mathematics

MDPI

Article

Smoothing the Subjective Financial Risk Tolerance: Volatility and Market Implications

Wookjae Heo [1] **and Eunchan Kim** [2,*]

[1] Division of Consumer Science, White Lodging-J.W. Marriot Jr. School of Hospitality & Tourism Management, Purdue University, West Lafayette, IN 47907, USA; heo28@purdue.edu
[2] Department of Information Systems, College of Engineering, Hanyang University, Seoul 04763, Republic of Korea
* Correspondence: eckim@hanyang.ac.kr

Abstract: This study explores smoothing techniques to refine financial risk tolerance (FRT) data for the improved prediction of financial market indicators, including the Volatility Index and S&P 500 ETF. Raw FRT data often contain noise and volatility, obscuring their relationship with market dynamics. Seven smoothing methods were applied to derive smoothed mean and standard deviation values, including exponential smoothing, ARIMA, and Kalman filter. Machine learning models, including support vector machines and neural networks, were used to assess predictive performance. The results demonstrate that smoothed FRT data significantly enhance prediction accuracy, with the smoothed standard deviation offering a more explicit representation of investor risk tolerance fluctuations. These findings highlight the value of smoothing techniques in behavioral finance, providing more reliable insights into market volatility and investor behavior. Smoothed FRT data hold potential for portfolio optimization, risk assessment, and financial decision-making, paving the way for more robust applications in financial modeling.

Keywords: financial risk tolerance; volatility; smoothing; behavioral finance; time series analysis; Kalman Filter; machine learning; market forecasting; portfolio optimization

MSC: 37M10; 68T07

Academic Editor: Davide Valenti

Received: 20 January 2025
Revised: 15 February 2025
Accepted: 18 February 2025
Published: 19 February 2025

Citation: Heo, W.; Kim, E. Smoothing the Subjective Financial Risk Tolerance: Volatility and Market Implications. *Mathematics* **2025**, *13*, 680. https://doi.org/10.3390/math13040680

1. Background

Financial risk tolerance (FRT) is a scale that represents investors' willingness to take investment risks [1]. FRT is often understood as the inverse concept of risk aversion, although it has been shown in previous research that it correlates strongly with risk aversion [2]. This underscores the importance of understanding FRT's connection to market dynamics, a key aspect of investment risk assessment [3,4].

One approach calculates risk aversion using probabilities, while the other uses a psychometric method developed by Grable and Lytton [1] to measure investors' subjective willingness to invest. This psychometric method is comparatively and empirically valid [5]. Through the usefulness of the psychometric measurement of FRT, this study aims to enhance the utility of FRT measured with the psychometric FRT measurement by Grable and Lytton [1], which is a key methodology in this research.

Specifically, while much research considers the correlation between FRT score (i.e., average) and stock volatility [3,4,6,7], a limitation of previous research is that most studies did not consider FRT's volatility. The core concept of FRT is to measure the willingness of investors to tolerate risky investments, but this willingness can fluctuate depending on

market conditions or various external factors [8]. This implies that FRT's volatility may remain due to market conditions and other environmental factors. When volatility is present, there might be a clear distinction between refining the measured FRT signals to observe patterns or to view FRT as it is. Based on this, FRT can be divided into Equations (1)–(3) as follows:

$$FRT_t = FRT_{smooth,t} + FRT_{abnormal,t} + \epsilon_t,\ \epsilon_t \ \sim\ N(0,\ \sigma^2) \quad (1)$$

$$FRT_{smooth,t} = \mu_t + v_t,\ v_t \ \sim\ N(0,\ \sigma^2_{smooth}) \quad (2)$$

$$FRT_{abnormal,t} = \sum_{i=1}^{n} \delta_i I(t = t_i) \quad (3)$$

where FRT_t is the total financial risk tolerance observed at the time of t; $FRT_{smooth,t}$ is the financial risk tolerance that does not contain abnormal fluctuation; $FRT_{abnormal,t}$ is the financial risk tolerance that contains abnormal fluctuation; ϵ_t is the white noise that follows the normal distribution; μ_t is the average of financial risk tolerance by time; v_t is random fluctuation as an error to follow a normal distribution; δ_i is the size of abnormality in the observed financial risk tolerance; i represents the sequential number of the days on which FRT is observed, starting from 1; n is the number corresponding to the last day on which FRT is observed; and $I(t = t_i)$ is the binary indicator to show whether the abnormality happens or not.

This study aims to maximize the advantages of Equation (2) by evaluating how refined the smoothed FRT signal is and whether it can become a valid indicator to show significant relationships with market indicators. As mentioned earlier, most previous studies have used FRT in its raw form. However, this study proposes to increase FRT's utility by smoothing it, rather than using it in its raw state. Smoothing is an appropriate method for refining volatile time series data such as FRT, collected daily from hundreds of people, to identify useful patterns [9–11].

However, various smoothing methods exist, each performing smoothly in different ways depending on the nature of the data. For instance, time series-based methods like autoregressive (AR) and autoregressive integrated moving average (ARIMA) smooth data by connecting them seamlessly over time [12]. Exponential smoothing (ES), on the other hand, applies higher weights to more recent records, reflecting the latest trends [13].

In contrast, moving average (MA) and discrete wavelet transform (DWT) utilize diverse approaches tailored to specific data characteristics. MA calculates the average of data over a specified period to remove sharp fluctuations or outliers and highlight long-term patterns [14]. It is a simple yet widely effective smoothing technique. Discrete wavelet transform (DWT) transforms data in the frequency domain to remove noise or extract key patterns [15]. This method is particularly advantageous for handling nonlinear or non-stationary characteristics in the data. The Kalman filter, meanwhile, is used to estimate the state of dynamic systems in time series data. It is a robust tool capable of predicting and updating estimates even in noisy data environments [16].

Those smoothing techniques make it especially suitable for real-time data processing and forecasting. These smoothing methods are selected based on the characteristics of the data and the analytical objectives, with each method applied to either emphasize or mitigate the variability and patterns within the data. It can be particularly suitable for applications to data like FRT, which measures the diverse psychological fluctuations of many individuals, to extract an indicator representing the group.

1.1. Research Purpose

As explained, raw FRT data have a challenge with their sensitivity to noise. They often include behavioral noise, random fluctuations, and outliers, making it difficult to identify

genuine trends and patterns [17]. As a result, decision-making based on unprocessed data can lead to misinterpretation and potentially flawed conclusions.

