Empirical Limitations of Elementary Machine Learning in Similarity-Based Hedging: kNN Fails to Identify Effective Hedges in Equity Portfolios

Jackson McBride*
Ellison Institute of Technology, Oxford
jmcbride@eit.org

October 2025

Abstract

This paper evaluates the effectiveness of the k-nearest neighbor (kNN) algorithm in identifying effective long-short hedged portfolios based on historical return-series vectors discounted over time. Using correlation distance, defined as one minus the Pearson correlation between exponentially weighted and standardized return vectors, we generate nearest neighbors for each target asset. We randomly sample 2,340 U.S. equities spanning 45 years and construct 84,240 candidate hedge portfolios through the kNN procedure, each evaluated up to 252 trading days into the future. Across all model specifications, the constructed portfolios tend on average to increase rather than reduce variance, providing little evidence of meaningful hedge effectiveness. Despite this systematic underperformance, heterogeneity in outcomes offers additional insight. Certain sectors exhibit greater persistence in variance structure, suggesting that past volatility may occasionally inform future dynamics. Moreover, increasing the number of neighbors or extending the testing horizon yields modest improvement in mean performance but also increases the dispersion of results.

These findings indicate that historical similarity in return behavior, when measured by correlation distance, does not reliably translate into forward-looking co-movement stability. The analysis provides a large-scale empirical benchmark for the limitations of nonparametric similarity-based methods in portfolio construction and highlights the need for adaptive or structural modeling approaches when attempting to identify effective hedges in nonstationary financial environments.

*Code and replication materials available at `https://github.com/jacksonmcbride/trading-knn` (archived at Zenodo, DOI: `https://doi.org/10.5281/zenodo.17404455`). Research Conducted at Columbia University, Department of Computer Science, and Norges Bank Investment Management.

Keywords: k-nearest neighbors, hedging, portfolio construction, machine learning, nonparametric methods, correlation instability

1 Introduction

Constructing hedges that reduce portfolio variance requires models whose estimated dependence structures remain stable out of sample. In practice, equity co-movements evolve across regimes, with correlations rising in market stress and exhibiting asymmetric behavior that standard static estimators fail to capture [13, 1]. These facts motivate time-varying covariance models such as Dynamic Conditional Correlation, which explicitly parameterize correlation dynamics [5], and they also motivate statistical remedies that reduce estimation error in high dimensions, including shrinkage toward structured targets and random-matrix-based cleaning [12, 11, 16]. Parallel work has explored the geometry of correlation spaces and hierarchical market structure to inform clustering and distance measures [14]. At the same time, distance-based selection has seen practical adoption in relative-value strategies such as pairs trading, where historical proximity of normalized price paths guides portfolio construction [6].

Against this backdrop, simple machine learning methods continue to be applied to financial data. Their appeal is interpretability and flexibility, yet the literature has repeatedly cautioned that nonstationarity, low signal-to-noise, and backtest overfitting often undermine apparent success [4, 3, 10]. A particularly common heuristic is k-nearest neighbors, which selects instruments by historical similarity under a chosen distance. Despite its popularity for classification and short-horizon prediction tasks, there is limited systematic evidence on whether plain kNN applied to raw return histories can identify hedges that reduce forward variance across assets, horizons, and regimes.

This paper provides such an evaluation. We study a large universe of U.S. equities over multiple decades and implement a rolling out-of-sample design with exponential weighting in the lookback window consistent with industry practice [8]. For each date, we select neighbors by correlation distance and estimate static hedge weights from the weighted window. We vary neighbor count, half-life, and investment horizon to assess robustness. Anticipating the literature on correlation instability, our central question is deliberately narrow and empirical: can a nonparametric, similarity-based rule recover stable forward covariance structure well enough to deliver effective hedges?

Our results indicate that it does not. Across years, sectors, horizons, and parameterizations, kNN-based hedges fail to reduce variance on average and do not improve alternative risk measures. The evidence is consistent with a simple interpretation: past return proximity is not a reliable indicator of future covariance stability in equities. The contribution of this study is therefore twofold. First, it establishes a strong empirical baseline that documents the limitations of elementary similarity learning for hedging at scale. Second, it situates those findings within a broader body of work on dynamic and noisy dependence,

clarifying why methods that rely on static historical distances are unlikely to generalize without additional structure, cleaning, or explicit modeling of time variation.

2 Data and Summary Statistics

2.1 Data Sources and Universe Construction

We construct our asset universe from official symbol directories published by NASDAQ and other U.S. exchanges.[1] All currently listed tickers across NASDAQ, NYSE, NYSEMKT, and NYSEARCA are compiled and converted into Yahoo-compatible formats (e.g., "BRK.B" $\rightarrow$ "BRK-B"). Assets are classified as *stocks* based on Yahoo's `quoteType` and name metadata. Raw daily price histories are retrieved via the `yfinance` API, which provides Adjusted Close, Close, Volume, Dividend, and Split series at daily frequency for the full available record of each ticker.[2] The earliest valid observation in our dataset is January 3, 1962, giving an overall coverage period from 1962 through 2025.

Our final sample includes 6,493 active tickers with non-empty price histories. Yahoo Finance provides data only for currently listed or recently delisted securities. Many delisted or bankrupt microcaps are missing, but a large number of inactive or near-zero microcaps remain. This inclusion of defunct or low-priced tickers leads to negatively skewed cross-sectional returns, an artifact that we document explicitly below.

2.2 Return Construction

We construct daily log returns for each asset as

$$r_{i,t} = \ln(P_{i,t}^{\text{adj}}) - \ln(P_{i,t-1}^{\text{adj}}),$$

where $P_{i,t}^{\text{adj}}$ is Yahoo's Adjusted Close, incorporating dividends and stock splits. Each return series is cleaned by removing non-positive prices, eliminating duplicate timestamps, and sorting by trading date. Tickers with too few valid observations or entirely missing Adjusted Close values are excluded. We then compute summary statistics including the mean and standard deviation of daily returns, daily return volatility, the number of valid trading days, lifetime start and end dates, and the compound annual growth rate, trimmed at the 1st and 99th percentiles to mitigate outliers from penny-stock baselines.

2.3 Filtering and Sample Characteristics

Before inclusion, tickers are filtered to ensure data quality. Securities with fewer than 250 valid daily observations or those that never trade above \$0.50 are discarded. In addition, tickers with implausible price dynamics—such as price

[1] Sources: `https://nasdaqtrader.com/dynamic/SymDir/nasdaqlisted.txt` and `https://nasdaqtrader.com/dynamic/SymDir/otherlisted.txt`.

jumps exceeding ±80% in a single day, inconsistent timestamps, or entirely constant price histories—are removed. The remaining universe of stocks thus represents active or semi-active listed equities with sufficient history to compute return-based measures reliably.

