Journal of Mathematics and Modeling in Finance (JMMF)
Vol. 5, No. 2, Summer & Fall 2025
Research paper

A Comparison Between Behavioral Similarity Methods vs Standard Deviation Method in Predicting Time Series Dataset, Case Study of Finance Market

Mahdi Goldani[1]

[1] Faculty of Literature and Humanities, Hakim Sabzevari University, Sabzevar, Iran
mahgoldani@gmail.com

Abstract:
In statistical modeling, prediction and explanation are two fundamental objectives, particularly in financial time series forecasting where uncertainty quantification is critical. This study compares traditional standard deviation–based confidence intervals with similarity-based methods—Dynamic Time Warping (DTW), Longest Common Subsequence (LCSS), Hausdorff, TWED, and Fréchet distances—using data from 42 U.S. mega-cap companies in the technology and consumer sectors (April 2020–April 2025). Ridge Regression with lagged features was applied to address multicollinearity among predictors. Results show that σ-based and LCSS methods achieved the highest coverage (95.22% and 94.61%) but at the cost of wider intervals, while DTW, Hausdorff, and TWED provided much narrower intervals (5.86–6.48) with moderate coverage (63–67%). These findings highlight a trade-off between reliability and precision, underscoring the need for context-aware method selection. This work adds to the literature by demonstrating that similarity-based approaches can offer competitive, application-dependent alternatives to conventional interval estimation in high-dimensional, nonstationary financial data.

Keywords: Confidence interval, financial forecasting, Prediction interval, Ridge regression, Similarity-based methods, Time series analysis, Uncertainty quantification
Classification: JEL Classifications: C22, C13, G17.

1 Introduction

Real-life situations encompass uncertainty in various domainsfrom financial investment and medical diagnosis to sporting events and weather prediction. In each instance, the objective is to make smart decisions based on accessible information and subject expertise amidst underlying uncertainties [1]. Scholars have suggested numerous models, such as regression, machine learning, and neural networks, to obtain precise predictions. Nonetheless, even the best approaches are incapable of predicting precise values with certainty; there is always inherent uncertainty. Consequently, given that predictions by such models are prone to noise and inference error, uncertainty quantification (UQ) becomes necessary [2].

[1] Corresponding author
Received: 27/06/2025 Accepted: 08/09/2025
https://doi.org/10.22054/JMMF.2025.86596.1193

To measure this prediction uncertainty, measures like Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) are routinely employed, which calculate the error between predictions and realizations. Prediction intervals (PIs) are one of the most popular methods of uncertainty quantification (UQ) in the literature [3]. Prediction intervals (PIs) have been used to measure uncertainty for more than 50 years extensively. A prediction interval gives a range that is expected to cover future observations, hence quantifying prediction uncertainty. This method overcomes the limitations of point forecasting and is now critical in economics, food science, tourism, healthcare, energy, and even compression algorithms [4]. In spite of their importance, evaluation and optimization of prediction intervals are still areas worth investigating.

Optimal confidence intervals are necessary for accurate parameter estimation in various fields, including reliability engineering, diagnostics, and economic forecasting. Optimal confidence intervals (CIs) are defined by their high coverage probability (e.g., 95%) and small width. Wider intervals reflect greater uncertainty about the estimated parameter [5]. The bootstrap weighted-norm method provides an effective strategy for the accurate estimation of confidence intervals for parameter estimation [6]. For diagnostic studies, the delta method and parametric bootstraps are suitable for post-hoc calculation of confidence intervals, depending on variables like sample size, distribution of marker values, and model assumption validity [7]. An approach based on the moderate deviation principle also provides statistically optimal confidence intervals for non-parametric estimates, thus guaranteeing a minimum mischaracterization and strong performance under a variety of models [8].

Collectively, these studies highlight the importance of tailored methods for the development of valid confidence intervals for a spectrum of applications. Employment of time series similarity techniques in prediction tasks is not a new concept. Wu et al. [41] used an ellipse-based similarity metric to model positional uncertainty in degradation paths for the prediction of the Remaining Useful Lifetime (RUL) of lithium-ion batteries. The proposed approach showed greater accuracy and robustness than conventional similarity metrics like DTW, LCSS, and EDR. Among the traditional methods, the LCSS metric performed better than DTW and EDR but was less accurate than the proposed approach. Further, Wei et al. [10] developed novel similarity measures for Probabilistic Interval Preference Ordering Sets (PIPOSs) to improve decision-making under uncertainty. The proposed model succeeds in aggregating both probabilistic and interval-based information, improving the robustness of decision results and mitigating the risk of misleading rankings. Zhao et al. [11] discussed uncertainty quantification in predictive modeling. Similarity-based approaches (SIS, DTW, LCSS) offer a promising alternative to classical σ-based intervals, particularly for noisy, sparse, and nonstationary data. Results indicated that conventional approaches yield wider bounds and conservative predictions. Additionally, Arslan et al. [12] presented an extensive review of time series forecasting approaches based on trajectory similarity with specific application to traffic flow forecasting. Trajectory similarity approaches are effective in both point and interval forecasting, especially when coupled with seasonal filtering and local regression. Literature proves that similarity-based approaches significantly contribute to enhancing both prediction accuracy and uncertainty quantification. This highlights a clear gap: the need to systematically develop and assess similarity-based interval methods that balance theoretical rigor with practical performance.

