

Comparative Performance of ML-Enhanced Statistical Arbitrage Across the S&P 500 and NIFTY 500

Sankar Ganesh M, Dr. S. Johnsi

Firebird Institute of Research in Management, Coimbatore
sankarganesh.m@firebird.net.in, johnsi.s@firebird.ac.in

Abstract-

This study proposes and empirically validates an advanced algorithmic trading framework for statistical arbitrage using a pairs-trading strategy enhanced with unsupervised machine learning. The framework is tested across four market cohorts: NIFTY 500 Long (2007–2025), NIFTY 500 Broad (2015–2025), S&P 500 Long (2007–2025), and S&P 500 Broad (2015–2025). Principal Component Analysis (PCA) is applied to neutralize systematic risk factors and isolate idiosyncratic residual returns, after which the OPTICS (Ordering Points To Identify the Clustering Structure) algorithm performs density-based clustering to group assets with similar residual behaviour without pre-specifying the number of clusters. The presence of a tradeable long-run relationship is confirmed through the Engle-Granger two-step cointegration test. The backtesting engine incorporates institutional-level constraints, including a 15-basis-point round-trip transaction cost and a hard –5% stop-loss, and benchmarks performance against sovereign hurdle rates of 7.2% (India) and 4.2% (United States). The results show substantial alpha generation in the NIFTY 500 Long cohort, with a net ROI of 485.3% under the traditional benchmark, while the machine-learning-enhanced framework delivers a superior risk-adjusted outcome in the NIFTY 500 Broad cohort (Sharpe ratio of 1.090 versus 0.869) alongside a materially reduced maximum drawdown. In the S&P 500, the ML-enhanced framework improves the Sharpe ratio from 0.440 to 0.583 (Long) and 0.424 to 0.479 (Broad), though absolute magnitudes remain smaller, consistent with a more factor-cohesive and efficiently arbitrated developed market. Statistical significance tests, including a one-sided t-test, the Jobson-Korkie test, and bootstrap resampling, confirm that these performance gains are not attributable to random chance. The findings indicate that while market efficiency compresses arbitrage windows, the PCA-OPTICS framework is effective at isolating economically meaningful, cost-adjusted idiosyncratic opportunities in both emerging and developed equity markets.

Keywords-Statistical Arbitrage; Pairs Trading; Principal Component Analysis; OPTICS Clustering; Cointegration; Machine Learning; NIFTY 500; S&P 500

I. Introduction

A. Introduction

Statistical arbitrage is a quantitative, market-neutral investment framework designed to identify and exploit persistent, though often transient, deviations from established pricing equilibria among related financial instruments, based on the expectation of mean reversion. Unlike pure arbitrage, it accepts that price discrepancies may widen before converging, and its effectiveness depends on distinguishing structural price change from stochastic noise - a task complicated by the shift of global markets toward

electronic, algorithmically driven structures. This study compares the viability of an ML-enhanced statistical arbitrage strategy across two structurally distinct markets: the S&P 500, characterized by deep liquidity and rapid informational efficiency, and the NIFTY 500, which despite India's market maturation retains features of an emerging market - greater volatility sensitivity, thinner institutional participation in mid- and small-cap segments, and structural inefficiencies that may offer more fertile ground for arbitrage. The study benchmarks viability against the localized cost of capital: a 7.2% sovereign hurdle in India versus a leaner 4.2% hurdle in the United States, and complements classical cointegration-based

arbitrage with Principal Component Analysis (PCA) for factor neutralization and OPTICS density-based clustering for resilient pair formation, while using SHAP (SHapley Additive Explanations) to keep trading signals interpretable rather than a black-box artefact of overfitting.

B. Background of the Study

The theoretical roots of statistical arbitrage trace to the mid-1980s Automated Proprietary Trading desk at Morgan Stanley, where a team led by Nunzio Tartaglia applied signal processing to identify pairs of historically correlated stocks and trade their reversion to the mean - the earliest systematic monetization of idiosyncratic mean reversion. Through the 1990s and early 2000s, researchers such as Avellaneda and Lee extended this to multi-factor, PCA-based models that traded a stock against a synthetic portfolio representing its latent risk factors rather than a single partner. The period 2007–2025 offers a natural comparative laboratory: the U.S. economy passed through the 2008 financial crisis, a decade of near-zero rates, and a 2020s inflationary cycle, depressing the cost of capital for arbitrageurs, while India's persistently higher policy rate regime kept its 10-year government bond yield closer to 6.5–8%, a very different risk-free benchmark.

This yield spread is not incidental: a strategy returning 6% annually would be a success in New York and a capital-destructive failure in Mumbai, compelling the Indian arbitrageur toward deeper mispricings or greater leverage, either of which raises model risk. Market microstructure compounds this: the S&P 500 is uniformly liquid, while the NIFTY 500's bottom 450 constituents across 72 sectors are less liquid, with wider bid-ask spreads and less efficient information flow - the classic precursors of arbitrage opportunity, but also of the roughly 15 basis point per-leg transaction cost that erodes them.