Smoothing techniques are necessary to address these issues individually or in combination. Individual smoothing methods offer specific advantages. For example, Kalman filters excel in tracking dynamic systems [18], and Savitzky-Golay filters effectively preserve peaks in noisy datasets [19]. By reducing noise, these methods clarify data trends, enabling better interpretation. However, individual smoothing techniques can introduce bias. Each method carries inherent assumptions and biases that may overemphasize certain features while suppressing others [20]. Additionally, method-specific patterns or artifacts may be introduced, potentially distorting the results.

Given the limitations of individual methods, combining multiple smoothing techniques and averaging their results were suggested in this study. The combination of numerous smoothing techniques can secure robustness [21]. Combining smoothing outputs reduces the sensitivity to the limitations or biases of any single method. Averaging mitigates over-reliance on individual techniques, minimizing anomalies introduced by specific methods. This approach is also scientifically justified. First, it is expected to reduce errors by lowering the variance in the smoothed results and providing more precise estimates. Second, it prevents overfitting, while a single method might overfit specific noise or patterns in the data. Lastly, combining methods inherently allows for cross-validation, validating the consistency of the observed patterns.

Combining multiple smoothing techniques into an averaged approach strikes a balance between preserving the integrity of raw data and eliminating noise or artifacts. This methodology enhances interpretability, robustness, and reliability, making it the optimal choice for deriving precise, actionable insights from FRT data. Therefore, this study will apply an average smoothing approach to FRT data to evaluate their practical effectiveness and real-world utility. By employing this method, this study seeks to ensure that FRT becomes more realistic and practically valuable.

1.2. Smoothing Financial Risk Tolerance and Research Question

The daily FRT responses from hundreds of individuals can be divided into an average and a standard deviation (SD), as shown in Equation (4). Given the characteristics of FRT, the average changes in a small fraction, which suggests that it may have limited relevance to market indicators. However, a lot of previous research utilized the average of FRT to check how FRT is associated with the finance market [2–4,6,7]. On the other hand, the SD is expected to fluctuate, as it reflects the variation in the responses from hundreds of individuals daily. A wide SD range would indicate that investors' willingness to take risk is highly uncertain, while a narrow SD range would indicate that investors' willingness is relatively consistent. Therefore, smoothing the SD would yield better explanations of the market. Reflecting these considerations, if the two components of FRT are expressed, it would be as shown in Equation (4).

$$FRT_{smooth,t} = \mu_{smooth,t} + \sigma_{smooth,t} \cdot Z_t \quad (4)$$

where $\mu_{smooth,t}$ is the smoothed mean FRT at time t; $\sigma_{smooth,t}$ is a smoothed standard deviation of FRT at time t; and Z_t is a standard normal random variable to capture deviation from the mean. $\mu_{smooth,t}$ represents the central tendency of FRT among respondents on day t. $\mu_{smooth,t}$ is expected to change slowly over time. In terms of $\sigma_{smooth,t}$, it is a fluctuation in respondents' FRT. In Equation (4), the smoothed mean and smoothed standard deviation are different. The smoothed mean describes the central tendency of FRT, while the standard deviation describes the variability of FRT. These two components should not be added directly but treated as independent factors.

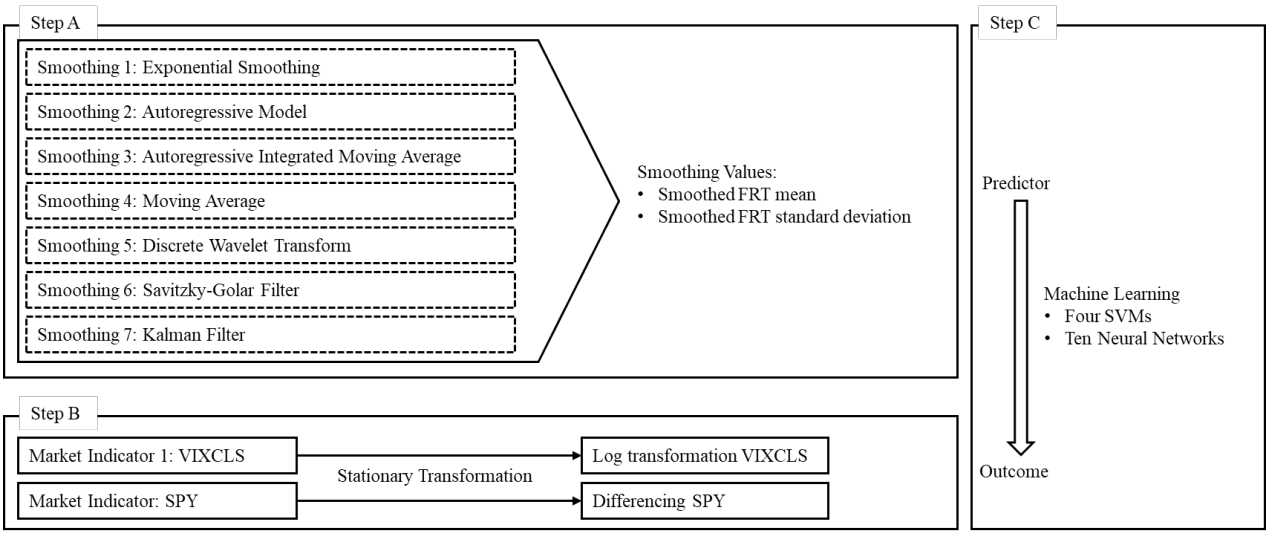
The smoothed standard deviation captures risk fluctuation. The smoothed FRT SD may capture the changing risk uncertainty in FRT responses, while the smoothed FRT average tracks the static trend. Therefore, by multiplying the standard by a random variable (Z_t), the model (Equation (4)) accounts for daily deviations around the smoothed mean, scaled by the smoothed standard deviation. Both the means of FRT and standard deviation are to be smoothed. The following vector matrix (Equation (5)) will be utilized in all of the following mathematical processes

$$FRT_{smooth,t} = \begin{bmatrix} \mu_{smooth,t} \\ \sigma_{smooth,t} \end{bmatrix} \quad (5)$$

As a result, the core research question of this study is whether FRT data, containing inherent volatility, can be effectively smoothed. If so, how much better do smoothed data perform compared to raw FRT data. Furthermore, if the smoothed FRT data provide a better fit, this study aims to investigate whether they explain market dynamics more effectively than the raw data. These questions arise from the need to refine volatile time series data like FRT, which fluctuate daily due to external factors, and assess whether smoothing methods enhance the interpretation of FRT regarding financial market indicators. This study evaluates the validity of smoothed FRT to detect long-term patterns while minimizing the effects of short-term volatility.