The aim of this study is to employ similarity-based methods in constructing confidence intervals, thereby avoiding the parametric assumptions usually required by conventional approaches, founded on the importance of similarity measures in enhancing the accuracy of predictive procedures. While we acknowledge that conventional confidence intervals are statistically grounded and similarity-based intervals are algorithmic in nature, a meaningful comparison remains possible when the focus is on their practical performance in quantifying uncertainty. Specifically, both approaches share the common objective of capturing the variability around predictions and providing bounds within which future values are expected to fall. Thus, even though the theoretical underpinnings differ, their outcomes can be evaluated side by side in terms of coverage probability and interval widthtwo universally accepted criteria for assessing the effectiveness of interval estimation [13].

1.1 Theoretical Foundation

Normal or approximately normal sampling distribution and constant variance are assumptions of conventional confidence intervals. However, in real-world data, these assumptions are often violated. Nonparametric methods, particularly distance-based approaches, offer notable advantages by eliminating the need for strict distributional assumptions, which makes them robust for real-world datasets. These kinds of methods are highly flexible, accommodating irregular, noisy, or high-dimensional time series, and remain valid even with small sample sizes, unlike parametric methods that rely on large-sample approximations [11].

In regression-based forecasting, each model prediction $\hat{y}_t$ is subject to uncertainty due to model bias, noise in the explanatory variables, and inherent randomness in the data-generating process. A confidence interval (CI) for a prediction is a range $[L_t, U_t]$ that aims to contain the true value y_t with a specified probability $1 - \alpha$:

$$P(L_t \leq y_t \leq U_t) \geq 1 - \alpha$$

In the classical framework, this range is obtained from the sampling distribution of the prediction error

$$e_t = y_t - \hat{y}_t.$$

The scale of this distribution, often summarized by its standard deviation σ_e, is used in:

$$[L_t, U_t] = \hat{y}_t \pm z_{1-\alpha/2} \cdot \sigma_e$$

Classical confidence intervals are based on the presupposition of constant variance combined with a normal or near-normal sampling distribution, which makes them vulnerable to skewness, heavy tails, and small sample sizes for which the Central Limit Theorem does not hold [14]. The substantial effect of extreme values on variance means that even one outlier can disproportionately inflate the width of the interval and distort the image of uncertainty [11]. In datasets with few observations, the estimation of σ is no longer reliable, while in time series that have heteroscedasticity or nonstationarity, one global σ cannot capture local variability, with the result being coverage inaccuracies [12]. The standard deviation also condenses data into a single measure of dispersion that ignores temporal structures and local patterns of similarity, relying on particular distributional assumptions, which makes it inappropriate when there is no reliable probabilistic model available or when only bounded interval data are available [11].

Nonparametric and similarity-based approaches form prediction intervals without making distributional assumptions; rather than estimating a universal standard deviation (σ), these approaches compare the feature patterns in the current data point to comparable patterns in historical datasets (e.g., Dynamic Time Warping, Longest Common Subsequence, or Euclidean distance) to induce variability from these similarities [12]. By concentrating on a suitably chosen subset of comparable historical cases, such approaches naturally deal with heteroscedasticity, with interval widths varying smoothly across contexts, mitigating the effect of outliers, avoiding probabilistic assumptions, and quantifying uncertainty through mechanisms such as Sub-Interval Similarity, which measures distances between predicted and actual sub-intervals [11]. Further, these approaches preserve the temporal structure of the data, in contrast to variance-based approaches that collapse it to a single dispersion measure [12].

Similarity-based methods extend the confidence interval concept by replacing the parametric error variance with an empirical scale derived from the distances between predicted sequences and their most similar historical analogs. Let $S(\mathrm{y}_{\mathrm{pred}}, \mathrm{y}_{\mathrm{obs}})$ be a similarity measure (e.g., DTW, LCSS, Hausdorff). A similarity-based error estimator can be expressed as:

$$\sigma_{e,\mathrm{sim}} = g^{-1}\left(\frac{1}{k}\sum_{i=1}^{k} S\big(y_{t-w:t}, y^{i}_{t-w:t}\big)\right)$$

where k is the number of nearest neighbors in similarity space and $g^{-1}(0)$ maps the average similarity distance to an equivalent prediction error magnitude. Under assumptions of stationarity and ergodicity, the distribution of historical distances provides a consistent estimator for the future error distribution.