C. Problem Statement

The central problem in contemporary quantitative finance is the ongoing decline of alpha in the face of rising market efficiency and cost of capital. Institutional managers judge strategies net of all costs relative to risk, and the classical statistical arbitrage model carries systemic weaknesses that intensify when a developed

market is compared against an emerging one. First, an efficiency-hurdle gap emerges: the S&P 500's hyper-efficient, high-frequency environment leaves spreads too thin to clear even a 4.2% hurdle net of 15 bps costs, while the NIFTY 500 offers more mispricing but demands clearance of a much higher 7.2% hurdle - a profitability trap in which the market with more inefficiency has the highest cost of capital. Second, survivorship bias remains widespread in backtesting: using only current index constituents excludes delisted or bankrupt firms and can inflate annual returns by an estimated 1.0% to 4.94%, understate drawdown by roughly 14 percentage points, and overstate Sharpe ratios by about 0.5 points, with as much as 82.5% of the small-cap universe removed from India's index over nine years. Third, the growing opacity of machine-learning models creates a "black box" trust gap for regulators and investment committees who require an economic rationale for automated trading signals. Fourth, statistical arbitrage strategies are notoriously fragile in systemic shocks, when correlations spike toward one and meanreversion relationships break down, making a disciplined stop-loss essential without being so tight that it exits positions on ordinary volatility. Finally, most academic studies treat transaction costs as negligible, whereas a realistic 15 bps round-trip cost can erode over half of an arbitrage strategy's excess return, particularly in the broader, less liquid NIFTY 500 universe.

D. Research Questions and Objectives

The study is structured around five research questions: (1) Can a machine-learning-enhanced statistical arbitrage framework using PCA-based residual extraction and OPTICS clustering be developed and evaluated across four market universes? (2) Can such a strategy produce returns in excess of the respective 7.2% and 4.2% hurdle rates, net of 15 bps transaction costs? (3) To what extent does survivorship bias overstate Sharpe ratios and annual returns in each market? (4) What is the impact of a hard -5% stop-loss on the long-term equity curve, and does it adequately reduce tail risk without unnecessarily impairing performance? (5) Using SHAP, what underlying characteristics most influence successful arbitrage trades, and how do these differ between developing and developed markets?

Correspondingly, the study's objectives are to construct a multi-stage ML arbitrage pipeline combining PCA-based factor extraction with OPTICS clustering; to conduct a comparative efficiency analysis of meanreversion speed and arbitrage duration between the Indian and U.S. markets; to implement a survivorship-bias-free backtesting engine spanning 2007–2025; to validate strategy performance against the 7.2% and 4.2% institutional hurdles; and to quantify the efficacy of the –5% stop-loss risk control across multiple market conditions.

E. Scope and Contribution of the Study

The study is confined to two major equity indices - the NIFTY 500 as a developing-market benchmark and the S&P 500 as a developed-market benchmark - covering 2007 to 2025 and segmented into Long (2007–2025) and Broad (2015–2025) cohorts for each market to capture performance across differing macroeconomic and liquidity regimes. Methodologically, the machinelearning component is limited to unsupervised learning (PCA and OPTICS) supplemented by SHAP-based explainability, and does not extend to supervised price forecasting. Financially, the backtest is conducted net of a 15 bps round-trip transaction cost and a –5% hard stop-loss, benchmarked against country-specific hurdle rates to compute an Excess Sharpe Ratio grounded in professional investment practice rather than theoretical gross performance. The study contributes the first large-scale, index-wide comparison of PCA-OPTICS arbitrage between a developing market with a higher cost of capital and a developed market with higher informational efficiency; demonstrates the relative advantage of density-based OPTICS clustering over centroid-based methods such as K-means in financial applications; operationalizes a 5% stop-loss and SHAP-based interpretability standard for responsible quantitative trading; and provides quantified evidence (a 4.94% return-inflation estimate) of the practical cost of survivorship bias in backtesting.

II. Literature Review

A. Overview

This review examines the theoretical and empirical foundations for a systematic statistical arbitrage model, beginning with theories of market efficiency that explain why mispricing can arise and persist, followed by the historical evolution of pairs trading and its erosion under increasingly competitive markets. It then reviews advances in factor neutralization through PCA and systematic stock grouping through density-based clustering, a comparative assessment of structural differences between the S&P 500 and NIFTY 500, and the practical considerations of transaction costs, capital hurdles, and survivorship bias that shape real-world arbitrage viability. The chapter closes by identifying gaps that motivate the present study's empirical design.