1.3. Research Structure

The research process follows a structured sequence comprising three main steps: Step A, Step B, and Step C, as outlined in Figure 1.



FRT: Financial Risk Tolerance
VIXCLS: Chicago Board Options Exchange Volatility Index
SPY: Standard & Poor's Depository Receipts S&P 500 Exchange Traded Fund Trust
SVM: Support Vector Machine

Figure 1. Study structure.

In Step A, various smoothing techniques are applied to refine FRT data. These methods aim to remove noise and highlight meaningful patterns in the data. Seven distinct smoothing techniques are used, each with unique strengths: ES emphasizes recent data to capture the latest trends, while AR and ARIMA models focus on time series relationships. MA smooths short-term fluctuations, and DWT isolates patterns in frequency domains. The Savitzky-Golay filter preserves the shape of the data while reducing noise, and the Kalman filter dynamically estimates states for noisy, real-time data. These methods produce two

primary smoothed values: the mean and standard deviation of FRT, providing a robust basis for further analysis.

Step B focuses on preparing market indicators for modeling. Two key indicators, VIXCLS (Chicago Board Options Exchange Volatility Index) and SPY (S&P 500 Exchange Traded Fund Trust), undergo transformations to meet the statistical requirements. The VIXCLS is log-transformed to stabilize variance, while SPY is differenced to achieve stationarity (see the Section 3 below). These transformations ensure that the market indicators are suitable for predictive modeling and align with the statistical assumptions of the subsequent steps.

In Step C, the smoothed FRT values from Step A and the transformed market indicators from Step B serve as inputs for machine learning models. Specifically, four support vector machine (SVM) models and ten neural networks (NNs) are employed to capture complex relationships and enhance predictive performance. This predictive modeling step integrates the processed data to generate accurate and reliable outcomes, demonstrating the effectiveness of the prior smoothing and transformation processes.

This sequential approach, combining data refinement, transformation, and machine learning, ensures robust and interpretable results, offering insights into the dynamics of FRT and market indicators.

1.4. Market Indicators as Outcomes

As shown in Figure 1 with Step B, this study utilized the Volatility Index (VIXCLS) released by the Chicago Board Options Exchange to reflect the financial market (CBOE). CBOE measures the expected volatility of the S&P 500 index using SPX options [22]. It incorporates both standard SPX options (i.e., expiring on the third Friday of each month) and weekly SPX options (i.e., expiring on all other Fridays). Intraday VIXCLS Index values are updated every 15 s during trading hours, reflecting real-time volatility expectations, and are calculated between 3:15 a.m. ET and 4:15 p.m. ET during different trading sessions. These VIXCLS data can be downloaded from CBOE's official webpage.

The S&P 500, which forms the basis of SPY, has been used as a representative index of the U.S. stock market since 1923 [23]. As it adopts a market capitalization-weighted method, it effectively reflects the impact of major companies' performance on the market. SPY, an ETF based on this S&P 500 index, was created in 1993 and has since become one of the most widely used benchmarks for analyzing the U.S. stock market [24]. Its high liquidity and narrow bid–ask spread make it a quick and efficient proxy for the market. Moreover, because the S&P 500 is not centered on a specific industry but considers a wide range of sectors such as technology, finance, healthcare, energy, and consumer goods, SPY also represents a broad cross-section of industries. Finally, SPY is used as a market indicator instead of the S&P 500 index itself because the S&P 500 is not a tradable index but rather a benchmark that reflects market performance. In contrast, SPY, as an ETF, is a tradable product, allowing market reactions to be more directly and visibly reflected compared to the S&P 500.

1.5. Checking the Usefulness of Smoothing by Machine Learning

As shown in Figure 1 with Step 3, this study utilizes machine learning (ML) methods to analyze the relationship between smoothed FRT and two market indicators (VIXCLS and SPY). This choice is justified by repeated processes, such as using time series characteristics and nonlinear relationships while smoothing processes are made. Analyzing the relationship between smoothed FRT and market indicators with simple time series models would not only duplicate the smoothing process but also overlap with the procedures used to ensure the stationarity of market indicators. This redundancy could lead to circular errors, where the same techniques repeatedly measure the same outcomes. ML methods

were applied to study the relationship between smoothed FRT and market indicators to overcome this issue.

Additional reasons like those listed below result in making the choice to use ML methods. First, while time series methods such as AR, ARIMA, Kalman filter, and DWT are used to smooth FRT data, the inherent noise is removed, and long-term trends or key patterns are reinforced. However, this smoothing process may weaken particular time series characteristics, such as dependencies or periodicity, within the data. Consequently, it becomes more suitable to employ ML-based models to learn complex relationships between data points rather than relying solely on time series-based forecasting. Second, traditional time series models typically assume linear relationships, which may not be sufficient to capture the more complex, nonlinear relationships between FRT and market indicators. In contrast, ML algorithms (such as SVM and NN) stand out in learning nonlinear patterns and higher-order interactions within data. Given that the relationship between FRT and market indicators can be nonlinear or multidimensional, ML models are well suited for providing higher predictive accuracy in this context. Third, after applying smoothing techniques, FRT data become relatively independent and stable, making it easier for ML algorithms to learn the structural relationships between smoothed FRT and market indicators. Even when time dependency is weakened due to the smoothing process, ML models excel in understanding and predicting outcomes based on the "snapshot" of the data. In addition, ML models improve prediction accuracy and provide additional insights through feature importance analysis. This promotes our understanding of the role of FRT in predicting market indicators. Beyond merely providing predictions, these models contribute to explaining the relationship between FRT and market indicators meaningfully.

Furthermore, the ML approaches utilized in this study were more likely simple algorithms, including four types of support vector machines (SVMs) and ten different neural network (NN) settings. The decision to use simplified SVM and NN models rather than more complex ML techniques was made to avoid issues such as overfitting. Many of the latest, more sophisticated methods are designed to maximize predictive power. Maximized predictiveness can lead to overfitting, especially with the simple data structure in this study. Given that this research involves straightforward predictors and outcome variables, there is a risk that overly complex methods might produce artificially inflated predictions that do not reflect the true utility of the smoothed FRT.

By employing simpler SVM and NN models, this study aims to ensure a clear and accurate assessment of the smoothed FRT's predictive capability without being skewed by the complexities of advanced algorithms. This approach aligns with this study's goal of evaluating the usefulness of smoothed FRT and provides a robust framework for interpreting its relationship with market indicators in a controlled and unbiased manner.