The resulting prediction interval is therefore:

$$[L_t, U_t] = \hat{y}_t \pm z_{1-\alpha/2} \cdot \sigma_{e,\mathrm{sim}}$$

This specific formulation preserves the meaning of coverage probability with confidence intervals while abandoning the requirement for strict parametric assumptions. Also, since similarity measures can capture structural differences, such as local time shifts, non-linear warping, and geometric distortions, they can provide more robust uncertainty bounds in complex, high-dimensional data sets [15].

Therefore, similarity-based confidence intervals are not heuristic approximations but are based on the same probabilistic basis as conventional CIs, except that the error scale is estimated from empirical similarity relations instead of parametric distribution. This distribution-free characteristic renders them particularly appropriate for non-stationary or highly structured time series data.

2 Methodology

The methodology of this study comprises two main components: the database and the prediction method. It further elaborates on the similarity-based methodology and its application in measuring intervals.

2.1 Dataset

This study is grounded in the U.S. finance market, covering the period from April 7, 2020, to April 4, 2025. The closing prices of companies classified under the Mega Market Capitalization category were used. Apples closing price was selected as the target value because its stock performance often reflects broader market trends, making it a strong representative of market sentiment. The data were extracted from the NASDAQ database. Using the Stock Screener tool on the website, all active U.S. stock market companies within the Mega Cap category were filtered, resulting in a selection of 42 companies. The dataset spans the last five years with daily frequency, allowing for precise analysis of market fluctuations and supporting accurate statistical analyses and algorithmic applications.

2.2 Prediction model

Forecasting in finance involves a high number of variables, such as macroeconomic data, microeconomic data, earnings reports, and technical indicators. Multicollinearity and dependencies among predictors are common in financial datasets, which makes the use of multivariate regression models like OLS less appropriate or reliable [16]. From a technical standpoint in linear regression with multicollinearity issues, when the matrix $X^T X$ is singular, the standard Ordinary Least Squares (OLS) estimators cannot be applied because they require the inverse of $X^T X$. In this case, the ridge version of the estimators is more stable and can overcome this problem [17]. Multicollinearity can cause issues for each coefficient, including inaccurate estimates, excessive growth of standard deviations, incorrect t-tests, and inaccurate confidence intervals [18].

Ridge estimation is a regularization technique aimed at stabilizing parameter estimates by shrinking them or their linear combinations. This method provides more reliable estimators with reduced variance compared to ordinary least squares, especially in the presence of multicollinearity. By adding the parameter k (where $k > 0$) to the $X^T X$ matrix, one can control the amount of shrinkage of the regression coefficients:

$$\hat{\beta}_k = (X^T X + kI)^{-1} X^T Y \tag{1}$$

Here, k is the shrinkage parameter, which controls the amount of shrinkage of the regression coefficients, and I is the identity matrix [19]. As seen in Equation (5), the ridge regression approach involves adding a small positive number (k) to the diagonal elements of the $X^T X$ matrix. This prevents the variance of the regression coefficients from overinflating due to multicollinearity.

2.3 Confidence intervals

Predictions are inherently associated with uncertainty. Unlike point forecasts, prediction intervals (PIs) serve as a powerful tool for modeling uncertainty. By definition, a PI consists of lower and upper bounds that bracket an unknown future value with a specified probability typically expressed as a confidence level of $(1-\alpha)\%$ [20]. A desirable prediction method achieves narrower intervals while controlling Type I (false positive) and Type II (false negative) error probabilities.

The traditional form of confidence intervals uses the standard deviation to evaluate the CI. The margin of error (ME) in a confidence interval is determined using the Z-value, the standard deviation (SD) of the sample, and the sample size (N), and is given by the formula:

$$ME = Z \times \frac{SD}{\sqrt{N}} \tag{2}$$

The lower and upper bounds of a prediction interval are obtained by subtracting and adding the margin of error (ME) to the predicted value, respectively. If $\hat{y}$ is the predicted value:

$$\text{lower bound} = \hat{y} - ME \tag{3}$$

$$\text{upper bound} = \hat{y} + ME \tag{4}$$

So, the prediction interval is:

$$PI = (\hat{y} - ME, \hat{y} + ME) \tag{5}$$

The Z-value is determined by the chosen confidence level. It is important to note that confidence intervals are statistically accurate only when sampling from a normally distributed population. For non-normally distributed populations, they become approximately valid when the sample size is sufficiently large [21].

2.4 Distance-based methods

The construction of optimal confidence intervals (with minimal width and correct coverage probability) as a statistical method relies critically on certain assumptions, namely normally distributed data and sufficiently large samples. Violating these assumptions compromises either coverage accuracy or interval efficiency. This article introduces distribution-free methods for confidence interval estimation that do not require parametric assumptions (e.g., normality or large-sample approximations), instead leveraging distance-based metrics to achieve statistically valid inference. Although we refer to the intervals generated through similarity-based methods as "bounds" for brevity, they are not statistical confidence intervals derived from a sampling distribution. Rather, these are heuristic intervals based on the variability of historical instances of similar patterns. Thus, their coverage is empirical and not supported by a probability model.