B. Theoretical Framework: Market Efficiency and Adaptation

The Efficient Market Hypothesis (EMH), developed independently by Fama and Samuelson in the 1960s, holds that market prices fully and immediately incorporate available information, implying that no investor can consistently earn risk-adjusted returns above the market average and, in its strict form, that statistical arbitrage should not be possible. Persistent anomalies, bubbles, and crashes have drawn sustained criticism of the EMH's assumption of investor rationality. Andrew Lo's (2004) Adaptive Market Hypothesis (AMH) reframes market efficiency through an evolutionary lens: investors behave as competing species adapting to a changing environment, so that efficiency is not constant but cyclical, admitting transient inefficiencies when the environment shifts faster than participants can adapt. This offers a theoretical rationale for statistical arbitrageurs as opportunistic exploiters of short-lived mispricing. The theory of limits to arbitrage (Shleifer & Vishny, 1997) further explains why even clear mispricing may go uncorrected - capital constraints, risk aversion, and short-selling restrictions can leave markets in a persistent state of partial disequilibrium, particularly in less mature indices such as the NIFTY 500.

C. Evolution of Statistical Arbitrage: Distance and Cointegration Methods

Gatev, Goetzmann, and Rouwenhorst (2006) provided the first rigorous academic documentation of pairs trading, matching stocks by minimum sum-of-squared deviations in normalized price series - the nonparametric "distance method" - and reporting excess returns of up to 11% annually on self-financing portfolios. However, the distance method assumes a constant price-level relationship between two stocks and offers no information on expected holding period or convergence probability, and its profitability has declined as the strategy became more widely known, consistent with the AMH's prediction of competed-away alpha. The Engle-Granger cointegration procedure addressed some of these weaknesses by testing for a genuine long-run equilibrium relationship rather than mere historical proximity, offering a stronger statistical basis for arbitrage, though as a bivariate method it does not scale efficiently to markets with hundreds of candidate stocks - motivating multivariate factor models and dimensionality reduction.

D. Factor Neutralization Using PCA

Avellaneda and Lee (2010) addressed the multivariate challenge by applying Principal Component Analysis to decompose stock returns into a systematic component driven by market, sector, and macroeconomic forces, and a stock-specific idiosyncratic residual. Arbitrage is then pursued on the residual, isolating stock-level mispricing while offsetting broad market exposure. Their results showed that PCA-based strategies achieved strong risk-adjusted performance in the S&P 500 during the late 1990s and early 2000s but weakened as competition increased, confirming both the value of factor neutralization and the absence of a permanent statistical edge in competitive markets. A key limitation is that PCA-extracted factors are mathematical constructs that do not map cleanly onto interpretable economic themes, complicating governance reporting, and the literature offers limited evidence on the temporal stability of these factors across changing market regimes - motivating the addition of a clustering layer.

E. Density-Based Clustering and OPTICS

While PCA removes broad market forces, it does not indicate which stocks should be paired or grouped, and traditional sector classifications are often too coarse to capture the true economic linkages driving comovement. The OPTICS algorithm (Ankerst et al., 1999) addresses this by discovering natural, density-based clusters directly from the data, without prespecifying the number of clusters and without assuming uniform density across the market - allowing tightly correlated sectors (e.g., utilities) and more diffuse ones (e.g., technology) to coexist in the same model. Unlike centroid-based methods such as K-means, OPTICS explicitly identifies noise points - stocks that do not belong to a structurally consistent group - and excludes them from candidate pair formation, reducing the risk of trading on spurious or transient correlations. Applied studies, including 2024 research from the University of Malta, found that combining PCA-based factor removal with density-based clustering produced more robust and consistent performance than PCA alone, and that the clustering-enhanced model remained profitable in the unstable post-COVID period when simpler residual-based models failed.

F. Comparative Market Dynamics: NIFTY 500 vs. S&P 500

The S&P 500 and NIFTY 500 represent markedly different stages of financial development. The S&P 500, at approximately USD 46 trillion in market size, exhibits extremely high liquidity, tight bid-ask spreads, and volatility that tends to persist once elevated as institutional investors gradually reprice risk. The NIFTY 500, at roughly USD 4 trillion, is more reactive to global and domestic shocks, with liquidity that is deep in large caps but variable across its mid- and small-cap constituents, and is comparatively more exposed to foreign institutional flows. These differences carry direct implications for arbitrage design: sharper but less predictable reversals in India favor faster execution and tighter risk controls, while the more gradual spread development in the U.S. favors precision and cost efficiency over raw magnitude.

G. Institutional Constraints, Hurdle Rates, and Backtesting Biases

A statistical arbitrage strategy is only institutionally viable if it clears the local cost of capital net of transaction costs. In 2025, U.S. market-neutral strategies are typically benchmarked against a 4.2% hurdle, while Indian strategies face a 7.2% hurdle reflecting higher growth and inflation expectations; a 5% gross return would be a success in the U.S. and a capital-destructive failure in India. Transaction costs differ structurally as well: India's Securities Transaction Tax, stamp duty, and higher impact costs contrast with SEC and FINRA fee structures in the U.S. Survivorship bias remains the most consequential backtesting trap, with research on the NIFTY Smallcap 250 finding annual returns artificially inflated by as much as 4.94 percentage points through survivor-only backtesting, and Sharpe ratios in the S&P 500 overstated by up to 0.5 points - underscoring the importance of point-in-time, survivorship-bias-free datasets, particularly given the substantially higher historical turnover (up to 82.5% over a decade) among NIFTY 500 constituents.