1.6. Following Structure

This paper is structured as follows. Section 2 provides a mathematical understanding of the computations. It explains how the computations shown in Figure 1—Steps A, B, and C—are performed by incorporating FRT and market indicators, detailing the specific calculations involved. Section 3 describes the materials used in this study, including FRT, VIXCLS, and SPY data, and presents the results of the smoothing and stationary transformations applied in Steps A and B of Figure 1. Section 4 focuses on the ML results corresponding to Step C in Figure 1, which constitutes the main findings of this study. If the smoothed FRT demonstrates the better predictive performance for the two market indicators compared to the raw FRT, it can be considered evidence of the utility of smoothed computation. Finally, the subsequent section discusses these results and their implications.

2. Computation

In this study, two types of variables were processed computationally. First, FRT was smoothed as a predictor variable. As mentioned earlier, this study aims to achieve a more refined smoothing of FRT. Specifically, three types of time series and four types of additional smoothing were employed. To this end, seven smoothing methods were applied to create the final smoothed FRT.

Market indicators were utilized to assess the utility of this smoothed FRT. VIXCLS and SPY were employed as market indicators. However, since unexpected market fluctuations or systematic risks can influence these indicators, it is necessary to convert them into stationary variables. To achieve this, the volatility of the market indicators was examined, and computational methods were applied to remove volatility for indicators where it was observed.

Time series methods such as AR and ARIMA were employed to smooth FRT. Additionally, time series techniques like differencing were used to assess market indicators' volatility. Therefore, a different approach should be applied to objectively estimate the predictive relationship between FRT and market indicators. Instead of relying on time series forecasting or regression methods, machine learning (ML) techniques will be utilized to evaluate how much more effective the smoothed FRT is compared to the original FRT in predicting market indicators.

This chapter provides a mathematical explanation of the smoothing formation process, the examination and removal of volatility, and the ML techniques.

2.1. Financial Risk Tolerance Smoothing

Combining multiple smoothing methods to create a robust method is an effective strategy in time series analysis, especially when dealing with noisy behavioral data like FRT [25]. Each smoothing method has strengths, so integrating them can lead to the better detection of the FRT patterns. Two sets of smoothing techniques were employed in this study. As the first set of smoothing techniques, time series-appropriate smoothing techniques such as exponential smoothing (ES), autoregressive (AR) model, and autoregressive integrated moving average (ARIMA) were applied. These methods are well-established forecasting and smoothing approaches that outdo time series data analysis [26,27], and they were used as the foundational smoothing techniques. As the second set of smoothing techniques, moving average (MA), wavelet transform (DWT), Savitzky-Golay filter, and Kalman filter were incorporated to complement these time series smoothing methods.

Specifically, MA is a simple approach that smooths by averaging recent data points. It effectively captures short-term trends, but it can lag behind significant shifts [28,29]. Second, DWT captures both frequency and time domain information, making it well suited for detecting patterns at different time scales [30]. It is more adaptive than a simple moving average. Third, the Savitzky-Golay filter smooths data while preserving higher-order features like peaks and valleys [31,32], making it ideal for a data-preserving shape. Fourth, the Kalman filter is optimal for systems with time-varying parameters, offering real-time updates and flexibility [33]. Combining these four smoothing techniques allows one to leverage the strengths of each approach while minimizing their weaknesses. For instance, the Kalman filter utilizes dynamics in real time. On the other hand, other methods, such as the moving average, work well in reducing noise over static periods.

Combining time series smoothing methods (ES, AR, ARIMA) with additional techniques (MA, DWT, Savitzky-Golay filter, Kalman filter) takes advantage of their complementary strengths. Time series methods are excellent for modeling trends and temporal dependencies [34,35], while the additional techniques are effective at addressing noise, localized patterns, and irregularities [36]. This integration enhances robustness, preserves

key data characteristics, and mitigates the limitations of individual methods, resulting in improved forecasting accuracy.

Moreover, combining these methods increases adaptability and noise reduction, as each technique performs optimally under different conditions. For instance, the Savitzky-Golay filter is better suited for high-volatility periods, while the moving average excels during stable periods. This hybrid approach ensures that the final smoothed data are less sensitive to specific conditions. Additionally, it strikes a balance between short-term responsiveness (capturing quick changes) and long-term stability (identifying overall trends), which is particularly valuable in financial contexts where trends and anomalies often coexist.

Time series smoothing methods include ES, AR, and ARIMA. To begin with ES (exponential smoothing), it is often used for forecasting time series data. Specifically, as shown in Equation (6), ES applies decreasing weights to older data while assigning weights closer to the mean for the most recent data. This approach makes it suitable for predicting the near-term future of time series data [37].

$$FRT_{smooth,t} = \alpha FRT_t + (1-\alpha)FRT_{t-1} \tag{6}$$

where α is the smoothing factor ($0 \leq \alpha \leq 1$), which determines the weight given to the most recent FRT.

Another time series smoothing method used is AR. It predicts values by linearly connecting past observations to the current value by creating a form of temporal dependency [37]. This approach provides a smoothing effect by mitigating outliers. This can be mathematically represented as shown in Equation (7).

$$FRT_{smooth,t} = c + \phi_1 FRT_{t-1} + \phi_2 FRT_{t-2} + \cdots + \phi_p FRT_{t-p} + \epsilon_t \tag{7}$$

where c is a constant; $\phi_1, \phi_2, \ldots, \phi_p$ are the coefficients of lagged FRT; p is the number of lagged observations; and ϵ_t is the error term.

The final time series smoothing method employed is ARIMA, which builds upon the AR model by incorporating two additional components: integration (to handle non-stationarity) and moving average (MA). This results in a model where the AR equation is extended with an MA component [38], as represented in Equation (8).

$$FRT_{smooth,t} = c + \phi_1 FRT_{t-1} + \phi_2 \mu_{t-2} + \cdots + \phi_p \mu_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{8}$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are the coefficients of MA terms; q is the size of the MA window. MA will be explained in the next section.

As previously mentioned, in addition to the three time series smoothing methods presented in Section 2.1, four additional smoothing techniques will be incorporated: MA, DWT, Savitzky-Golay filter, and Kalman filter. The features and equations of these methods are outlined as follows.