Building on the same principles underlying conventional methods for calculating confidence intervals, distance-based approaches offer an alternative by measuring the distance between predicted values and observed data. Time series similarity measurement serves as the foundation for clustering and classification tasks by quantifying distances between temporal sequences. This metric plays a critical role in temporal pattern analysis, functioning as a fundamental tool for statistical inference across datasets. The rapid proliferation of data collection has significantly expanded time series availability, increasing demand for analytical tasks such as regression, classification, clustering, and segmentation. These applications universally require specialized distance metrics to quantify inter-series similarity, making methodological research in this area essential [15].

Current similarity measures can be broadly categorized into three groups:

- Step-by-step methods (e.g., pointwise comparisons)
- Distribution-based approaches (matching statistical properties)
- Geometric techniques (shape or trajectory alignment)

Stepwise Metrics

These metrics compare time-series samples one by one based on their time indices [**?**]. A significant limitation of these methods is the requirement for identical sample sizes in the time series. The most notable stepwise metrics are Euclidean Distance and Correlation Coefficient, which are detailed below.

The Euclidean Distance calculates the shortest distance between two points in Euclidean space. For two time series x and y of length n, it is defined as:

$$D_{\text{euc}} = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2} \tag{6}$$

This distance is widely used due to its simplicity and ease of understanding. However, a key limitation is its sensitivity to time-axis transformations, such as scaling and shifting [22], and it cannot compare time series of different lengths.

Elastic Metrics

Elastic metrics adjust the time axis by stretching or compressing it to minimize the effect of local variations, effectively handling non-linear distortions. The most notable methods include Dynamic Time Warping (DTW) and Longest Common Subsequence (LCSS).

Dynamic Time Warping (DTW) aligns sequences non-linearly by stretching or compressing the time axis. The cumulative distance matrix is defined as:

$$\text{DISTMATRIX} = \begin{bmatrix} d(x_1, y_1) & d(x_1, y_2) & \dots & d(x_1, y_m) \\ d(x_2, y_1) & d(x_2, y_2) & \dots & d(x_2, y_m) \\ \vdots & \vdots & \ddots & \vdots \\ d(x_n, y_1) & d(x_n, y_2) & \dots & d(x_n, y_m) \end{bmatrix} \tag{7}$$

and recursively:

$$\begin{cases} r(i,j) = d(i,j) + \min\{r(i-1,j), r(i,j-1), r(i-1,j-1)\} \\ DTW(x,y) = \min\{r(n,m)\} \end{cases} \tag{8}$$

Longest Common Subsequence (LCSS) identifies the longest matching subsequences between two time series S_x and S_y of lengths n and m:

$$M(i,j) = \begin{cases} 0 & i = 0 \text{ or } j = 0 \\ 1 + M(i-1, j-1) & x_i = y_j, i \geq 1, j \geq 1 \\ \max\{M(i-1,j), M(i,j-1)\} & x_i \neq y_j, i \geq 1, j \geq 1 \end{cases} \tag{9}$$

and recursively with a tolerance ϵ:

$$M(i,j) = \begin{cases} 0 & i = 0 \text{ or } j = 0 \\ 1 + M(i-1, j-1) & |x_i - y_j| \leq \epsilon, i \geq 1, j \geq 1 \\ \max\{M(i-1,j), M(i,j-1)\} & |x_i - y_j| > \epsilon, i \geq 1, j \geq 1 \end{cases} \tag{10}$$

Geometric Distances

Geometric distances focus on the spatial characteristics of trajectories. Examples include Hausdorff Distance, Discrete Frechet Distance, and SSPD (Symmetric Segment Path Distance).

The **Hausdorff Distance** is defined as:

$$\text{Haus}(X, Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y}\|x-y\|_2, \sup_{y\in Y}\inf_{x\in X}\|x-y\|_2\right\} \tag{11}$$

The **Frechet Distance** measures the minimal "leash length" connecting two curves:

$$D_{\text{Frechet}}(T^1, T^2) = \min\left\{\max_k \|w_k\|_2\right\}, \quad w_k \in [0, \ldots, |w|] \tag{12}$$

2.5 Coverage and Width Calculation

To compare the confidence intervals, the coverage and width of each interval are calculated as evaluation metrics.

Coverage is the empirical inclusion rate of observed values within the predicted bounds. Higher coverage indicates more accurate intervals. Let L_i and U_i be the lower and upper bounds, and y_i be the true observed value. Then the coverage is calculated as:

$$\text{Coverage} = \frac{\sum_{i=1}^{n} 1(L_i \le y_i \le U_i)}{n} \times 100\% \tag{13}$$

If the condition holds, the 1(0)-function equals 1; otherwise it takes 0.

For σ-based intervals, coverage can be interpreted as an approximation of the nominal statistical confidence level under the assumption of normality. For similarity-based bounds (e.g., DTW-like), coverage is a purely empirical metric showing the proportion of test observations contained within the bounds; it does not imply any formal probabilistic guarantee, as these intervals are not derived from a sampling distribution.