H. Research Gap and Hypotheses Development

Despite substantial individual-technique literature on PCA, clustering, and pairs trading, several gaps remain. No study has directly compared a PCA-enhanced OPTICS framework across a mature market (S&P 500) and a large emerging market (NIFTY 500) using a single, consistent methodology. The use of OPTICS as pre-processing for factor-based arbitrage remains thinly evidenced at the scale of a full equity index, and the interaction between survivorship bias, transaction costs, and divergent sovereign hurdle rates (7.2% versus 4.2%) has not been incorporated into a comparative performance study. Nor has the temporal stability of OPTICS-identified clusters been examined across developing and developed markets. These gaps motivate three hypotheses:

H1: ML-enhanced, density-based clustering (OPTICS) produces a lower proportion of unstable or nonintegrated pairs than traditional centroid-based clustering (e.g., K-means), yielding more stable longrun relationships.

H2: The NIFTY 500 (developing market) generates higher gross returns than the S&P 500 (developed market), but this advantage is not proportionally sustained in risk-adjusted terms once transaction costs and sovereign hurdle rates are applied.

H3: Strategy performance differs significantly between the Long (2007–2025) and Broad (2015–2025) regimes, reflected in mean-reversion speed, trade duration, and drawdown behaviour.

III. Research Methodology

A. Introduction

This chapter outlines the systematic, multi-stage quantitative process used to identify, validate, and evaluate statistical arbitrage opportunities in the S&P 500 and NIFTY 500. The design follows progressive validation: at each stage, a statistical or econometric filter reduces noise and spuriousness before a trading signal is generated, and the strategy is tested across four research universes segmented by market type and time period to assess robustness across differing liquidity regimes and economic conditions.

B. Conceptual Framework (The Strategy Pipeline)

The pipeline comprises six stages. Stage 1 (data construction) segments the data into four cohorts to enable evaluation across differing market conditions while mitigating survivorship and selection bias. Stage 2 (signal engineering) applies Principal Component Analysis to neutralize systematic market exposure and isolate idiosyncratic residual returns. Stage 3 (asset selection) applies OPTICS density-based clustering to the residuals to exclude unstable or irregular assets. Stage 4 (econometric validation) applies the EngleGranger cointegration test to verify long-term mean reversion. Stage 5 (backtesting) simulates execution under a 15 bps institutional cost assumption. Stage 6 (model transparency) applies SHAP-based interpretation to support managerial oversight.

C. Research Variables

The dependent variable is overall strategy performance, assessed through Net Strategy Return (ROI net of 15 bps transaction costs), Excess Sharpe Ratio (riskadjusted return above the 7.2% or 4.2% sovereign hurdle), and

Maximum Drawdown (the largest peak-to-trough decline, used to assess the adequacy of the –5% stop-loss). The independent variables are Market Type (S&P 500 versus NIFTY 500, capturing differences in liquidity, regulation, and institutional participation) and Time Cohort (the 2007–2025 Long cohort, capturing systemic crises and macroeconomic turmoil, versus the 2015–2025 Broad cohort, capturing the contemporary, algorithmically driven trading environment). Fixed institutional control factors - 15 bps transaction cost per side, a –5% stop-loss per trade, a 7.2% India hurdle rate, a 4.2% USA hurdle rate, and a universe restricted to the top 500 equities per market - are applied uniformly to ensure that backtest outcomes are representative of real-world investment conditions.

D. Source of Data and Universe Construction

S&P 500 price data are sourced from publicly available financial databases with adjustments for dividends, splits, and bonus issues; NIFTY 500 data are drawn from the National Stock Exchange of India and comparable financial databases. All price series are converted to log returns for time-additivity and variance stabilization. Four research universes are constructed: S&P 500 Long (2007–2025, 394 stocks) and Broad (2015–2025, 467 stocks), and NIFTY 500 Long (2007–2025, 295 stocks) and Broad (2015–2025, 448 stocks). To mitigate survivorship bias, index constituent lists are reconstructed at the start of each cohort period; delisted securities remain in the dataset until their last trading date, with the terminal price treated as the exit value and capital reallocated accordingly, so that the backtest reflects the investable universe actually available to market participants at each point in time.

E. Signal Engineering: PCA Factor Neutralization

PCA decomposes standardized stock log returns into orthogonal components: $R_{i,t} = \sum \beta_{ij} F_{j,t} + \epsilon_{i,t}$, where $F_{j,t}$ are principal components, β_{ij} are factor loadings, and $\epsilon_{i,t}$ is the idiosyncratic residual. Empirical validation confirms PC1 as the dominant market factor in both universes: in the S&P 500 Long cohort, PC1 explains 39.40% of cross-sectional variance and correlates 0.9984 with an equal-weighted market proxy ($R^2 = 0.9968$); in the NIFTY 500 Long cohort, PC1 explains 23.06% of variance with a 0.9977 correlation ($R^2 =$

0.9954). By removing the leading ten principal components, the model isolates residual returns for use in clustering and cointegration testing, ensuring the strategy targets relative-value opportunities rather than directional market exposure.

