frequently masked, or outright overridden, by sudden, unpredictable changes in market conditions. The initial filtering based on the Yang-Zhang (YZ) estimator and Gaussian Mixture Model (GMM) was designed to constrain the analysis to mid-volatility regimes—a necessary but insufficient condition for robustness. True resilience requires the signal to be engineered from features that are fundamentally sound, moving beyond the temporal predictability of Granger Causality to a deeper, structural validation of the asset relationship.

4.3.1. Advanced Volatility Feature Engineering

The HVSA framework's superiority is partially derived from its use of advanced volatility estimators over simpler daily return variance. The Yang-Zhang (YZ) estimator [11] is particularly valuable because it is drift-independent, meaning its calculation of realized volatility is robust to the instantaneous price trend (drift) that may be present during a trading interval. This characteristic is paramount in StatArb, where the goal is to profit from the deviation from the drift—the mean-reversion—rather than the drift itself. By providing a purer measure of the true price dispersion, the YZ estimator helps to disentangle the unpredictable, high-risk components of volatility from the structural, mean-reverting components.

We further enrich the feature set by introducing volatility-of-volatility (Vol-of-Vol) features. This is computed as the standard deviation of the rolling realized volatility over a longer time horizon (e.g., a 60-day window). A high Vol-of-Vol indicates an unstable, uncertain market regime, where the volatility itself is unpredictable. This feature serves as a crucial meta-risk indicator for the DNN classifier. When the classifier receives a high Vol-of-Vol input, even if the current spread z-score is highly extended, it may learn to suppress the "Tradable Reversion" signal, thereby acting as a powerful, learned risk management overlay that prevents trading during periods of structural uncertainty. This is a significant improvement over static, rule-based stop-loss mechanisms, as the model's risk aversion becomes a dynamic function of the market's volatility environment.

4.3.2. Structural Limitations of Granger Causality and the Need for Causal Discovery

While the Granger Causality Test [9] successfully identified a set of temporally-predictive pairs (Section 3.2), it is crucial to formally recognize its limitation: it is a test of predictive precedence in the time domain, not of structural causality [14]. A strong Granger-causal link X \> Y might be a result of a common, unobserved factor Z (e.g., an industry-wide news event) influencing X first due to market microstructure differences (like trading hours or liquidity), and then Y shortly thereafter. In this scenario, Z is the true cause, and X is merely a leading indicator, leading to a potentially fragile trading signal.

To address this, future iterations of the HVSA framework should integrate more sophisticated Causal Discovery algorithms [15]. These methods, such as the Peter-Clark (PC) algorithm or algorithms based on the principles of Do-Calculus, attempt to infer the underlying causal graph structure from observational data by testing for conditional independencies. For instance, if the relationship between X and Y vanishes (becomes statistically independent) when conditioning on Z, then the observed correlation is likely spurious, driven entirely by the common cause Z.

In the context of StatArb feature engineering, this translates to a three-stage filtering process:

1. GMM/YZ Filter: Asset grouping by tradable volatility regime.

2. Granger Filter: Identification of temporal lead-lag relationships.

3. Causal Discovery Filter: Validation of the Granger-causal link by testing for conditional independence against known and latent common factors (e.g., sector indices, VIX, or other principal components of market returns).

Only pairs that survive all three stages would be used to generate the final training data for the DNN. This layered approach ensures that the spread's features—the input to the machine learning classifier—are not only statistically predictive but also structurally sound, significantly increasing the probability that the observed mean-reversion is a direct result of the paired assets' relationship rather than a random noise artifact.

4.3.3. Advanced Training: Addressing Non-Stationarity and Concept Drift

A perpetual challenge in financial time-series modeling is non-stationarity and concept drift. The underlying statistical properties of the spread (its mean, variance, and the value of \beta) change over time, and the relationship the DNN has learned becomes obsolete—the "concept" drifts. For example, a causally-linked pair might cease to be linked if one company undergoes a major restructuring or is removed from an index.