First, MA eliminates short-term fluctuations when the period does not show the abnormal fluctuation of FRT. For this to happen, the moving average rapidly changes [28,39]. MA for short-term smooth fluctuations is shown as the equation below (9):

$$FRT_{smooth,t} = MA_t = \frac{1}{w}\sum_{i=0}^{w-1} FRT_{t-i} \tag{9}$$

where w is the window size, such as the number of previous periods considered; FRT_{t-i} is the financial risk tolerance score at the time of $t - i$. Considering that there are five working days per week for the stock market, the 5-period window was set as Equation (10):

$$MA_t = \frac{1}{5}(FRT_t + FRT_{t-1} + FRT_{t-2} + FRT_{t-3} + FRT_{t-4}) \quad (10)$$

Application of the equation would smooth the data, and short-term outliers would be removed. This is particularly important for identifying long-term trends in FRT changes rather than focusing on short-term fluctuations.

Second, DWT captures the high frequencies as noise and keeps the low frequencies as trends [30], as shown in the equation below (11):

$$FRT_t = A_{j.t} + D_{j,t} \quad (11)$$

where $A_{j.t}$ is the approximation coefficient for a low-frequency trend at the level of j; $D_{j,t}$ is the detail coefficient for high-frequency noise at the level of j. After DWT, $D_{j,t}$ is considered noise to be dismissed, resulting in focusing on $A_{j.t}$. Therefore, Equation (8) changes to the trend equation focusing on $A_{j.t}$ as the equation below (12):

$$A_{j,t} = \sum_n FRT(n)\phi_{j,n,t} \quad (12)$$

where $\phi_{j,n,t}$ is the scaling function at the level of j with the number of observations n. Then, by considering the FRT vectors from Equation (5), Equation (12) can be further developed to consider each component as Equations (13) and (14):

$$A_{j,t} = \mu_t + \sum(FRT(n) - \mu_t)\phi_{j,n,t} \quad (13)$$

$$D_{j,t} = \sigma_t(\sum_n (\frac{FRT(n) - \mu_t}{\sigma_t})\psi_{j,n,t}) \quad (14)$$

where the standard deviation (σ_t) controls the volatility and $\psi_{j,n,t}$ is the basis function of the wavelet to obtain the high-frequency component. Applying DWT separates short-term volatility and long-term trends, allowing for the identification of more detailed fluctuations.

Third, in terms of the SG filter, after smoothing and eliminating the high frequencies, the SG filter preserves the observed FRT's dynamic characteristics by fitting a polynomial to the data points with a moving window. The SG filter maintains the dynamics [31,40] as Equation (15):

$$SG_t = \sum_{k=-m}^{m} C_k FRT_{t+k} \quad (15)$$

where C_k denotes the coefficients of the polynomial; m is the half-window size from the full size ($2m + 1$). If the degree of the polynomial of FRT is three, then the equation would be represented as Equation (16):

$$FRT \coloneqq c_0 + c_1(t+k)^1 + c_2(t+k)^2 + c_3(t+k)^3 \quad (16)$$

By doing this, the SG filter calculates the smoothed FRT estimate at a specific time point.

Finally, for the additional smoothing process, the Kalman filter supports the correction to predict the FRT [33] by using Equations (17)–(19). Equation (17) predicts the next FRT trend; Equations (18) and (19) update the predicted state with the observed FRT; and Equation (20) updates the error covariates:

$$x_{t|t-1} = F \cdot x_{t-1|t-1} + u \quad (17)$$

$$K_t = \frac{P_{t|t-1}H^T}{HP_{t|t-1}H^T + R} \tag{18}$$

$$x_{t|t} = x_{t|t-1} + K_t\left(FRT_t - Hx_{t|t-1}\right) \tag{19}$$

$$P_{t|t} = (I - K_tH)P_{t|t-1} \tag{20}$$

where $x_{t|t-1}$ is the predicted FRT trend; F is the state transition matrix; u is the control input, which is usually zero; K_t is the Kalman gain to be updated; $P_{t|t-1}$ is the predicted error covariance; H is the observation model; and R is the measurement noise covariance. The two components of the vector mentioned in Equation (5), the mean and standard deviation, are each smoothed using four different smoothing methods. As a result, the following matrix structures (Equations (21)–(23)) are formed:

$$FRT_{\mu,t} = \begin{bmatrix} \mu_t^{Exp} \\ \mu_t^{AR} \\ \mu_t^{ARIMA} \\ \mu_t^{MA} \\ \mu_t^{DWT} \\ \mu_t^{SG} \\ \mu_t^{K} \end{bmatrix} \tag{21}$$

$$FRT_{\sigma,t} = \begin{bmatrix} \sigma_t^{Exp} \\ \sigma_t^{AR} \\ \sigma_t^{ARIMA} \\ \sigma_t^{MA} \\ \sigma_t^{DWT} \\ \sigma_t^{SG} \\ \sigma_t^{K} \end{bmatrix} \tag{22}$$

$$FRT_{smooth,t}\ Matrix = \begin{bmatrix} f(\mu_t^{Exp}\ \mu_t^{AR}\ \mu_t^{ARIMA}\ \mu_t^{MA}\ \mu_t^{DWT}\ \mu_t^{SG}\ \mu_t^{K}) \\ f(\sigma_t^{Exp}\ \sigma_t^{AR}\ \sigma_t^{ARIMA}\ \sigma_t^{MA}\ \sigma_t^{DWT}\ \sigma_t^{SG}\ \sigma_t^{K}) \end{bmatrix} \tag{23}$$

where μ_t^{Exp} is the smoothed mean from ES; μ_t^{AR} is the smoothed mean from AR; μ_t^{ARIMA} is the smoothed mean from ARIMA; μ_t^{MA} is the smoothed mean from MA; μ_t^{DWT} is the smoothed mean from DWT; μ_t^{SG} is the smoothed mean from the SG filter; μ_t^{K} is the smoothed mean from the Kalman filter; σ_t^{Exp} is the smoothed standard deviation from ES; σ_t^{AR} is the smoothed standard deviation from AR; σ_t^{ARIMA} is the smoothed standard deviation from ARIMA; σ_t^{MA} is the smoothed standard deviation from MA; σ_t^{DWT} is the smoothed standard deviation from DWT; σ_t^{SG} is the smoothed standard deviation from the SG filter; and σ_t^{K} is the smoothed standard deviation from the Kalman filter. Equation (23) combines both the smoothed mean and smoothed standard deviation. Based on this, Equation (24) shows the value matrix to be inserted into the empirical testing:

$$Smoothing\ FRT_t\ Matrix = \begin{bmatrix} \overline{\mu_t} \\ \overline{\sigma_t} \end{bmatrix}\ Subject\ to\ \begin{bmatrix} \overline{\mu_t} = \frac{1}{7}\sum_1^7 \mu_t^{(s)} \\ \overline{\sigma_t} = \frac{1}{7}\sum_1^7 \sigma_t^{(s)} \end{bmatrix} \tag{24}$$