F. Algorithmic Clustering: OPTICS

Because K-means and other centroid-based methods assume spherical, evenly distributed clusters that poorly match the heterogeneous, variable-density structure of financial returns, the study applies OPTICS to the PCA-neutralized residual space using a correlation-derived distance metric, $d_{ij} = \sqrt{(2(1 - \rho_{ij}))}$, where ρ_{ij} is the residual correlation between assets i

and j . Observations that do not fit a stable density region are classified as noise (Cluster –1) and excluded from the tradable universe, reducing the risk of building pairs on unstable relationships.

G. Econometric Validation: Cointegration

Intra-cluster candidate pairs are tested for cointegration via the Engle-Granger two-step procedure: a static hedge ratio is estimated by OLS regression ($PA_{i,t} = \alpha + \beta PB_{i,t} + \epsilon_t$), and the residual spread ϵ_t is tested for stationarity using the Augmented Dickey-Fuller test, with the unit-root null rejected at the 5% significance level ($p < 0.05$) required for retention. Only pairs meeting this criterion are advanced to backtesting, ensuring that trades are grounded in genuine long-run equilibrium rather than short-term co-movement.

H. Backtesting Framework and Institutional Constraints

Entry and exit signals are generated using an S-score mechanism, $S_t = (\text{Spread}_t - \mu) / \sigma$, where μ and σ are the historical mean and standard deviation of the spread. A long-short position is opened when the Sscore crosses a specified entry threshold (e.g., ± 2.0) and closed when it returns to zero, indicating convergence. Each backtest incorporates a 15 bps perside transaction cost covering brokerage, bid-ask impact, slippage, and applicable taxes, and is benchmarked against the 7.2% (India, 10-year G-Sec) or 4.2% (USA, 10-year Treasury) hurdle rate. A hard – 5% stop-loss per trade is enforced regardless of the Sscore signal to limit downside

exposure in the event of a structural breakdown in the cointegrating relationship.

I. Model Governance and Interpretability

To address the black-box trust gap facing institutional and regulatory stakeholders, the study applies SHAP (SHapley Additive Explanations) to decompose the contribution of individual features - such as residual volatility, momentum, and PCA loadings - to each trading signal, supporting model-drift detection and fairness checks consistent with emerging AI governance frameworks. Performance evaluation is further supported by the Omega Ratio, which captures asymmetry in the distribution of gains and losses, and by an information-decay measure that tracks the erosion of net alpha by transaction costs.

Component	S&P 500 Variance (%)	NIFTY 500 Variance (%)	Interpretation
PC1	39.40	23.06	Stronger common factor dominance in U.S.
PC2	~7.5	~5.8	Sector-level thematic influences
PC3	~4.2	~3.5	Secondary systematic risks
Cumulative (PC1–PC10)	~65.0	~52.0	Total systematic risk captured

IV. Data Analysis and Results

This section presents the empirical results of the PCAOPTICS statistical arbitrage framework across the

TABLE I PCA VARIANCE DECOMPOSITION ACROSS MARKETS (LONG COHORT)

B. Clustering and Pair Formation

Table II summarizes OPTICS clustering outcomes. The S&P 500 exhibits substantially greater cluster density (82 clusters in the Broad cohort versus 26 in the NIFTY 500 Broad cohort), indicating a more structurally segmented, sub-sector-differentiated market, while the

four research universes, beginning with structural validation and proceeding through clustering, cointegration, backtesting, comparative analysis, and formal hypothesis testing.

A. Sample Construction and PCA Structural Validation

Table I summarizes the variance decomposition for the Long cohort of each market. PC1 explains 39.40% of cross-sectional variance in the S&P 500 versus 23.06% in the NIFTY 500, and the top ten components cumulatively explain approximately 65.0% of variance in the S&P 500 against approximately 52.0% in the NIFTY 500 - confirming that systematic risk factors are structurally more concentrated in the developed U.S. market, while the Indian market exhibits greater

crosssectional dispersion and correspondingly greater scope for idiosyncratic residual alpha.

NIFTY 500 shows a proportionally higher share of noise assets, consistent with greater idiosyncratic dispersion. Density-based clusters demonstrated approximately 18% higher membership persistence over rolling windows than pairs formed via traditional distance-based metrics.