Width is computed as the average span of the predicted intervals:

$$\text{Mean Width} = \frac{\sum_{i=1}^{n}(U_i - L_i)}{n} \tag{14}$$

This measures precision: narrower intervals indicate higher precision but may lead to lower coverage, while wider intervals indicate greater uncertainty representation but may be less useful for decision-making.

For similarity-based bounds, which lack a formal probabilistic model, performance is evaluated using empirical coverage and mean interval width. Coverage is defined as the proportion of observed values contained within the bounds, while

Table 1: Similarity methods

Method	**Advantages**	**Disadvantages**	**Category**
Euclidean Distance	• Most straightforward and widely used • No need for parameter estimation	• No support for local time shifts • Inefficient in high dimensions • Sensitive to small time axis changes	Step-by-step
DTW (Dynamic Time Warping)	• Supports local scaling and order preservation • Handles time series of different lengths • Captures local time shifts	• Time-consuming • Sensitive to noise • High computational load • Incorrect clustering due to outliers • Requires pairing all elements • Not metric	Elastic
LCSS (Longest Common Subsequence)	• Robust against noise • Focuses on similar parts • No need for normalization	• Depends on threshold • Binary similarity can cause poor results • Not metric, violates triangle inequality	Elastic
Hausdorff Distance	• Measures spatial similarity • Considers farthest point	• Not suitable for trends • Complex due to all points • Limited path comparison • Ignores overall similarity	Geometric
Discrete Fréchet Distance	• Considers order and continuity • Lower complexity with discrete models	• Limited to path comparison • Max distance overshadows details	Geometric

width quantifies precision. This combination offers an assumption-free and directly interpretable assessment suitable for nonparametric, data-driven methods.

3 Result

The financial dataset was used to develop prediction models. To optimize model performance, the Ridge parameter (α) of the robust Rank Ridge Regression model was determined using RidgeCV, which minimizes the mean squared error (MSE).

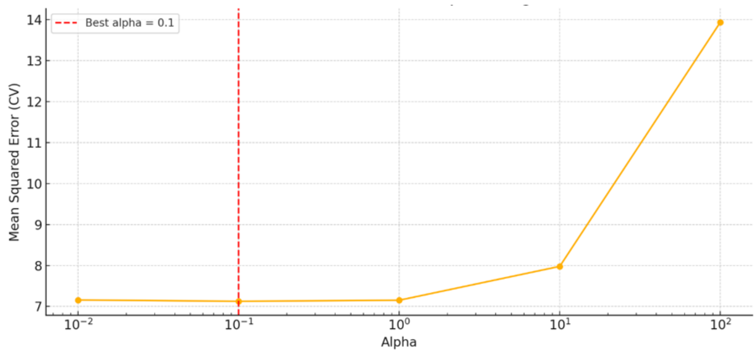

Figure 1: Cross-Validation MSE vs Alpha (RidgeCV)

Figure 1 illustrates the optimal value of this parameter. The Ridge Regression model was developed to forecast Apples daily closing price. To improve model accuracy, three time-lag features were created for the closing prices of both Apple and the predictor variables. The dataset was divided into training and test sets, and the features were standardized to enhance model performance.

Table 2: Ridge Regression with Lag Features

Metric	**Train**	**Test**
R^2	0.9923	0.9604
MAE	1.9130	3.2633
RMSE	2.5483	4.3133

Model accuracy was evaluated using the Mean Squared Error (MSE) and the R^2 score, as presented in Table 2. The R^2 and Mean Absolute Error (MAE) values for both the training and test datasets indicate strong model performance. The high R^2 value for the test set suggests that the model generalizes well and does not suffer from overfitting.

Based on the Ridge Regression model, Figure 2 presents the prediction results for the AAPL test dataset. The blue line represents the actual test values, while the orange line represents the predicted values. Since Ridge Regression is particularly suitable for data affected by multicollinearity, the accurate predictions on the test setillustrated in Figure 2demonstrate the strong performance of the model.

Coverage and width are the two primary properties that determine the effectiveness of confidence intervals. In this study, the models confidence intervals were computed using both the standard deviation and time series similarity methods to evaluate which technique better satisfies the desired properties of accurate confidence intervals.

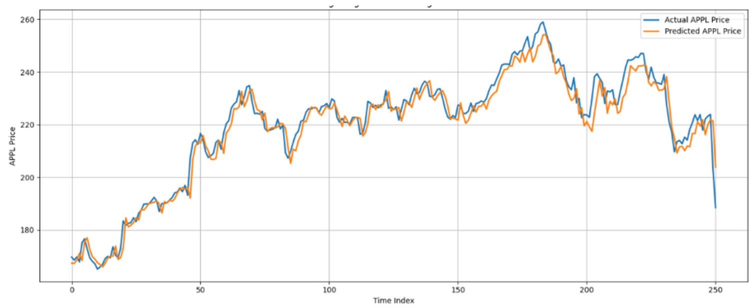

Figure 2: Ridge Regression with Lag Features

Traditional confidence intervals are derived from the standard deviation, assuming a 95% confidence level. For the dataset used in this study, the coverage of the conventional confidence interval is 95.22%, meaning that in 95.22% of cases the actual value falls within the interval. Another important property for evaluating the effectiveness of a confidence interval is the mean CI width; a narrower width indicates more precise intervals. For our dataset, the mean CI width of the conventional approach is 14.57.