Table 1 summarizes the cross-sectional characteristics of the final sample. The average daily return is 0.03%, and the average annualized return is 5.60%, where both values represent the equal-weighted mean of per-ticker averages rather than a capitalization-weighted aggregate. Daily volatility is higher and more right-skewed, consistent with the long-tailed distribution of equity risk.

Table 1: Stocks: Summary statistics

	n	Mean	Median	Std	p5	p95
Avg daily return (%)	3,383	0.02	0.03	0.11	-0.12	0.09
Daily return volatility (%)	3,383	3.41	2.86	2.51	1.40	7.19
Days of data	3,383	5,441.94	4,827.00	3,827.10	789.10	11,827.90
CAGR (%, trimmed 1–99)	3,315	5.60	7.70	14.91	-22.77	24.34

2.4 Date Coverage

Figure 1 shows the distribution of first available trading dates for all equities. Stocks span a wide historical range (minimum: 1962-01-03, median: 2014-11-03). Virtually all assets have data through the final observation date (2025-10-09).

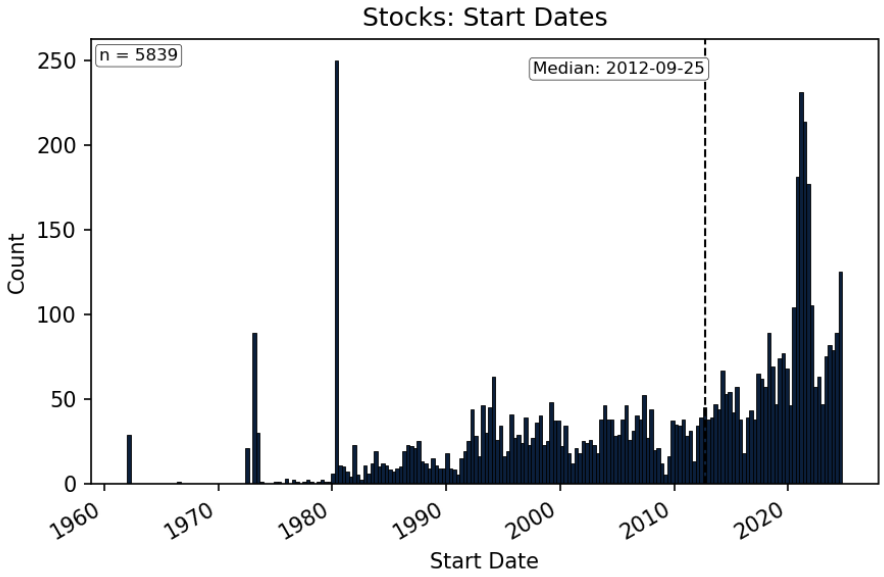

Figure 1: Stock first-available date distribution.

2.5 Distributional Properties

Figure 2 plots the cross-sectional histogram of average daily stock returns. Both the mean and median are modestly positive, but dispersion is substantial, reflecting the high variability of daily returns across firms.

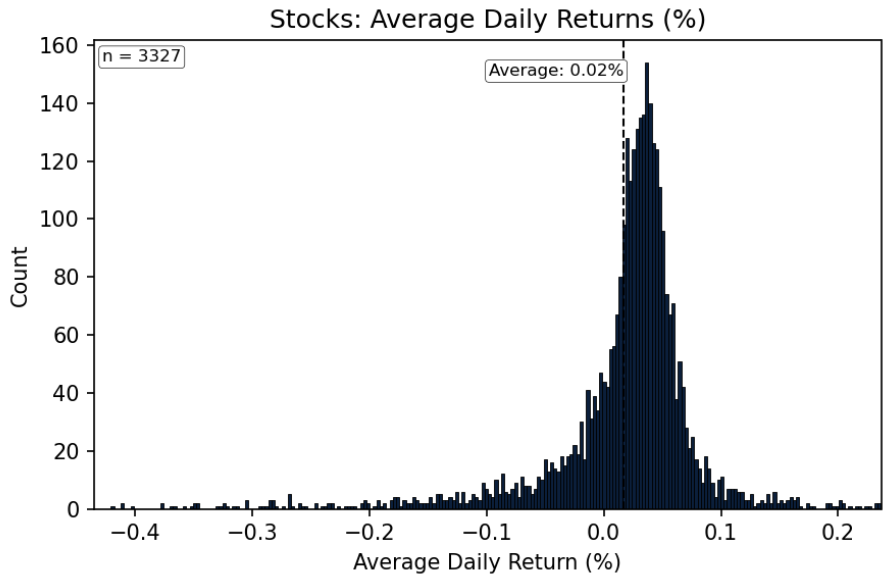

Figure 2: Histogram of stock average daily returns (%).

Figure 3 displays lifetime CAGRs (trimmed 1–99%). The left tail of the distribution is dominated not only by older or inactive microcaps, but also by a cohort of recently listed equities that experienced extreme valuation collapses following large reverse stock splits. Examples include Elevai Labs, Inc. (ELAB), which executed a 1-for-200 reverse split in November 2024, and Golden Heaven Group Holdings Ltd. (GDHG), which conducted multiple consolidations, including a 1-for-25 reverse split in 2025. Such corporate actions mechanically rescale historical prices upward, so the adjusted "first" observations in our dataset appear to start at extraordinarily high levels, and subsequent declines register as annualized losses of 90–99%. Because our filtering is intentionally permissive—designed to retain legitimate microcap names—these issuers remain within the sample and heavily influence the lower tail. Accordingly, the left tail reflects not just legacy delisted firms drifting toward zero, but also recent speculative listings, often microcap uplists or reverse-merger vehicles with limited revenue and significant dilution risk. These structural features produce unstable return dynamics and idiosyncratic volatility, contributing to the pronounced negative skew in the cross-section and to the divergence between mean and median returns.[7]