TABLE II OPTICS CLUSTERING SUMMARY

Market Universe	Cohort	Identified Clusters	Stocks Grouped	Noise Stocks (Cluster -1)
NIFTY 500	Long	24	66	229
NIFTY 500	Broad	26	76	252
S&P 500	Long	65	178	216
S&P 500	Broad	82	~220	~247

C. Cointegration Testing Results

the PCA-OPTICS-enhanced strategy across all four cohorts, reporting the single best-performing

Cohort	Intra-Cluster Pairs Scanned	% Cointegrated ($p < 0.05$)	Avg. ADF Statistic	Mean Half-Life (Days)
S&P Long	~2,500	19.7	-3.921	14.5
S&P Broad	~3,000	21.2	-4.015	11.2

Table III presents cointegration density across cohorts. Developed-market (S&P 500) cohorts show materially

cointegrated pair per cohort net of 15 bps transaction costs, a -5% stop-loss, and a +15% take-profit rule. In

TABLE III COINTEGRATION TEST SUMMARY ACROSS COHORTS

Cohort	Intra-Cluster Pairs Scanned	% Cointegrated ($p < 0.05$)	Avg. ADF Statistic	Mean Half-Life (Days)
NIFTY Long	~1,000	12.4	-3.459	24.2
NIFTY Broad	~1,200	13.1	-3.512	19.8

TABLE IV NET-OF-COST PERFORMANCE: TRADITIONAL VS. ML-ENHANCED STRATEGY

Market	Cohort	Strategy	Pair	ROI (%)	Sharpe	Max DD (%)
NIFTY 500	Long	Traditional	ACE.NS-ELECON.NS	485.35	0.913	-31.60
NIFTY 500	Long	ML-Enhanced	AXISBANK.NS-HDFCBANK.NS	8.61	0.367	-31.10
NIFTY 500	Broad	Traditional	ACE.NS-ELECON.NS	26.41	0.869	-31.22
NIFTY 500	Broad	ML-Enhanced	INDHOTEL.NS-EIHOTEL.NS	19.65	1.090	-21.27
S&P 500	Long	Traditional	A-PLD	6.96	0.440	-36.15
S&P 500	Long	ML-Enhanced	DG-DLTR	7.99	0.583	-45.66
S&P 500	Broad	Traditional	ADI-DOV	1.99	0.424	-23.95
S&P 500	Broad	ML-Enhanced	DG-DLTR	2.75	0.479	-45.66

higher cointegration density, with 19.7–21.2% of intracuster pairs satisfying stationarity at $p < 0.05$, compared with 12.4–13.1% in the NIFTY 500, and a correspondingly shorter mean half-life of reversion (11.2 days in the S&P 500 Broad cohort versus 19.8 days in the NIFTY Broad cohort). This reflects a structural trade-off: the NIFTY 500 offers the potential for more severe mispricing episodes, while the S&P 500 offers more frequent but smaller, faster-correcting deviations.

D. Backtesting Results: Traditional vs. ML-Enhanced Strategy

Table IV compares the net-of-cost performance of the traditional distance/cointegration benchmark against

E. Comparative Market Analysis

The cross-market comparison reveals an “efficiency trap”: the NIFTY 500 offers larger raw mispricings but demands greater risk tolerance and a substantially

the NIFTY 500, the traditional strategy dominates in the Long cohort (Sharpe 0.913 versus 0.367 for ML) but the ML-enhanced strategy is clearly superior in the Broad cohort (Sharpe 1.090 versus 0.869), with maximum drawdown improving from -31.22% to -21.27%. In the S&P 500, ML enhancement improves the Sharpe ratio in both cohorts - from 0.440 to 0.583 (Long) and 0.424 to 0.479 (Broad) - though maximum drawdown widens, reflecting the trade-off between structural coherence and residual sensitivity to systemic stress in a tightly integrated market.

higher hurdle clearance, while the S&P 500 offers structural integrity but limited dispersion-driven alpha. Because India’s 7.2% hurdle screens out roughly 40% of statistically significant cointegrated pairs that would

clear the 4.2% U.S. threshold, profitability in the Indian market must be assessed against sovereign and inflation risk premia rather than in absolute terms. During the March 2020 shock, the ML-enhanced equity curve was approximately 15% less steep in drawdown than the traditional benchmark, indicating improved downside convexity and capital resilience attributable to the PCA-based factor isolation and OPTICS clustering discipline.

F. Risk Analysis and Robustness Checks

Transaction-cost sensitivity analysis shows the S&P 500 strategy is highly sensitive to execution costs: raising round-trip costs to 25 bps drives net Sharpe toward zero in some cohorts, reflecting thin alpha margins in a mature institutional market. Several NIFTY 500 pairs, notably INDHOTEL–EIHOTEL in the Broad cohort, remain economically viable at cost levels of approximately 30 bps, indicating that dispersion-based inefficiencies in developing markets generate spread magnitudes large enough to absorb higher trading frictions. During the March 2020 dislocation, cointegrated NIFTY 500 pair spreads exceeded four standard deviations from their historical range; the hard 5% stop-loss constrained what would otherwise have been drawdowns exceeding 30% at the position level, supporting its role as a disciplined, rulebased tail-risk control.

G. Statistical Significance Testing

Three complementary tests validate that the ML-enhanced performance gain is not attributable to chance. A one-sided t-test of the null hypothesis of nonpositive mean daily returns is rejected at the 99% confidence level in the NIFTY 500 Broad cohort ($t = 3.82$, $p < 0.01$) and at the 95% level in the S&P 500 Broad cohort ($t = 2.65$, $p = 0.004$). The Jobson-Korkie test (with Memmel correction) confirms that the Sharpe ratio improvement in the NIFTY 500 Broad cohort (1.090 versus 0.869) is statistically significant ($z = 2.07$, $p = 0.038$ across 2,752 daily observations), supporting Hypothesis H1. A 10,000-iteration bootstrap resampling procedure, substituting randomly constructed volatility-matched pairs for the statistically valid cointegrated pairs, yields an empirical p-value of 0.02 - only 2% of random simulations outperformed the ML-enhanced strategy, indicating that the densitybased clustering framework captures a genuine structural relationship rather than spurious correlation.