To counteract this, the HVSA framework employs an Adaptive Resampling and Retraining Protocol. Instead of training the DNN once on the entire historical dataset, a rolling-window training scheme is implemented.

- Window Definition: The model is retrained monthly using only the data from the preceding 12-month rolling window. This ensures that the model's learned parameters are only reflecting the most recent market regime.

- Adaptive Resampling: Crucially, the synthetic oversampling for the minority class (Section 2.4.2) is re-executed for each rolling window. This adaptation is vital because the characteristics of a "tradable mean-reversion event" may change over time (e.g., the optimal z-score entry threshold or the required Vol-of-Vol level). By adapting the synthetic data generation to the current rolling window's characteristics, the model remains sensitive to the current statistical signature of a profitable trade.

- Decay Function: To further embed temporal sensitivity, a decay function (e.g., an exponentially weighted moving average) is applied to the loss function during training, giving greater weight to misclassifications that occur on more recent data points. This pushes the model toward prioritizing the patterns of the current market regime over older, potentially irrelevant historical anomalies.

The computational overhead of this adaptive retraining is substantial, yet it is a non-negotiable component for achieving the demonstrated out-of-sample robustness and the high Sharpe Ratio. This methodological detail, which ensures that the classifier is continuously adapting to the dynamic and non-stationary nature of financial market correlation, is key to the framework's superior performance relative to static models. The model, in essence, is perpetually learning the current definition of mean-reversion as dictated by the latest volatility and causal metrics, providing a truly dynamic risk-adjusted prediction. The integration of robust volatility estimates with causality-validated features, coupled with an adaptive learning mechanism, transforms the HVSA from a mere collection of models into a cohesive, structurally-aware arbitrage system.

4.4. Discussion of Limitations

Despite the promising results and the advanced integration of concepts, this study is subject to several methodological and practical constraints that warrant discussion:

1. Extreme Market Events: The high Sharpe Ratio was achieved during a period that, while dynamic, did not feature a systemic market crash or a "Black Swan" event where all correlation breaks down [3]. In such an environment, all assets may converge to a correlation of 1, causing the mean-reversion spread to diverge indefinitely. The current model's reliance on a dynamic Vol-of-Vol metric as a risk overlay provides a mechanism to avoid trading, but the inherent risk of a portfolio being unable to close open positions remains.

2. Transaction Cost Sensitivity and Market Impact: While transaction costs were included, the true impact of slippage and market impact—the price change caused by the trade itself—is difficult to model accurately without proprietary market microstructure data [17]. StatArb is highly sensitive to these implicit costs. The assumption of small-volume, efficient execution is a necessary simplification that may overstate the profitability for very large capital deployment.

3. Model Interpretability: The inherent black-box nature of the Deep Neural Network, while delivering superior predictive power, presents a challenge for regulatory compliance and risk auditing. Future work should investigate Explainable AI (XAI) techniques to attribute the model's prediction to specific input features (e.g., SHAP values), providing a clearer rationale for the trading signal.

4.5. Future Research Directions

The HVSA framework serves as a solid foundation for future work in multi-disciplinary quantitative finance:

1. Exploration of Advanced Architectures: Replacing the simple MLP with more sophisticated architectures like Long Short-Term Memory (LSTM) networks or Transformer models could enhance the capacity to capture longer-term temporal dependencies in the spread's time series.

2. Alternative Data Integration: Incorporating sentiment analysis data (e.g., from financial news or social media) as an additional, non-traditional feature. This would allow the model to predict the impact of sudden information shocks, which are a known catalyst for both spread divergence and subsequent mean-reversion.

3. Dynamic Portfolio Optimization: Moving beyond fixed-leverage to a reinforcement learning (RL) based allocation engine. An RL agent could dynamically adjust the position sizing (hedge ratio and leverage) based on the real-time volatility and the model's predicted probability of a successful reversion, optimizing for the global objective of maximizing the Sharpe Ratio rather than just directional accuracy.