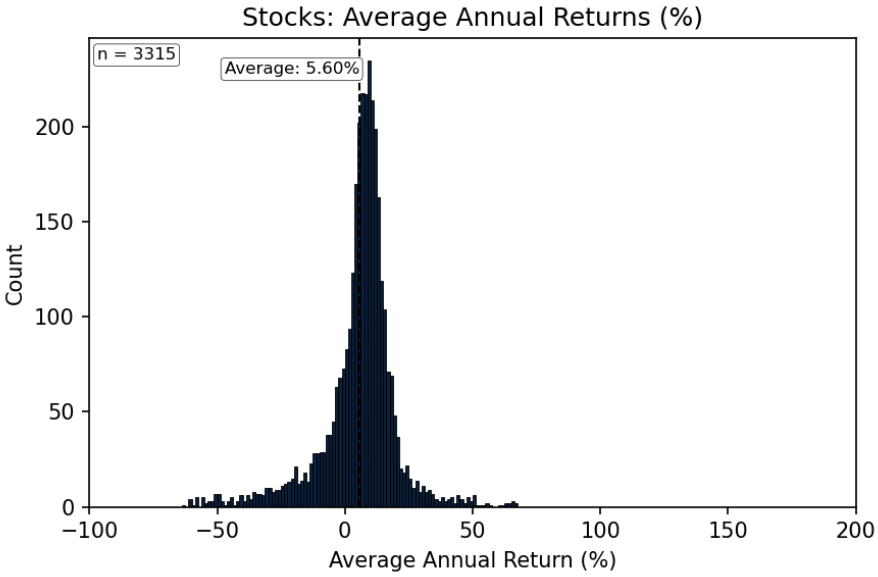

Figure 3: Distribution of lifetime stock CAGRs (trimmed 1–99%).

Figure 4 illustrates the number of valid price observations per ticker, confirming the long-tailed lifespan distribution for equities.

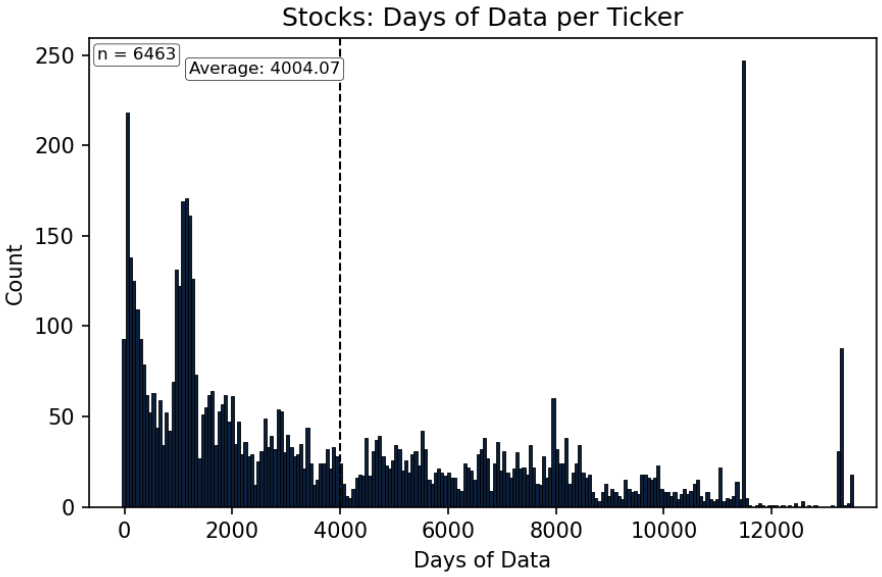

Figure 4: Distribution of days of valid data per stock.

2.6 Limitations

Yahoo Finance aggregates data from exchange feeds but does not provide complete delisting histories. Thus, while we capture many failed or microcap equities, survivorship bias is reduced but not eliminated. Adjusted Close values are sourced from Yahoo's adjustment algorithm, which generally aligns with CRSP-style adjustments but may occasionally mis-handle distributions. Equal-weighted averages are strongly influenced by the long left tail of defunct microcaps; a market-cap-weighted aggregation would yield higher average returns.

3 Methodology

3.1 Universe and Data Acquisition

The analysis uses the equity universe and historical price series described in Section 2. All securities are common stocks listed on major U.S. exchanges, with daily adjusted closing prices retrieved from Yahoo Finance. The dataset spans from 1962 through 2025 and includes only equities that pass the data-quality and liquidity filters previously outlined. These returns serve as the foundation for constructing the empirical hedging experiment.

3.2 Return Construction

Daily log returns are computed from adjusted closing prices as

$$r_{i,t} = \ln(P_{i,t}^{\text{adj}}) - \ln(P_{i,t-1}^{\text{adj}}),$$

where $P_{i,t}^{\text{adj}}$ incorporates corporate actions such as dividends and stock splits. All returns used in subsequent modeling are identical to those defined in Section 2, except where further weighting or normalization is introduced in the hedging procedure.

3.3 Out-of-Sample Evaluation Design

The study employs a rolling out-of-sample framework. Evaluation dates occur at weekly intervals throughout each calendar year. On each date d, a lookback window ending at d is formed. Eligible target securities are those with at least $6 \times \text{HL}_{\min}$ valid daily returns, where $\text{HL}_{\min}$ is the shortest half-life parameter used. One target is drawn uniformly at random from the eligible set. The hedge portfolio constructed on date d is held static for the full forward evaluation period with no rebalancing or weight updates.

Figure 5 summarizes the frequency with which each stock was selected as a target across all evaluation dates. The selection distribution is highly uneven, as expected when sampling over multiple decades. Long-lived, continuously listed firms appear more frequently because they remain eligible in a larger fraction of rolling windows. This confirms that the randomization procedure does not overweight particular time periods but naturally favors assets with complete and uninterrupted price histories. The most frequently sampled names, such as UniFirst (UNF), Altria (MO), Unifi (UFI), Chemed (CHE), and PG&E (PCG), are all long-established U.S. companies with uninterrupted trading records dating back to the 1980s or earlier. Their prevalence simply reflects persistent data availability rather than structural sampling bias.

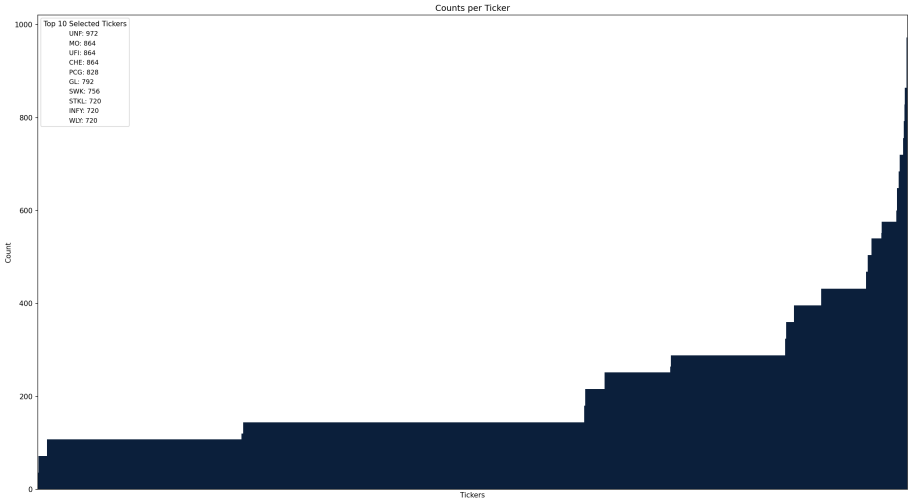

Figure 5: Distribution of the number of times each ticker was selected as a target across all evaluation dates. Long-lived firms dominate the upper tail because they are eligible in a larger fraction of rolling windows.