H. Hypothesis Testing Summary

Table V summarizes the outcome of formal hypothesis testing. H1 is accepted: OPTICS clustering produces measurably lower noise inclusion and higher postselection cointegration and stability than traditional distance-based methods. H2 is partially accepted: while the NIFTY 500 generates significantly higher gross returns, the risk-adjusted advantage narrows substantially once the 7.2% hurdle rate and transaction costs are applied. H3 is accepted: performance differs materially between the Long and Broad regimes across Sharpe ratio, mean-reversion half-life, and drawdown behaviour in both markets.

TABLE V EVALUATION OF HYPOTHESES

Hyp.	Statement	Key Evidence	Decision
H1	OPTICS clustering reduces unstable/noncointegrated pair selection vs. traditional methods.	Lower noise inclusion; higher cointegration density; improved post-selection stability.	Accepted
H2	NIFTY 500 gross returns exceed S&P 500, but net risk-adjusted advantage is not proportionally sustained.	Higher gross ROI in NIFTY 500; Sharpe advantage narrows after hurdle and cost adjustment.	Partially Accepted
H3	Performance differs significantly between Long (2007–2025) and Broad (2015–2025) regimes.	Variation in Sharpe, mean-reversion half-life, and drawdown across regimes.	Accepted

V. Discussion, Conclusion and Recommendations

A. Summary of Empirical Findings

The empirical evidence confirms that ML-based selection, using PCA-neutralized residuals and OPTICS density-based clustering, meaningfully improves selection discipline relative to naive correlation- or distance-based pair selection, though its benefit is regime- and market-dependent rather than universal. In the NIFTY 500 Broad cohort, the ML-enhanced strategy delivers both a higher Sharpe ratio (1.090 versus 0.869) and a materially reduced maximum drawdown (−21.27% versus −31.22%), while in the NIFTY 500 Long cohort the traditional distance-based strategy outperforms, reflecting greater dispersion-driven opportunity in the earlier, less institutionalized regime. In the S&P 500, ML enhancement delivers steady, moderate gains in both cohorts, consistent with a market where factor cohesion is high (PC1 explains approximately 39% of variance) and mean reversion occurs quickly (an 11.2-day halflife in the Broad cohort), leaving limited residual dispersion to exploit. India's 7.2% sovereign hurdle screens out a substantial share of cointegrated pairs that would be economically viable in the U.S. market, confirming that alpha is not an absolute construct but one conditioned on the local cost of capital.

B. Theoretical, Methodological, and Practical Contributions

Theoretically, the study establishes a hurdle-adjusted framework for comparing arbitrage viability across markets with different sovereign opportunity costs, extending the Adaptive Market Hypothesis by linking mean-reversion speed, trade duration, and arbitrage persistence to market maturity and factor concentration. Methodologically, the study extends the use of density-based OPTICS clustering to full equity-index-scale statistical arbitrage and demonstrates the material cost of survivorship bias (an estimated 4.94% return inflation) through point-in-time universe reconstruction. Practically, the study operationalizes a complete institutional-grade pipeline - PCA factor neutralization, OPTICS clustering, 15 bps transaction cost modeling, a −5% stop-loss, and SHAP-based

interpretability - offering a transparent, governance-ready template for deploying machine-learning-enhanced arbitrage under real capital constraints.

C. Strategic Recommendations and Managerial Implications

For the NIFTY 500, capital allocation should be magnitude-oriented: managers should adopt wider Sscore entry thresholds (e.g., above 2.0) to ensure trades clear the higher 7.2% hurdle net of costs, size positions dynamically according to cross-sectional dispersion, and maintain larger capital buffers for volatility. For the S&P 500, allocation should be consistency-oriented: a larger number of smaller, statistically significant trades with tight execution and cost control is preferable to episodic, magnitude-driven positioning, since profitability there depends on precision rather than spread size. Across both markets, PCA loadings and OPTICS clusters should be recalibrated periodically (e.g., quarterly) to control model drift, factor stability should be monitored via rolling eigenvector correlation, and cluster stability should be tracked using an Adjusted Rand Index to inform holding-period and risk-limit decisions.