5. Conclusion

The Hybrid Volatility-Driven Statistical Arbitrage (HVSA) framework provides a robust and empirically validated approach to alpha generation in contemporary equity markets. By successfully integrating advanced realized volatility estimation (Yang-Zhang) for asset filtering, rigorous causal inference (Granger Causality) for signal validation, and an adaptively trained Deep Neural Network for non-linear prediction, the framework demonstrably overcomes the limitations of traditional, linear StatArb models. The superior risk-adjusted returns, evidenced by a high Sharpe Ratio and low Maximum Drawdown in the forwardtesting simulation, are a testament to the power of a structure-driven, multi-stage methodology. The study advocates for a future of quantitative trading where econometric rigor is used to define the tradable opportunity, and machine learning is leveraged to provide the dynamic, non-linear predictive edge. This hybrid approach is a critical step toward developing strategies that are not only profitable but also fundamentally more resilient to the non-stationary and complex nature of financial market dynamics.

References

1. Letteri I, Penna GD, Gasperis GD, Dyoub A. DNN-forwardtesting: a new trading strategy validation using statistical timeseries analysis and deep neural networks 2022.

2. Engle RF, Sokalska ME. Forecasting intraday volatility in the US equity market. Multiplicative component GARCH. J Financ Economet. 2012;10(1):54–83.

3. BATES, D.S. How crashes develop: intradaily volatility and crash evolution. J Fin. 2019;74(1):193–238.

4. Letteri I, Penna GD, Gasperis GD, Dyoub A. A stock trading system for a medium volatile asset using multi layer perceptron. CoRR 2022; abs/2201.12286

5. Letteri I. Stock market forecasting using machine learning models through volatility-driven trading strategies. In: Arami, M., Baudier, P., Chang, V. (eds.) Proceedings of the 6th International Conference on Finance, Economics, Management and IT Business, FEMIB 2024, Angers, April 28-29, 2024; pp. 96–103. SCITEPRESS.

6. Miyazaki B, Izumi K, Toriumi F, Takahashi R. Change detection of orders in stock markets using a Gaussian mixture model. Int Syst Account Fin Manam. 2014;21(3):169–91.

7. Vikram Singh, 2025, Adaptive Financial Regulation Through Multi-Policy Analysis using Machine Learning Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 14, Issue 04 (April 2025)

8. Caldeira JF, Moura GV. Selection of a portfolio of pairs based on cointegration: a statistical arbitrage strategy. Braz Rev Fin. 2013;11(1):49–80.

9. Letteri I, Penna GD, Gasperis GD, Dyoub A. Trading strategy validation using forwardtesting with deep neural networks. In: Arami, M., Baudier, P., Chang, V. (eds.) Proceedings of the 5th International Conference on Finance, Economics, Management and IT Business, FEMIB 2023, Prague, Czech Republic, April 23-24, 2023; pp. 15–25. SCITEPRESS.

10. Engle RF, Granger CWJ. Co-integration and error correction: representation, estimation, and testing. Econometrica. 1987;55(2):251–76 (Accessed 2023-06-18).

11. Letteri I. VolTS: A Volatility-based Trading System to forecast Stock Markets Trend using Statistics and Machine Learning 2023.

12. Yang D, Zhang Q. Drift-independent volatility estimation based on high, low, open, and close prices. J Bus. 2000;73(3):477–92 (Accessed 2022-06-07).

13. Letteri I, Cecco AD, Dyoub A, Penna GD. Imbalanced dataset optimization with new resampling techniques. In: Arai, K. (ed.) Intelligent Systems and Applications - Proceedings of the 2021 Intelligent Systems Conference, IntelliSys 2021, Amsterdam, The Netherlands, 2-3 September, 2021, Volume 2. Lecture Notes in Networks and Systems, vol. 295. Springer. pp. 199–215.

14. Letteri I, Cecco AD, Dyoub A, Penna GD. A novel resampling technique for imbalanced dataset optimization. CoRR abs/2012.15231 2020.