3.4 Weighted Standardization and Lookback Window

For each target and half-life h, a lookback length of $L = 6h$ daily observations is used. Exponentially weighted moving average row weights are applied, with decay parameter $\lambda = \ln(2)/h$.[9] Within this window, each return series is standardized to zero mean and unit variance using the weighted mean and variance. This normalization ensures comparability across assets with differing volatilities before computing distances or fitting regressions.

3.5 Neighbor Selection

Potential hedge candidates are identified through a nearest-neighbor search conducted within the standardized lookback window. For each target asset i and candidate j, distance is defined as

$$d(i, j) = 1 - \rho(i, j),$$

where $\rho(i, j)$ is the Pearson correlation coefficient between the weighted return vectors. This metric favors assets that are highly positively correlated with the target, consistent with a long–short hedge structure in which the target is held long and the neighbors are shorted. The k assets with the smallest distances are selected as neighbors, with $k \in \{1, 3, 5, 10, 30, 60\}$. We use $1 - \rho$, a monotone transformation of the metric $\sqrt{2(1 - \rho)}$ in Mantegna [15].

3.6 Hedge Weight Estimation

Hedge weights are estimated with ridge regression to produce stable and robust coefficients out of sample. The neighbor assets are deliberately selected for high correlation with the target, which creates strong multicollinearity among their returns. Under ordinary least squares this collinearity leads to unstable weights and excessive leverage. Ridge regularization introduces a controlled penalty that shrinks coefficients toward zero, improving numerical stability and reducing sensitivity to transient correlations while adding minimal bias.

The optimization problem is defined as

$$\min_{\beta} (y - X\beta)^\top W (y - X\beta) + \lambda \beta^\top \beta,$$

where y is the target return vector, X is the matrix of neighbor returns, and W is a diagonal matrix of exponentially weighted moving-average row weights. The regularization parameter is set as

$$\lambda = 10^{-4} \frac{\operatorname{trace}(X^\top W X)}{k}.$$

This scaling ties the penalty to the typical eigenvalue magnitude of $X^\top W X$, ensuring that shrinkage intensity adapts to the overall variance level and to the number of neighbors. The small constant provides mild regularization, sufficient to stabilize the inversion of $X^\top W X$ without materially biasing exposures. This trace-based normalization follows the same logic as trace-scaled penalties in kernel ridge regression and covariance shrinkage estimators, making the degree of penalization comparable across parameter settings.

The estimated coefficients $\hat{\beta}$ define a simulated long–short hedge portfolio:

$$r_t^{\text{hedged}} = r_{T,t} - \sum_{j=1}^{k} \hat{\beta}_j r_{j,t}.$$

The target asset is held long, while the neighbor assets are held short in proportion to the estimated weights. The position is dollar-neutral and remains static throughout the forward evaluation period.

3.7 Forward Evaluation

Forward performance is evaluated over horizons of $H \in \{5, 21, 126, 252\}$ trading days, capped at 365 calendar days. Only overlapping, nonmissing observations between the target and its neighbors are included. The hedge remains unchanged throughout each evaluation window.

3.8 Performance Metrics

Hedging outcomes are summarized by a set of standard risk and return measures. Hedge effectiveness is defined as

$$HE = 1 - \frac{\operatorname{Var}(r_T - \hat{r}_P)}{\operatorname{Var}(r_T)},$$

where r_T is the target return and $\hat{r}_P$ is the hedge portfolio return. Tracking error is the annualized standard deviation of $r_T - \hat{r}_P$. Additional metrics include maximum drawdown reduction, reduction in the 95th percentile Value at Risk, annualized alpha, and directional hit rate. All variances are computed using overlapping valid data and are comparable across horizons after annualization.

3.9 Aggregation and Summary

Each evaluation yields a record containing the target, half-life, neighbor count, horizon, and associated metrics. Results are aggregated across targets and time to obtain means, medians, and shares of positive hedge effectiveness. Cross-sectional and temporal summaries are computed for combinations of half-life, neighbor count, and horizon to characterize the stability of performance across parameter settings.

4 Results

4.1 Overview of Hedge Effectiveness

Table 2 and Figures 6–7 summarize mean hedge effectiveness (HE) across parameter settings and time. Values are averaged across all horizons and years for combinations of neighbor count k and half-life (HL). Across nearly all configurations, mean HE is strongly negative, indicating that the constructed hedges systematically increase rather than reduce variance relative to the unhedged target. Short half lives exhibit more erratic results, but longer half lives do not yield meaningful improvement. Likewise, increasing the number of neighbors introduces additional variability without producing a clear trend toward stability or higher performance.

Table 2: Mean Hedge Effectiveness by Neighbor Count k and Half-Life (HL). Values are averaged across all years and horizons. Negative values indicate variance increases relative to the target.

k	HL=5	HL=30	HL=90	HL=180	HL=365	HL=720
1	-15.739	-11.495	-6.808	-6.948	-7.014	-6.817
3	-3.150	-3.454	-2.913	-2.964	-2.993	-2.971
5	-1.843	-1.784	-1.572	-1.652	-1.638	-1.641
10	-0.840	-0.799	-0.833	-0.869	-0.844	-0.855
30	-0.273	-0.222	-0.196	-0.195	-0.197	-0.207
60	-0.164	-0.098	-0.075	-0.076	-0.076	-0.089

The pattern in Table 2 shows that as k increases, the average HE becomes less negative, approaching zero but never turning positive. This progression

implies that portfolios with fewer neighbors consistently generate worse hedges, particularly over longer evaluation horizons. When only a few neighbors are used, the hedge weights are concentrated on highly correlated assets whose relationships are unstable, leading to amplified residual variance over longer windows. As k increases, these idiosyncratic effects average out, producing slightly less negative mean values but at the cost of higher variance in performance across time. The greater spread in HE among higher k values is visible in Figure 6, where larger k lines fluctuate more widely around the mean. This pattern reflects the fact that including more neighbors introduces additional weak correlations and noise, yielding hedges that are more sensitive to transient co-movements rather than to persistent structure.