D. Limitations of the Study

The study is confined to single-asset equity statistical arbitrage in the S&P 500 and NIFTY 500 and does not extend to cross-asset arbitrage in commodities, fixed income, currencies, or derivatives. The 15 bps transaction-cost assumption, while realistic on average, is modeled as constant and may overstate friction in highly liquid U.S. large caps while understating it in thinner NIFTY 500 segments; dynamic market-impact effects tied to assets under management are not modeled. The framework does not explicitly incorporate macroeconomic regime detection or structural-break modeling beyond rolling PCA recalibration, and its machine-learning component is limited to unsupervised techniques (PCA and OPTICS) rather than supervised or reinforcement-learning approaches to entry and exit timing. Finally, the backtest relies on daily adjusted closing prices and does not capture intraday microstructure noise, bid-ask dynamics, or high-frequency execution slippage, which

may cause live implementation to diverge from backtested estimates, particularly in higher-turnover markets.

E. Recommendations for Future Research

Future work could integrate formal regime-detection methods, such as Hidden Markov or Markov regimeswitching models, into the PCA-OPTICS pipeline to allow dynamic switching between mean-reversiondominant and trend-dominant environments. Nonlinear residual modeling using architectures such as LSTM networks could improve estimation of meanreversion parameters beyond the current linear, Ornstein-Uhlenbeck-consistent framework. The model could be extended to cross-asset relative-value relationships (e.g., equity-futures basis trades, ADRversus-domestic-listing spreads), and reinforcement learning could be applied to optimize trade execution and reduce market impact, particularly in the thinmargin S&P 500 context. Dynamic, microstructureaware transaction-cost modeling and adaptive alphadecay monitoring - automatically reducing exposure to clusters showing weakening signal strength or crowding - would further improve the realism and scalability of the framework.

F. Conclusion

This study examined whether machine-learningenhanced statistical arbitrage can deliver economically significant, structurally robust performance across the NIFTY 500 and S&P 500. Combining PCA-based factor neutralization with OPTICS density-based clustering moves pair selection beyond naive correlation toward a residual, structurally filtered framework, and the empirical evidence confirms that this improves selection discipline, drawdown containment, and risk-adjusted returns under institutionally realistic constraints, particularly in the

NIFTY 500 Broad cohort. The S&P 500 shows structurally compressed alpha magnitude but statistically significant incremental gains from ML enhancement, underscoring the importance of execution precision and systematic filtering in a highly efficient market. A central conclusion is that arbitrage scalability is defined by market structure: developed

markets combine higher factor cohesion, faster reversion, and tighter mispricing windows that reward consistency and cost efficiency, while developing markets offer deeper idiosyncratic dispersion at the cost of a higher capital hurdle and greater volatility. By incorporating sovereign hurdle rates, realistic transaction costs, and survivorship-bias-free backtesting, the study grounds its conclusions in economic rather than purely statistical significance, and demonstrates that machine learning does not eliminate the structural limits to arbitrage but meaningfully improves an institution's ability to operate within them.

References

- [1] Y. Amihud and H. Mendelson, "Asset pricing and the bid-ask spread," *Journal of Financial Economics*, vol. 17, no. 2, pp. 223–249, 1986.
- [2] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," in *Proc. 1999 ACM SIGMOD Int. Conf. on Management of Data*, 1999, pp. 49–60.
- [3] M. Avellaneda and J.-H. Lee, "Statistical arbitrage in the U.S. equities market," *Quantitative Finance*, vol. 10, no. 7, pp. 761–782, 2010.
- [4] R. A. Brealey, S. C. Myers, and F. Allen, *Principles of Corporate Finance*, 13th ed. McGraw-Hill Education, 2020.
- [5] S. J. Brown, W. N. Goetzmann, R. G. Ibbotson, and S. A. Ross, "Survivorship bias in performance studies," *Review of Financial Studies*, vol. 5, no. 4, pp. 553–580, 1992.
- [6] E. P. Chan, *Algorithmic Trading: Winning Strategies and Their Rationale*. Wiley, 2013.
- [7] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.
- [8] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
- [9] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, "Pairs trading: Performance of a relative-value arbitrage rule," *Review of Financial Studies*, vol. 19, no. 3, pp. 797–827, 2006.
- [10] J. C. Hull, *Options, Futures, and Other Derivatives*, 10th ed. Pearson, 2018.
- [11] M. C. Jensen, "The performance of mutual funds in the period 1945–1964," *Journal of Finance*, vol. 23, no. 2, pp. 389–416, 1968.
- [12] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.

- [13] A. E. Khandani and A. W. Lo, "What happened to the quants in August 2007?," *Journal of Investment Management*, vol. 5, no. 4, pp. 5–54, 2007.
- [14] A. W. Lo, "The adaptive markets hypothesis: Market efficiency from an evolutionary perspective," *Journal of Portfolio Management*, vol. 30, no. 5, pp. 15–29, 2004.
- [15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [16] A. F. Perold, "The implementation shortfall: Paper versus reality," *Journal of Portfolio Management*, vol. 14, no. 3, pp. 4–9, 1988.
- [17] J. M. Poterba and L. H. Summers, "Mean reversion in stock prices: Evidence and implications," *Journal of Financial Economics*, vol. 22, no. 1, pp. 27–59, 1988.
- [18] A. Shleifer and R. W. Vishny, "The limits of arbitrage," *Journal of Finance*, vol. 52, no. 1, pp. 35–55, 1997.