2.2. Making Market Indicators Stationary

To evaluate how well the smoothed FRT explains the market, this study examines the selected market indicators, VIXCLS and SPY. Two market indicators must first undergo a volatility check. For this, Augmented Dickey-Fuller (ADF) and Phillips-Perron tests are conducted, and appropriate volatility adjustments are made based on the test results [41,42]. By processing the selected market indicators to achieve stationarity, the relationship between the smoothed FRT and the market indicators can be assessed more clearly, aligning with the objective of this study. The ADF and Phillips-Perron tests used for the volatility check are conducted based on the following formulas. First, ADF utilized the following Equation (25):

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-1} + \epsilon_t \quad (25)$$

where Δy_t is the first difference in the series y_t (VIXCLS and SPY) by using $y_t - y_{t-1}$; α is the constant; β is the coefficient of time, which accounts for time-related patterns in the data; γ is the coefficient of lagged y_t, which is the indicator of stationary; p is the number of lags in the dataset; ϕ_i is the coefficients of past differences; and ϵ_t is the error term. By using this equation, γ can be smaller than zero or equal to zero. When γ is revealed to be lower than zero, the market indicator is stationary. If it is the same as zero, the market indicator is non-stationary.

The Phillips-Perron test is a simplified version of ADF by using Equation (26):

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \epsilon_t \quad (26)$$

where the components are the same as the ADF equation, while the Phillips-Perron test does not have the lagged difference term ($\sum_{i=1}^{p} \phi_i \Delta y_{t-1}$). Instead, the Phillips-Perron test utilizes the t-statistic from the simplified version equation as below (27):

$$Z_T = \frac{t_\gamma}{\sqrt{\hat{S}_e}} \quad (27)$$

where the denominator is the square root of a correction term ($\hat{S}_e$); $\hat{S}_e$ accounts for the autocorrelation and heteroskedasticity in the residuals; and t_γ is the t-statistics for the coefficient γ. When Z_T is larger than the critical value, the market indicator is stationary. Otherwise, it is revealed as non-stationary.

Based on the results of the ADF and Phillips-Perron tests, appropriate measures will be taken for the market indicators. These measures will result in the indicators being transformed into either a natural logarithm form or a different form depending on the outcomes. In this study, the transformations will be applied appropriately according to the results presented in the Materials Section below.

2.3. Machine Learning Prediction

As mentioned in the Introduction, this study employs four types of SVMs and ten configurations of NNs. Each SVM and NN is represented by the following. First, the SVM can be explained [43,44] with these Equations (28) and (29) by considering the smoothed FRT (see Equation (5)):

$$VIXCLS_t = f(FRT_{smooth,t}) = \sum_{i=1}^{N} \alpha_i VIXCLS_i K(FRT_{smooth,i}, FRT_{smooth,t}) + b \quad (28)$$

$$SPY_t = f(FRT_{smooth,t}) = \sum_{i=1}^{N} \alpha_i SPY_i K(FRT_{smooth,i}, FRT_{smooth,t}) + b \quad (29)$$

where $VIXCLS_t$ and SPY_t are the predicted market indicators (VIXCLS and SPY) at the time of t; $VIXCLS_i$ and SPY_i are the observed values of market indicators for support

vector $i = 1, 2, \ldots, N$; K is the kernel function, including linear, radial basis function (RBF), polynomial, and sigmoid; α_i denotes the support vector weights; and b is the bias term. Four SVMs are made by each K by using Equations (30)–(33) (for Linear kernel, polynomial kernel, RBF kernel, and sigmoid kernel, respectively):

$$K(u, v) = u^T v \tag{30}$$

$$K(u, v) = (u^T v + c)^d \tag{31}$$

$$K(u, v) = exp(-\gamma||u - v||^2) \tag{32}$$

$$K(u, v) = tanh(\gamma u^T v + c)^d \tag{33}$$

where u and v are the feature vectors representing the data points from the input (e.g., u = ith FRT_{smooth} and v = jth FRT_{smooth}); c is the bias term; d is the degree of nonlinear interaction; γ is the control factor that decreases the distance; *tanh* is the hyperbolic tangent; and T represent the mapping function in the kernel transformation.

NNs with ten settings (neurons = 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100) were utilized [45] with the below Equations (34)–(36). Equation (34) indicates the input and Equations (35) and (36) show the NNs with a rectified linear unit (ReLU):

$$Input\ Layer(X) = FRT_{smooth,t} = \begin{bmatrix} \mu_{smooth,t} \\ \sigma_{smooth,t} \end{bmatrix} \tag{34}$$

$$VIXCLS_t = ReLU(W_{VIXCLS,\,l} \sum_{l=1}^{2} FRT_{smooth,l} + b_{VIXCLS}) \tag{35}$$

$$SPY_t = ReLU(W_{SPY,\,l} \sum_{l=1}^{2} FRT_{smooth,l} + b_{SPY}) \tag{36}$$

where $VIXCLS_t$ and SPY_t are the predicted market indicators (VIXCLS and SPY) at the time of t; l includes the mean and standard deviation as inputs; and b_{VIXCLS} and b_{SPY} are bias terms for each number of neurons. ReLU activation is widely utilized in neural networks, which select the maximized input like in the below Equation (37):

$$f(x) = \max(0, x) \tag{37}$$

where x is input to the activation.

2.4. United Forms for Smoothing, Stationary Market, and Machine Learning

In Section 2.1, we demonstrated how FRT, as an input variable, is smoothed into a single variable. Equation (5) serves as the raw input, while Equation (6) through (20) outline the smoothing transformations. The integration of these transformations into a single representation is described in Equations (21)–(24). When united into a single equation, it can be represented as Equation (38) below.

$$FRT_{smooth,t} = \begin{bmatrix} \mu_{smooth,t} \\ \sigma_{smooth,t} \end{bmatrix} = \left[\frac{\sum_{i=1}^{m} w_i FRT_{smooth,i,t}}{\sum_{i=1}^{m} w_i}, \sqrt{\frac{\sum_{i=1}^{m} w_i (FRT_{smooth,i,t} - FRT_{smooth,t})^2}{\sum_{i=1}^{m} w_i}} \right] \tag{38}$$

where $FRT_{smooth,i,t}$ is the smoothed FRT from method i (ES, AR, ARIMA, MA, DWT, Savitzky-Golay filter, and Kalman filter); w_i represents each method(i)'s parameters.