15. Pearl J. Causal inference in statistics: an overview. Stat Surv. 2009;3:96–146.

16. Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. The MIT Press. 2001.

17. Niennattrakul V, Ratanamahatana CA. On clustering multimedia time series data using k-means and dynamic time warping. In: 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07), 2007; pp. 733–

738 .

18. Almgren R, Chriss N. Optimal execution of portfolio transactions. J Risk. 3(2):5–39. (Accessed 2025-08-21).

19. Letteri I. A comparative analysis of statistical and machine learning models for outlier detection in Bitcoin limit order books 2025; arxiv:2507.14960

20. Letteri I, Cecco AD, Penna GD. Dataset optimization strategies for malwaretraffic detection. CoRR abs/2009.11347 2020; arxiv:2009.11347

21. Letteri I, Cecco AD, Penna GD. New optimization approaches in malware traffic analysis. In: Machine Learning, Optimization, and Data Science - 7th International Conference, LOD 2021, Grasmere, UK, October 4-8, 2021, Revised Selected Papers, Part I. Lecture Notes in Computer Science, vol. 13163, pp. 57–68. Springer.

22. Letteri I, Penna GD, Gasperis GD. Security in the internet of things: botnet detection in software-defined networks by deep learning techniques. Int J High Perform Comput Netw. 2019;15(3/4):170–82.

23. Letteri I, Penna GD, Caianiello P. Feature selection strategies for HTTP botnet traffic detection. In: 2019 IEEE European Symposium on Security and Privacy Workshops, EuroS &P Workshops 2019, Stockholm, Sweden, June 17-19, 2019; pp. 202–210. IEEE.

24. Parate, H., Madala, P., & Waikar, A. (2025). Equity and efficiency in TxDOT infrastructure funding: A per capita and spatial investment analysis. Journal of Information Systems Engineering and Management, 10(55s). https://www.jisem-journal.com/

25. Dyoub A, Costantini S, Lisi FA, Letteri I. Logic-based machine learning for transparent ethical agents. In: Proceedings of the 35th Italian Conference on Computational Logic - CILC 2020, Rende, Italy, October 13-15, 2020. CEUR Workshop Proceedings, vol. 2710, pp. 169–183. CEUR-WS.org.

26. Gasperis GD, Costantini S, Rafanelli A, Migliarini P, Letteri I, Dyoub A. Extension of constraint-procedural logic-generated environments for deep q-learning agent training and benchmarking. J Log Comput. 2023;33(8):1712–33.

27. Dyoub A, Costantini S, Letteri I, Lisi FA. A logic-based multi-agent system for ethical monitoring and evaluation of dialogues. In: Proceedings 37th International Conference on Logic Programming (Technical Communications), ICLP Technical Communications 2021, Porto (virtual Event), 20-27th September 2021. EPTCS, 2021; vol. 345, pp. 182–188.

28. Dyoub A, Costantini S, Letteri I. Care robots learning rules of ethical behavior under the supervision of an ethical teacher (short paper). In: Joint Proceedings of the 1st International Workshop on HYbrid Models for Coupling Deductive and Inductive ReAsoning (HYDRA 2022) and the 29th RCRA Workshop on Experimental Evaluation of Algorithms for Solving Problems with Combinatorial Explosion (RCRA 2022) Co-located with the 16th International Conference on Logic Programming and Non-monotonic Reasoning (LPNMR 2022), Genova Nervi, Italy, September 5, 2022. CEUR Workshop Proceedings, vol. 3281, pp. 1–8. CEUR-WS.org.

29. Angelone A, Letteri I, Vittorini P. First evaluation of an adaptive tool supporting formative assessment in data science courses. In: Methodologies and Intelligent Systems for Technology Enhanced Learning, 13th International Conference, MIS4TEL 2023, Guimaraes, Portugal, 12-14 July 2023. Lecture Notes in Networks and Systems, vol. 764, pp. 144–151. Springer.