Figures 6 and 7 display time-series of mean HE across years. In general, regardless of parameter choice, KNN fails to find effective hedges at shorter horizons (window = 5), while performance becomes slightly less negative at longer horizons. The short-horizon outcomes are dominated by high-frequency, unpredictable variation in returns, which prevents the algorithm from capturing stable relationships. By contrast, at longer evaluation windows the average HE values are still negative but exhibit less volatility, indicating that the apparent performance is dominated by transient, short-term noise rather than persistent co-movement.

The second figure, which varies the half-life, reveals substantial variance and almost no temporal consistency across all parameterizations. The alternating sign and magnitude of HE over time suggest that the exponential weighting scheme captures local, short-lived return patterns rather than enduring relationships between assets. This instability is especially pronounced at longer horizons, where performance alternates between weakly negative and strongly negative values, implying that decaying-weight history does not translate into predictive information about future covariance. Overall, both plots indicate that hedge effectiveness is driven primarily by short-term, unpredictable variation in returns rather than by any lasting structure in cross-sectional correlations.

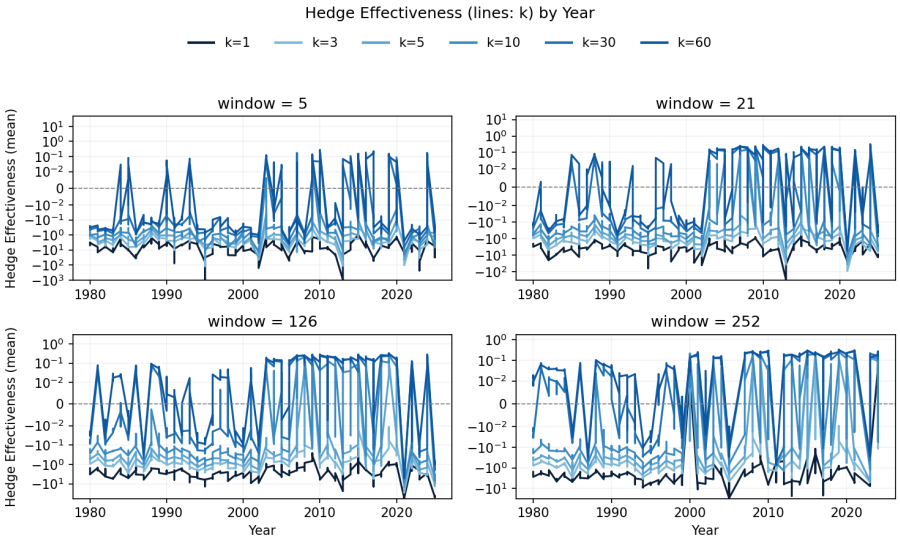

Figure 6: Mean hedge effectiveness by year, grouped by neighbor count k. Larger k values exhibit greater variance and remain consistently negative, though slightly less extreme than for small k.

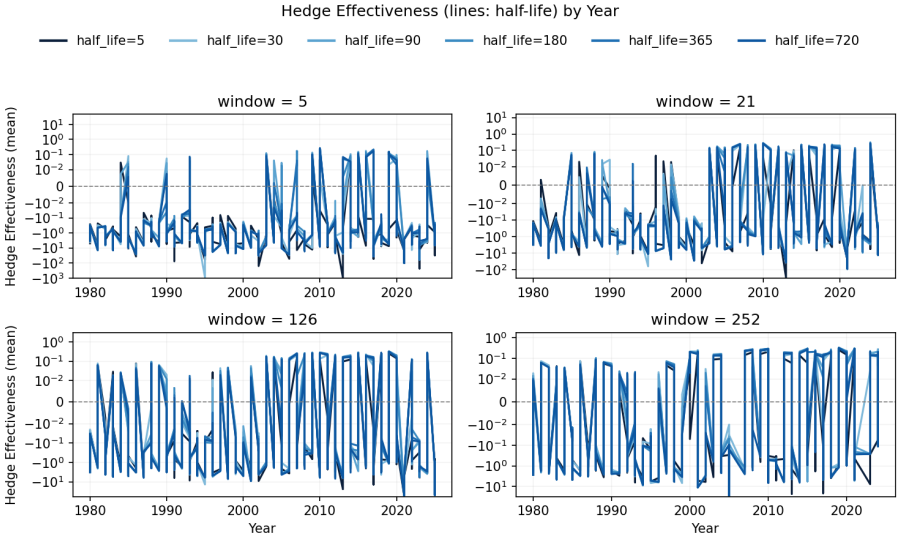

Figure 7: Mean hedge effectiveness by year, grouped by half-life. The high volatility and alternating sign of values across all horizons indicate that weighting history by half-life fails to capture persistent co-movement.

Although HE is our primary measure, other risk metrics—tracking error, maximum drawdown (MDD) reduction, 95% Value-at-Risk (VaR_{95}) reduction, annualized alpha, and hit rate—show the same qualitative pattern when summarized across the same $k\times$ half-life grid. All exhibit near-zero or negative mean values and no statistically meaningful dependence on parameterization. These additional results, consistent with the findings presented here, are reported in Appendix 5.

4.2 Dependence on Horizon and Parameterization

Table 3 and Table 4 examine how hedge effectiveness varies across evaluation horizons and model parameters. The results show that performance remains uniformly poor across all horizons and parameter combinations. Short-horizon tests exhibit the most extreme losses in hedge effectiveness, while longer-horizon evaluations show less negative values, suggesting that high-frequency noise dominates the short-term outcomes. Nevertheless, even over a full trading year, no configuration yields consistent or statistically significant improvement.

Table 3: Mean Hedge Effectiveness by Half-Life and Evaluation Horizon. Values represent averages across all years and configurations; negative values indicate variance increases relative to the target.

Half-Life	5-day	21-day	126-day	252-day
5	-9.276	-2.852	-1.484	-0.879
30	-7.416	-2.279	-1.420	-0.633
90	-3.761	-2.171	-1.438	-0.813
180	-3.858	-2.218	-1.468	-0.844
365	-3.881	-2.226	-1.474	-0.842
720	-3.716	-2.223	-1.498	-0.870

The pattern across half-lives demonstrates that the exponential weighting scheme does not stabilize hedge performance. All half-life configurations produce strongly negative mean hedge effectiveness, especially at short horizons, where daily-level variation dominates. Although longer half-lives slightly smooth the results, the sign remains negative across every horizon. The p-values corresponding to Table 3 confirm that none of these values are statistically significant at conventional confidence levels, underscoring that differences across horizons arise from noise rather than structural predictability. The variation observed in Figure 7 therefore reflects instability rather than improvement. At longer horizons, the less negative mean values likely result from averaging over uncorrelated fluctuations, not from genuine hedging success.