In Section 2.2, the usefulness of the smoothed FRT will be examined as an input variable using market indicators (VIXCLS and SPY) as the outcome variable. However, since market indicators can often be non-stationary due to various factors such as systematic

risk, Equation (25) through (27) were used to transform them into stationary variables. When consolidated into a single equation, Equation (39) below is represented.

$$Y_t = \begin{cases} ln(Y_t), & if\ tranformation\ needed \\ \Delta Y_t, & if\ differencing\ needed \\ Y_t, & otherwise \end{cases} \tag{39}$$

where Y_t is the market indicators (VIXCLS and SPY).

Finally, Section 2.3 used ML techniques to examine how well the input variable predicts the outcome variable. Four types of SVMs and ten different settings of NNs were employed. When consolidated into a single equation, they are represented as Equation (40) below.

$$Y_t = f(FRT_{smooth,t}) = \begin{cases} K(u,v),\ SVM\ with\ kernel\ function \\ ReLu(W \cdot FRT_{smooth,t} + b),\ Neural\ Network \end{cases} \tag{40}$$

This study employs four standard SVM kernels (linear, polynomial, radial basis function, and sigmoid) to capture a broad range of potential functional relationships between the input data and market indicators. Although the linear and RBF kernels are frequently favored for their balance of performance and interpretability, polynomial and sigmoid kernels can reveal additional forms of nonlinearity. To further explore predictive performance, this research also considers ten neural network configurations with hidden-layer sizes ranging from 10 to 100 neurons. This systematic variation allows for an examination of how increasing model complexity affects predictive accuracy, mitigating the risk of underfitting when the model is too simple and overfitting when it becomes overly complex. The analysis aims to identify the most robust modeling approach for linking smoothed FRT data to market outcomes by comparing the results across multiple kernel types and network sizes.

3. Materials and Smoothing Process Results

In this study, the data comprise three types: one type is FRT data, and the other two types are market indicators, VIXCLS and SPY. VIXCLS is a market Volatility Index provided by the Chicago Board Options Exchange (CBOE) [46], while SPY is an Exchange Traded Fund (ETF) based on the S&P 500 provided by Standard & Poor's. Among the ETFs, SPY, the Standard & Poor's Depositary Receipt (SPDR), is widely regarded as the ETF that best represents the market [47].

First, the input variables (mean and standard deviation) for FRT were collected through a publicly available survey hosted on a website by a university extension station in the Midwest region [48]. This survey, approved by the IRB of the extension station, allows individuals across the United States to participate in checking their FRT scores. The survey itself is publicly accessible, allowing anyone to visit at any time to measure and check their own FRT. The survey includes 13 items developed by Grable and Lytton [1]. This scale uses the sum of the 13 items, with scores ranging from a minimum of 13 to a maximum of 47, where higher scores indicate a higher risk tolerance. The details of the data are summarized in Table 1.

As shown in Table 1, data were collected over 1394 working days from 8 January 2018 to 24 July 2023. The counted data only include working days to allow for a comparison with the market indicators (VIXCLS and SPY), which will be discussed later. During this period, the survey gathered responses from 463,337 individuals. These responses were restructured into the daily data, averaging 271.81 participants per day (SD = 141.32). Given that the number of daily respondents ranged from a minimum of 25 to a maximum of 771, and the daily average exceeded 270 respondents—well above the threshold generally required for

the law of large numbers to apply ($n > 30$)—the dataset satisfies the assumptions of the law of large numbers, ensuring that the sample average closely approximates the population mean. Moreover, given that the respondents participated from across the United States, this dataset can be considered a robust reflection of the FRT of U.S. individual investors. When examining the daily average of the respondents' FRT scores, the mean of the averages was 27.46 (SD = 0.57), indicating relatively consistent daily averages. In contrast, the standard deviation of the respondents on any given day had a mean of 4.94 (SD = 0.30), showing more variation than the daily average.

Table 1. Descriptive information of dataset (total sample = 463,337 over 1,394 days).

	Mean	SD	Min	Max
FRT	26.65	6.60	13	47
Daily samples (count)	271.81	141.32	25	771
Daily FRT mean	27.46	0.57	25.79	29.62
Daily FRT SD	4.94	0.30	3.64	6.05
VIXCLS	21.03	8.13	9.52	82.69
ln(VIXCLS)	2.99	0.33	2.25	4.42
SPY	351.44	66.70	222.95	477.71
SPY differencing	0.13	4.39	−29.47	21.21
1 month T-Bill	1.57	1.52	0	6.02

Note. FRT: financial risk tolerance; SD: standard deviation; VIXCLS: Volatility Index; SPY: SPDR ETF; T-Bill: Treasury Bill as risk-free market inflation; a further explanation about ln(VIXCLS) and SPY differencing will be provided in Section 3.2.

As explained in the background, VIXCLS and SPY are used as the market indicators in this study. As shown in Table 1, the daily average of FRT exhibits relatively low volatility. The minimum average daily FRT is 25.79, and the maximum average daily FRT is 29.62 compared to the average daily FRT (27.46); therefore, the volatility is 13.95%. On the other hand, the daily SD of FRT shows significant fluctuations, ranging from 3.64 to 6.05, with a mean of 4.94. The volatility range is 48.79%. This substantial variability in individuals' willingness to take investment risks aligns with the VIXCLS and SPY characteristics, which measure market volatility and liquidity. Hence, VIXCLS and SPY were selected as suitable market volatility indicators that reflect changes in investor risk tolerance. Descriptive information about VIXCLS and SPY are shown in Table 1.

Additionally, while using VIXCLS as an input variable, it is necessary to consider the inflation rate during the observed period from December 2017 to July 2023. Therefore, this study includes inflation as a control variable in addition to market indicators when it is VIXCLS modeling. Inflation rates can be derived from sources such as the Consumer Price Index (CPI) or Treasury Bills (T-Bills). CPI, however, is measured monthly and categorized by product groups [49], making it unsuitable for this study, which is conducted on a daily basis. On the other hand, T-Bills serve as a risk-free interest rate in financial markets and are commonly used as an inflation-reflecting indicator. Typically, either the 1-month T-Bill or the 3-month T-Bill is used for this purpose. However, since a significant portion of the 2018 data for the 3-month T-Bill are missing, this study employs the 1-month T-Bill instead [50]. Descriptive information regarding this variable is presented in Table 1.