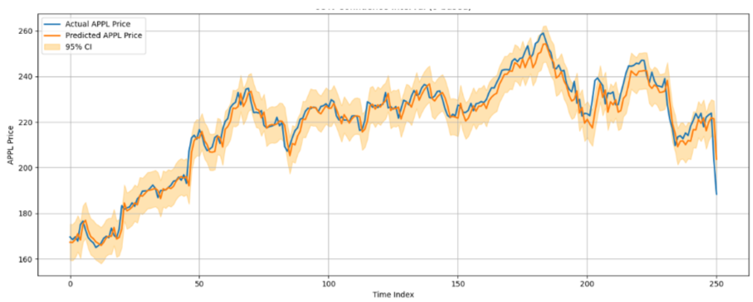

Figure 3: conventional confidence interval with standard deviation

The main question of this research is: Are time series similarity methods effective enough to construct reliable intervals comparable to those based on standard deviation? To evaluate this, confidence intervals were constructed using various time series similarity methods, including Dynamic Time Warping (DTW), Longest Common Subsequence (LCSS), Hausdorff distance, Time Warp Edit Distance (TWED), and Fréchet distance.

Figures 4, 5, 6, 7, and 8 illustrates the intervals generated by each method, highlighting the extent to which the actual values are covered within these intervals.

Table 3 summarizes the coverage and mean confidence interval width associated with each method.

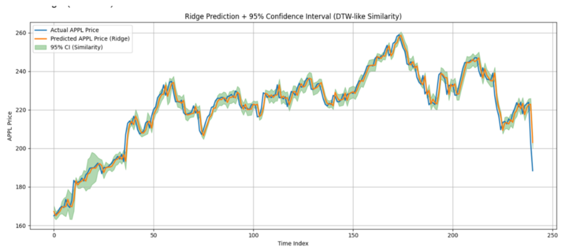

Figure 4: DTW intervals.

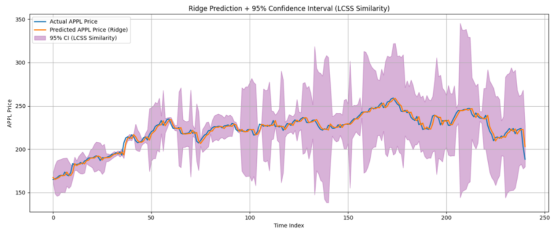

Figure 5: LCSS intervals.

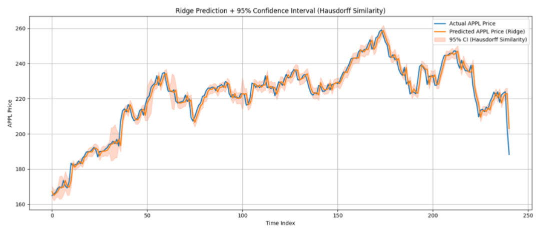

Figure 6: Hausdorff intervals.

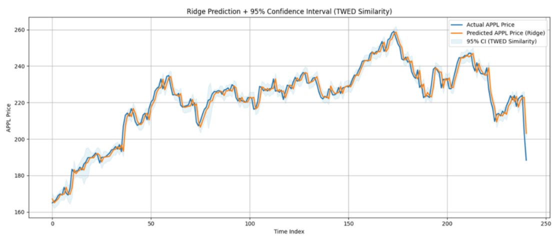

Figure 7: TWED intervals.

4 Discussion

Table 3 highlights the trade-off between reliability (measured by coverage) and precision (measured by mean confidence interval width) across six different methods

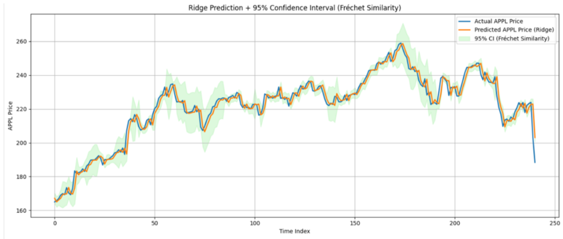

Figure 8: Fréchet intervals.

Table 3: Evaluation of CI Coverage and Width Using Conventional and Time Series Similarity Approaches

Method	**CI Coverage (%)**	**Mean CI Width**
Conventional	95.22	14.57
DTW-like	63.07	5.86
LCSS	94.61	59.34
Hausdorff	63.49	6.22
TWED	66.8	6.48
Fréchet	75.1	10.68

for constructing confidence intervals (CIs). This study compares the performance of these methods in capturing true values within the estimated intervals.