Table 4: Mean Hedge Effectiveness by Neighbor Count k and Evaluation Horizon. Values represent averages across all years and configurations; negative values indicate variance increases relative to the target.

k	5-day	21-day	126-day	252-day
1	-19.306	-7.906	-5.420	-3.551
3	-6.032	-3.477	-1.750	-0.894
5	-3.642	-1.631	-0.954	-0.445
10	-2.028	-0.686	-0.474	-0.126
30	-0.581	-0.189	-0.124	0.052
60	-0.319	-0.079	-0.058	0.083

The relationship between neighbor count and performance shows a clear and persistent pattern. At small k, hedge effectiveness is strongly negative across all horizons, reaching values as low as -19.3 at the 5-day horizon and remaining near -3.6 even at 252 days. This reflects the instability of highly concentrated hedges, where the small number of neighbors amplifies idiosyncratic noise and short-term correlation shifts. As k increases, the average HE becomes progressively less negative, and by $k = 30$ and $k = 60$, the values turn slightly positive for the longest evaluation window, approximately 0.05 and 0.08, respectively. This apparent improvement, however, is minimal in magnitude and lacks consistency across horizons. It likely arises from statistical averaging rather than genuine hedge performance, as the inclusion of many weakly correlated assets reduces the influence of short-term noise but does not capture any stable forward relationship.

Across all horizons, the pattern is clear: hedge effectiveness is most negative at short horizons, dominated by unpredictable high-frequency variation in return correlations, and becomes less extreme as the evaluation window lengthens. At longer horizons, the influence of short-term noise is diluted, producing values closer to zero or marginally positive, but still lacking any systematic or economically meaningful gain. The higher variance in performance among large k portfolios further indicates that adding neighbors introduces instability in weight estimation, as the distance metric incorporates increasingly irrelevant or weakly related assets.

4.3 Cross-Sectional and Sectoral Patterns

We next evaluate whether hedge performance differs systematically across sectors or industries. Figure 8 and Figure 9 present the distributions of mean hedge effectiveness across major sectors and the twenty most represented industries. The results reveal that the failure of the KNN approach is pervasive rather than confined to a specific part of the market. Median hedge effectiveness is negative or near zero across all sectors, with wide interquartile ranges indicating high dispersion even within each group. No single sector demonstrates consistent positive performance, suggesting that correlation-based similarity does not

translate into stable hedging relationships in any domain.

The Technology and Health Care sectors exhibit slightly wider variability in outcomes, occasionally reaching small positive values, but these instances appear idiosyncratic rather than systematic. Traditional defensive sectors such as Utilities and Consumer Non-Durables tend to cluster below zero, implying that their return structures are no more predictable than those of cyclical sectors. The broad overlap in interquartile ranges across categories suggests that industry-specific fundamentals play little role in determining KNN hedge quality; instead, the results are dominated by random correlation fluctuations at the asset level. This lack of separation between sectors indicates that the inefficacy of the approach arises from the underlying statistical instability of return co-movements rather than from differences in market structure or asset class characteristics.

At the industry level, a similar picture emerges. While a few industries, such as Metal Fabrications and Semiconductors, display slightly positive medians, most industries cluster around zero with wide tails extending into negative territory. This dispersion implies that even within narrowly defined groups where one might expect stronger structural relationships, historical return similarity provides no reliable basis for forward-looking hedges. The heterogeneity of outcomes across industries further supports the interpretation that the correlation landscape is highly dynamic, and that KNN applied to raw return vectors cannot consistently identify assets that provide meaningful variance reduction.

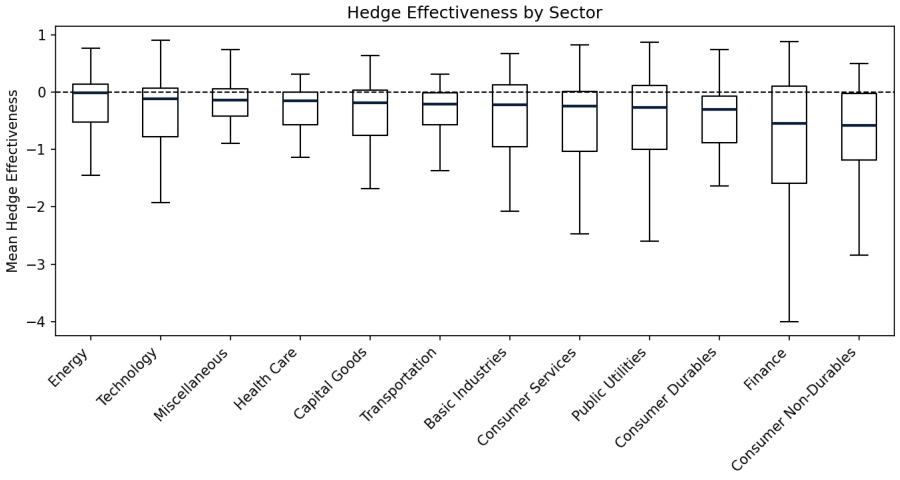

Figure 8: Distribution of mean hedge effectiveness by sector. Median hedge effectiveness is negative or near zero across all sectors, with no evidence of consistent outperformance in any category.

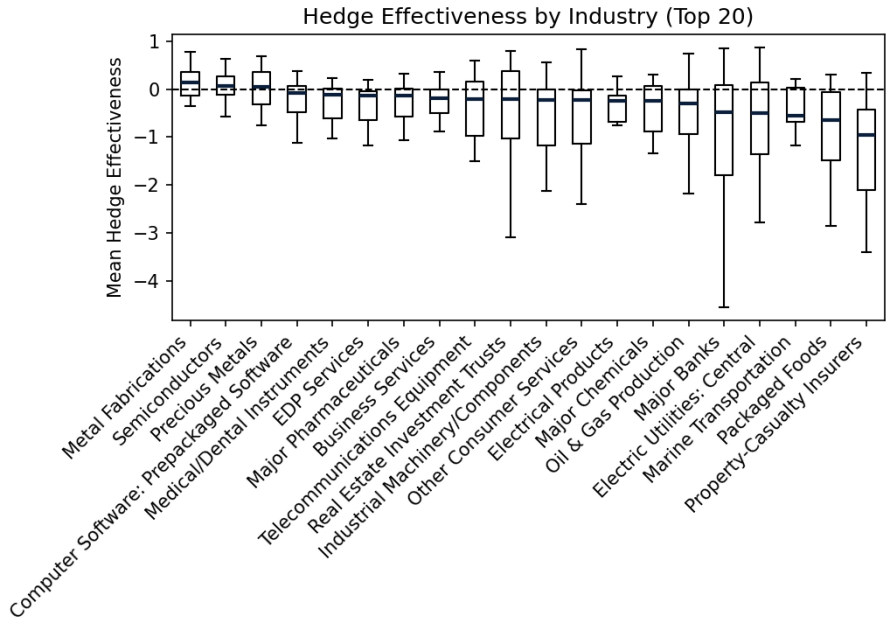

Figure 9: Distribution of mean hedge effectiveness of top 20 industries by sample count. Slightly positive medians in a few industries such as Metal Fabrications and Semiconductors are outweighed by broad variability and negative tails across most groups.