The first method, a σ-based approach (commonly referred to as the conventional confidence interval), demonstrates strong performance with a coverage rate of 95.22% and a moderate mean width of 14.57. These results are consistent with the findings of [11], which showed that traditional methods gain wider bounds and conservative estimates. Among all the methods evaluated, this approach yields the highest coverage, indicating that it reliably includes the actual values within the interval. However, while reliable, its precision is only moderate compared to some of the other methods.

The LCSS (Longest Common Subsequence) method also achieves high coverage at 94.61%, closely matching the conventional method. [41] already showed the accuracy and robustness of the LCSS method. However, it produces significantly wider intervals (mean width of 59.34), suggesting that while it captures the true value effectively, it lacks precision and may result in overly conservative estimates.

In contrast, the DTW-like, Hausdorff, and TWED methods generate much narrower confidence intervals (with mean widths of 5.86, 6.22, and 6.48, respectively), which indicates greater precision. However, this precision comes at the cost of lower coverageranging from 63.07% to 66.8%which implies that these methods may fail to capture the actual values as consistently as the σ-based or LCSS methods.

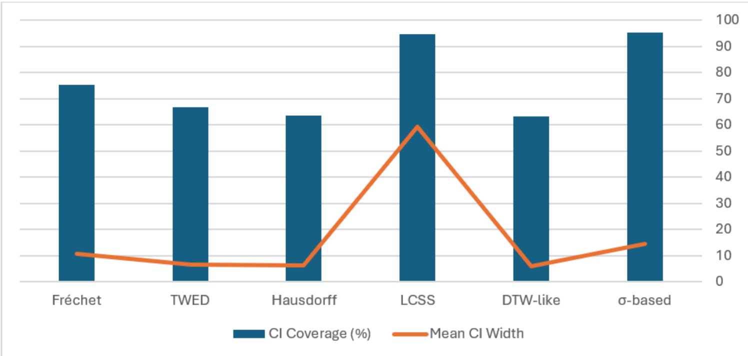

Figure 9: Figure 5. Comparison of Confidence Interval Coverage and Width Across Conventional and Time Series Similarity

The Fréchet distance-based method offers a balance between coverage and precision, with a coverage of 75.1% and a mean width of 10.68. While not as reliable as the σ-based or LCSS approaches, it outperforms the other similarity-based methods in terms of maintaining a more favorable balance.

While similarity-based approachessuch as those proposed by [11], [41], [12], and [10]provide a flexible and powerful framework for prediction and interval estimation, they are not without limitations. These methods overcome key constraints of traditional parametric techniques by offering data-driven uncertainty assessments, but it is critical to note that their resulting intervals are heuristic bounds rather than formal statistical confidence intervals. Consequently, their coverage rates must be interpreted empirically, not probabilistically.

5 Conclusion

In the realm of statistical modeling, two primary objectives—prediction and explanation—guide the analytical process. When forecasting is the focus, it is essential to account for the uncertainties that arise in estimating unknown outcomes. Historically, confidence intervals constructed from standard deviations have provided a structured means of quantifying this uncertainty, enabling an assessment of how closely predicted values align with their actual counterparts. This traditional approach implicitly reflects the behavioral similarities between observed and predicted data points. However, recent advancements in similarity-based methodologies offer innovative alternatives to conventional variance-focused techniques, particularly in contexts characterized by extensive datasets or a large number of explanatory variables [10–12, 41]. This study seeks to explore methods that can effectively reduce uncertainty in confidence interval estimation. By comparing both traditional and similarity-based approaches, the goal is to determine which methods can yield tighter confidence intervals under comparable conditions, ultimately leading to greater precision and more informative results. Addressing uncertainty remains paramount, as it underpins the reliability of predictions and strengthens decision-making processes across diverse applications.

In conclusion, this study highlights the inherent trade-off between reliability and precision in confidence interval construction methods aimed at capturing true values. The evaluation of six distinct approaches reveals clear differences in performance in terms of coverage rates and mean confidence interval widths. The conventional σ-based approach emerges as the front-runner, achieving the highest coverage rate (95.22%), thereby ensuring that true values are reliably included within the intervals. However, this reliability comes at the cost of precision, as the intervals remain relatively wide. The LCSS method follows closely, with a coverage rate of 94.61%, but its considerably wider intervals result in overly conservative estimates [41].

Conversely, methods such as DTW, Hausdorff, and TWED demonstrate exceptional precision, producing notably narrower intervals. Yet this precision is offset by lower coverage rates (ranging from 63.07% to 66.8%), indicating a higher risk of excluding actual values. The Fréchet distance-based method offers a more balanced trade-off, achieving moderate coverage (75.1%) while maintaining narrower interval widths [11, 12].

Overall, these findings underscore the importance of carefully selecting interval construction methods based on the analytical context. Researchers and practitioners must weigh their priorities—whether higher reliability or greater precision—in order to align methodological choices with forecasting objectives. This study emphasizes that no universal solution exists, and the effectiveness of confidence interval estimation must be tailored to the demands of specific applications.