4.4 Other Evaluation Metrics

Although hedge effectiveness is our primary measure, the same patterns appear for other performance metrics. Tables summarizing tracking error, maximum drawdown reduction, 95% Value-at-Risk reduction, annualized alpha, and hit rate show similarly weak or negative results. These additional metrics are computed in the same format as the HE tables: $k\times$ half-life, $k\times$ window, and half-life $\times$ window matrices, along with associated p-value tables. None display statistically meaningful positive performance.

4.5 Time Variation and Robustness

Finally, Figure 6 and Figure 7 summarize the temporal evolution of HE across the full sample. The consistency of near-zero mean values across years, including major market regimes (e.g., 2000, 2008, 2020), indicates that the failure of the KNN-based hedge identification is robust to time period, volatility regime, and parameterization.

4.6 Interpretation

Taken together, these results suggest that KNN applied to historical return vectors fails to produce viable forward-looking hedges. The method primarily identifies assets that were historically similar rather than oppositely co-moving, and the return correlation structure changes too rapidly for past similarity to imply future covariance stability. The universality of the null result across time, horizons, and sectors implies that the limitation is intrinsic to the approach rather than a sampling artifact.

5 Conclusion

The results of this study show that k-nearest neighbor methods, when applied to historical return vectors, do not yield effective or reliable hedges. Across all horizons, sectors, and parameterizations, the constructed portfolios fail to reduce variance and exhibit no statistically significant improvement across alternative risk metrics. This outcome persists through multiple robustness checks and time periods, suggesting that the limitation is inherent to the approach rather than a function of data or sampling choices. The kNN algorithm, by design, identifies neighbors based on historical similarity, yet financial relationships among assets evolve too quickly for such similarity to remain predictive. The apparent closeness in past returns does not imply stability in future co-movements, and thus the method captures transient noise rather than durable structure.

This finding offers a useful caution for the application of nonparametric learning techniques in portfolio construction. It illustrates that models built solely on historical proximity may struggle in environments characterized by regime shifts and nonstationary dependence structures. The adage that past relationships are not a reliable guide to future dynamics is confirmed here in a systematic, data-driven setting. The results do not imply that all machine learning approaches to hedging are futile, but they do indicate that methods grounded in static similarity metrics are unlikely to generalize across time.

Future work may focus on adapting similarity-based learning to account for temporal evolution in asset behavior, potentially through dynamic distance measures or state-dependent representations. For now, this analysis highlights an important boundary of applicability: simple nonparametric algorithms such as kNN may describe what has happened in markets, but they provide little information about what is likely to happen next.

Appendix A: Additional Evaluation Metrics

This appendix reports complementary performance measures beyond hedge effectiveness, including tracking error, maximum drawdown reduction, 95% Value-at-Risk reduction, and hit rate. All metrics are evaluated over identical grids of neighbor count k, half-life, and evaluation horizon.

A.1 Tracking Error

Tracking error quantifies the volatility of residual returns between the target and its hedge. It decreases with larger k and longer horizons, reflecting diversification and temporal smoothing rather than improved hedge performance.

Table 5: Mean Tracking Error by Neighbor Count k and Half-Life.

k	HL=5	HL=30	HL=90	HL=180	HL=365	HL=720
1	0.569	0.580	0.585	0.593	0.599	0.603
3	0.474	0.469	0.473	0.478	0.481	0.483
5	0.445	0.437	0.441	0.444	0.447	0.449
10	0.421	0.412	0.415	0.417	0.418	0.420
30	0.402	0.396	0.396	0.396	0.398	0.399
60	0.397	0.393	0.393	0.393	0.394	0.395

Table 6: Mean Tracking Error by Half-Life and Evaluation Horizon.

Half-Life	5-day	21-day	126-day	252-day
5	0.380	0.443	0.478	0.507
30	0.378	0.443	0.476	0.500
90	0.378	0.444	0.478	0.504
180	0.382	0.448	0.481	0.507
365	0.385	0.451	0.484	0.509
720	0.386	0.452	0.486	0.512

Table 7: Mean Tracking Error by Neighbor Count k and Evaluation Horizon.

k	5-day	21-day	126-day	252-day
1	0.490	0.586	0.627	0.654
3	0.401	0.472	0.505	0.530
5	0.375	0.438	0.470	0.495
10	0.354	0.410	0.441	0.466
30	0.336	0.390	0.421	0.448
60	0.332	0.385	0.418	0.446

A.2 Maximum Drawdown Reduction

Maximum drawdown reductions are negative across all configurations. Portfolios with fewer neighbors suffer the deepest losses, while larger k values reduce their magnitude but not their sign.

Table 8: Mean Maximum Drawdown Reduction by Neighbor Count k and Half-Life.

k	HL=5	HL=30	HL=90	HL=180	HL=365	HL=720
1	-1.185	-1.089	-1.161	-1.219	-1.236	-1.265
3	-0.632	-0.571	-0.599	-0.614	-0.634	-0.628
5	-0.468	-0.399	-0.419	-0.425	-0.439	-0.440
10	-0.320	-0.257	-0.248	-0.261	-0.271	-0.275
30	-0.190	-0.149	-0.135	-0.141	-0.152	-0.159
60	-0.156	-0.125	-0.113	-0.118	-0.125	-0.132

Table 9: Mean Maximum Drawdown Reduction by Half-Life and Evaluation Horizon.

Half-Life	5-day	21-day	126-day	252-day
5	-0.802	-0.458	-0.365	-0.332
30	-0.703	-0.418	-0.324	-0.271
90	-0.710	-0.439	-0.334	-0.291
180	-0.746	-0.446	-0.347	-0.303
365	-0.771	-0.462	-0.353	-0.309
720	-0.762	-0.470	-0.368	-0.321

Table 10: Mean Maximum Drawdown Reduction by Neighbor Count k and Evaluation Horizon.

k	5-day	21-day	126-day	252-day
1	-1.734	-1.148	-0.990	-0.878
3	-0.961	-0.609	-0.469	-0.399
5	-0.718	-0.425	-0.313	-0.260
10	-0.493	-0.261	-0.180	-0.146
30	-0.316	-0.138	-0.079	-0.080
60	-0.274	-0.111	-0.061	-0.062

A.3 Value-at-Risk Reduction

VaR_{95} reductions mirror the drawdown results. Short-horizon and small-k hedges show strongly negative values, while larger k and longer horizons bring results closer to zero or marginally positive.

Table 11: Mean 95% Value-at-Risk Reduction by Neighbor Count k and Half-Life.

k	HL=5	HL=30	HL=90	HL=180	HL=365	HL=720
1	-1.250	-1.117	-1.015	-1.238	-1.264	-1.277
3	-0.793	-0.583	-0.552	-0.601	-0.611	-0.591
5	-0.643	-0.449	-0.428	-0.419	-0.365	-0.350
10	-0.476	-0.296	-0.303	-0.272	-0.234	-0.238
30	-0.208	-0.163	-0.155	-0.163	-0.149	-0.154
60	-0.184	-0.131	-0.135	-0.148	-0.104	-0.107

Table 12: Mean 95% Value-at-Risk Reduction by Half-Life and Evaluation Horizon.

Half-Life	5-day	21-day	126-day	252-day
5	-1.632	-0.321	-0.199	-0.192
30	-1.230	-0.267	-0.174	-0.134
90	-1.085	-0.277	-0.184	-0.162
180	-1.225	-0.286	-0.191	-0.173
365	-1.137	-0.294	-0.196	-0.172
720	-1.097	-0.304	-0.207	-0.185

Table 13: Mean 95% Value-at-Risk Reduction by Neighbor Count k and Evaluation Horizon.

k	5-day	21-day	126-day	252-day
1	-2.480	-0.890	-0.706	-0.663
3	-1.498	-0.419	-0.288	-0.258
5	-1.172	-0.271	-0.165	-0.141
10	-0.961	-0.140	-0.060	-0.036
30	-0.678	-0.027	0.025	0.033
60	-0.616	-0.002	0.044	0.047

A.4 Hit Rate

The hit rate, the share of periods where the hedge outperforms the target, remains near 0.49 for all horizons and k values. It shows no sensitivity to parameter choice, confirming that the model provides no directional advantage and performs indistinguishably from chance.

Table 14: Mean Hit Rate by Neighbor Count k and Evaluation Horizon.

k	5-day	21-day	126-day	252-day
1	0.477	0.478	0.480	0.474
3	0.498	0.497	0.497	0.494
5	0.496	0.497	0.497	0.495
10	0.498	0.496	0.496	0.493
30	0.497	0.493	0.493	0.489
60	0.493	0.490	0.491	0.487

Across all additional metrics, the results are consistent with those presented in the main text. No configuration or horizon produces statistically meaningful improvement, indicating that KNN-based hedges neither enhance stability nor reduce downside or tail risk. Performance remains near random across all dimensions.

References

[1] Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494, 2002. doi: 10.1016/S0304-405X(02)00068-5. URL `https://business.columbia.edu/sites/default/files-efs/pubfiles/1516/corr.pdf`.

[2] Ran Aroussi. yfinance: Download market data from yahoo! finance. `https://github.com/ranaroussi/yfinance`, 2017. Accessed October 2025.

[3] David H. Bailey, Jonathan M. Borwein, Marcos López de Prado, and Qiji Jim Zhu. The probability of backtest overfitting. *Notices of the American Mathematical Society*, 61(5):458–471, 2014. URL `https://www.ams.org/notices/201405/rnoti-p458.pdf`.

[4] Marcos López de Prado. *Advances in Financial Machine Learning*. Wiley, 2018. URL `https://archive.org/download/massimo_motta_competition_policy_theory_and_prabookfi-org/Marcos%20Lopez%20de%20Prado%20-%20Advances%20in%20Financial%20Machine%20Learning-Wiley%20%282018%29.pdf`.

[5] Robert F. Engle. Dynamic conditional correlation: A simple class of multivariate GARCH models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002. URL `https://faculty.washington.edu/ezivot//econ589/EngleDCCJBES.pdf`.

[6] Evan Gatev, William N. Goetzmann, and K. Geert Rouwenhorst. Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, 19(3):797–827, 2006. URL `https://repec.som.yale.edu/icfpub/publications/2573.pdf`.

[7] Investing.com and Nasdaq Filings. Reverse stock split announcements: Elevai labs (1-for-200, 2024) and golden heaven group (1-for-25, 2025). https://ph.investing.com/news/sec-filings/elevai-labs-announces-1for200-reverse-stock-split-93CH-1554478, 2025. Accessed October 2025.

[8] J.P. Morgan and Reuters. Riskmetrics technical document, fourth edition. Technical report, J.P. Morgan, 1996. URL https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a.

[9] J.P. Morgan/Reuters. Riskmetrics technical document. https://www.msci.com/www/research-report/1996-riskmetrics-technical/018482266, 1996. Fourth Edition, December 1996.

[10] Bryan T. Kelly, Seth Pruitt, and Yinan Su. Financial machine learning. Technical report, Becker Friedman Institute Working Paper 2023-100, 2023. URL https://bfi.uchicago.edu/wp-content/uploads/2023/07/BFI_WP_2023-100.pdf.

[11] Laurent Laloux, Pierre Cizeau, Jean-Philippe Bouchaud, and Marc Potters. Noise dressing of financial correlation matrices. *Physical Review Letters*, 83(7):1467–1470, 1999. doi: 10.1103/PhysRevLett.83.1467. URL https://arxiv.org/abs/cond-mat/9810255.

[12] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004. doi: 10.3905/jpm.2004.110. URL https://www.pm-research.com/content/iijpormgmt/30/4/110.

[13] François Longin and Bruno Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676, 2001. doi: 10.1111/0022-1082.00340. URL https://longin.fr/Recherche_Publications/Articles_pdf/Longin_Solnik_Extreme_corelation_of_international_equity_market.pdf.

[14] Rosario N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B*, 11(1):193–197, 1999. doi: 10.1007/s100510050929. URL https://link.springer.com/article/10.1007/s100510050929.

[15] Rosario N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B*, 11:193–197, 1999. doi: 10.1007/s100510050929.

[16] Vasiliki Plerou, Parameswaran Gopikrishnan, Bernd Rosenow, Luis A. Nunes Amaral, and H. Eugene Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126, 2002. doi: 10.1103/PhysRevE.65.066126. URL https://link.aps.org/doi/10.1103/PhysRevE.65.066126.