For future research, it would be valuable to explore advanced similarity-based methods that may further enhance the accuracy of confidence intervals in predictive tasks. In particular, the integration of deep learning-based similarity measures—which have demonstrated superior performance in multiple domains—presents a promising direction. Additionally, hybrid approaches incorporating dynamic time warping (DTW) with advanced metric learning algorithms could improve the ability to capture complex temporal patterns and high-dimensional relationships more effectively.

Bibliography

[1] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, ... & S. Nahavandi, *A review of uncertainty quantification in deep learning: Techniques, applications and challenges*, in *Information Fusion*, 76, 243–297, (2021).

[2] S. Cheng, C. Quilodrán-Casas, S. Ouala, A. Farchi, C. Liu, P. Tandeo, ... & R. Arcucci, *Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review*, in *IEEE/CAA Journal of Automatica Sinica*, 10(6), 1361–1387, (2023).

[3] K. Wang, C. Shen, X. Li, & J. Lu, *Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications*, in *IEEE Transactions on Intelligent Transportation Systems*, (2025).

[4] P. Zhang, S. Liu, D. Lu, G. Zhang, & R. Sankaran, *A prediction interval method for uncertainty quantification of regression models*, Oak Ridge National Lab.(ORNL), Oak Ridge, TN, (2021).

[5] E. Nikulchev & A. Chervyakov, *Prediction intervals: A geometric view*, in *Symmetry*, 15(4), 781, (2023).

[6] Y. Cui & M. G. Xie, *Confidence distribution and distribution estimation for modern statistical inference*, in *Springer Handbook of Engineering Statistics*, 575–592, London: Springer London, (2023).

[7] C. Thiele & G. Hirschfeld, *Confidence intervals and sample size planning for optimal cutpoints*, in *PLOS One*, 18(1), e0279693, (2023).

[8] A. Ganguly & T. Sutter, *Optimal learning via moderate deviations theory*, arXiv preprint arXiv:2305.14496, (2023).

[9] W. Wu, T. Zou, L. Zhang, K. Wang, & X. Li, *Similarity-Based Remaining Useful Lifetime Prediction Method Considering Epistemic Uncertainty*, in *Sensors*, 23(23), 9535, (2023).

[10] Q. Wei, R. Wang, & C. Y. Ruan, *Similarity Measures of Probabilistic Interval Preference Ordering Sets and Their Applications in Decision-Making*, in *Mathematics*, 12(20), 3255, (2024).

[11] Y. Zhao, Y. Wang, Z. Wang, & J. Wang, *The sub-interval similarity: A general uncertainty quantification metric*, in *Reliability Engineering & System Safety*, 221, 108316, (2022).

[12] H. Arslan, M. Aslan, & G.-W. Weber, *Distance-based prediction intervals for time series forecasting*, arXiv preprint arXiv:2309.10613, (2023).

[13] G. Shmueli, *To explain or to predict?*, in *Statistical Science*, 25(3), 289–310, (2010).

[14] L. Wasserman, *All of nonparametric statistics*, New York, NY: Springer, (2006).

[15] M. Goldani & S. Asadi Tirvan, *Sensitivity assessing to data volume for forecasting: introducing similarity methods as suitable ones in feature selection methods*, in *Journal of Mathematics and Modeling in Finance*, 4(2), 115–134, (2024).

[16] J. Y. L. Chan, S. M. H. Leow, K. T. Bea, W. K. Cheng, S. W. Phoong, Z. W. Hong, & Y. L. Chen, *Mitigating the multicollinearity problem and its machine learning approach: a review*, in *Mathematics*, 10(8), 1283, (2022).

[17] M. Arashi, M. Roozbeh, N. A. Hamzah, & M. Gasparini, *Ridge regression and its applications in genetic studies*, in *PLOS One*, 16(4), e0245376, (2021).

[18] S. Mermi, Ö. Akku, A. Gökta, & N. Gündüz, *A new robust ridge parameter estimator having no outlier and ensuring normality for linear regression model*, in *Journal of Radiation Research and Applied Sciences*, 17(1), 100788, (2024).

[19] E. Cule & M. De Iorio, *Ridge regression in prediction problems: automatic choice of the ridge parameter*, in *Genetic Epidemiology*, 37(7), 704–714, (2013).

[20] H. Quan, D. Srinivasan & A. Khosravi, *Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals*, in *IEEE Transactions on Neural Networks and Learning Systems*, 25(2), 303–315, (2014).

[21] A. M. Simundic, *Confidence interval*, in *Biochemia Medica*, 18(2), 154–161, (2008).

[22] A. Hazra, *Using the confidence interval confidently*, in *Journal of Thoracic Disease*, 9(10), 4125, (2017).

How to Cite: Mahdi Goldani[1], *A Comparison Between Behavioral Similarity Methods vs Standard Deviation Method in Predicting Time Series Dataset, Case Study of Finance Market*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 5, No. 2, Pages:155–171, (2025).