3.1. FRT Smoothing Process

Taking all seven smoothing techniques, Figure 2 shows the smoothed mean of FRT by ES compared to the raw mean of FRT; Figure 3 shows the smoothed mean of FRT by AR compared to the raw mean of FRT; Figure 4 shows the smoothed mean of FRT by ARIMA compared to the raw mean of FRT; Figure 5 shows the smoothed mean of FRT by MA compared to the raw mean of FRT; Figure 6 shows the smoothed mean of FRT by DWT compared to the raw mean of FRT; Figure 7 shows the smoothed mean of FRT by the SG

filter compared to the raw mean of FRT; and Figure 8 shows the smoothed mean of FRT by the Kalman filter compared to the raw mean of FRT.

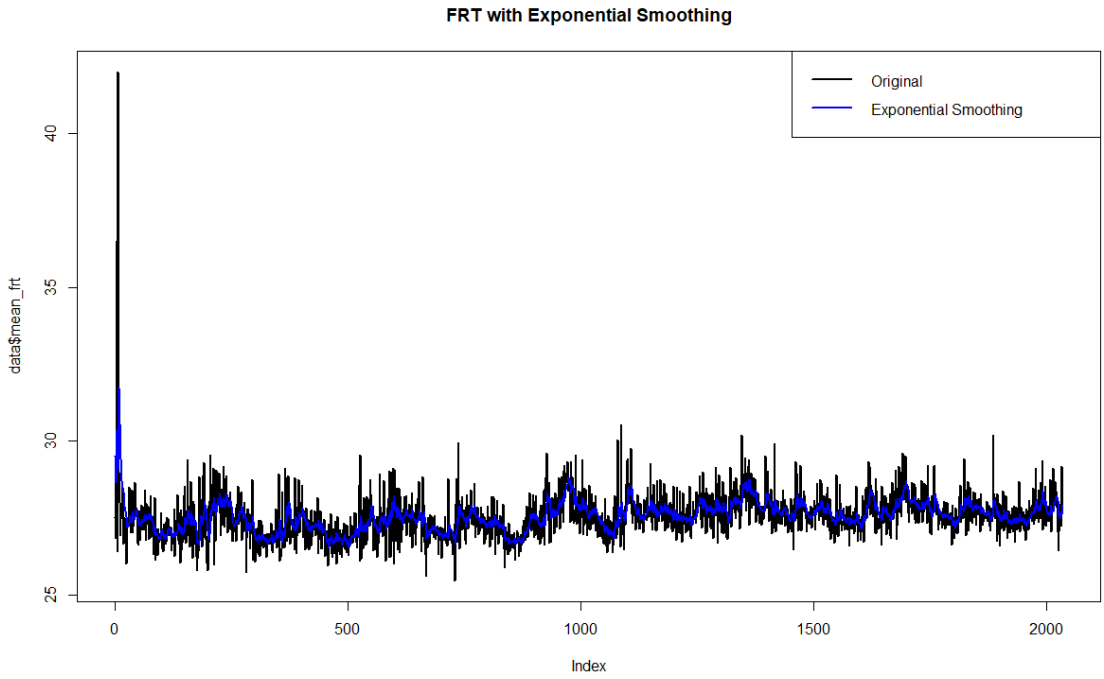


Figure 2. The smoothed mean of FRT by ES.

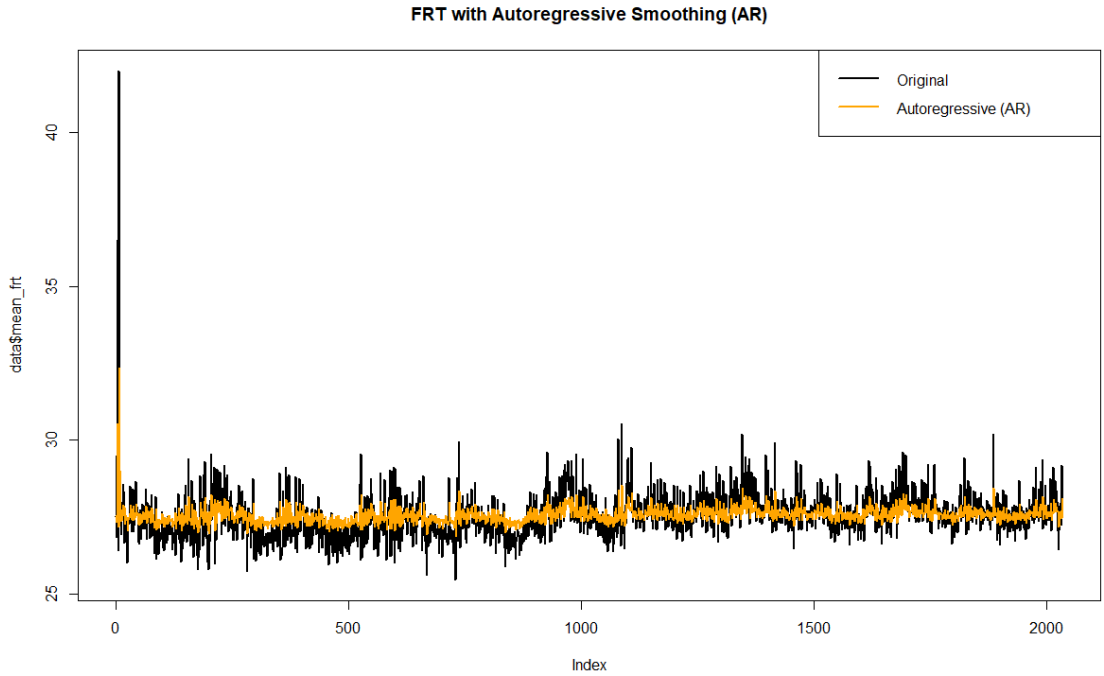


Figure 3. The smoothed mean of FRT by AR.

Figure 2 illustrates the FRT data over time, comparing the original raw data (black line) with their exponentially smoothed version (blue line). The original data exhibit significant short-term fluctuations and volatility. The exponentially smoothed data effectively reduce the noise in the original data, highlighting the underlying trends more clearly. While the smoothed line closely follows the overall trajectory of the original data, it filters out sharp spikes and day-to-day volatility. This makes it particularly useful for identifying long-term patterns and trends in the data. As a result, ES shows its utility in reducing noise and

providing a clearer perspective on the data. This approach is valuable for analysts focusing on broader patterns rather than short-term variability in FRT data.

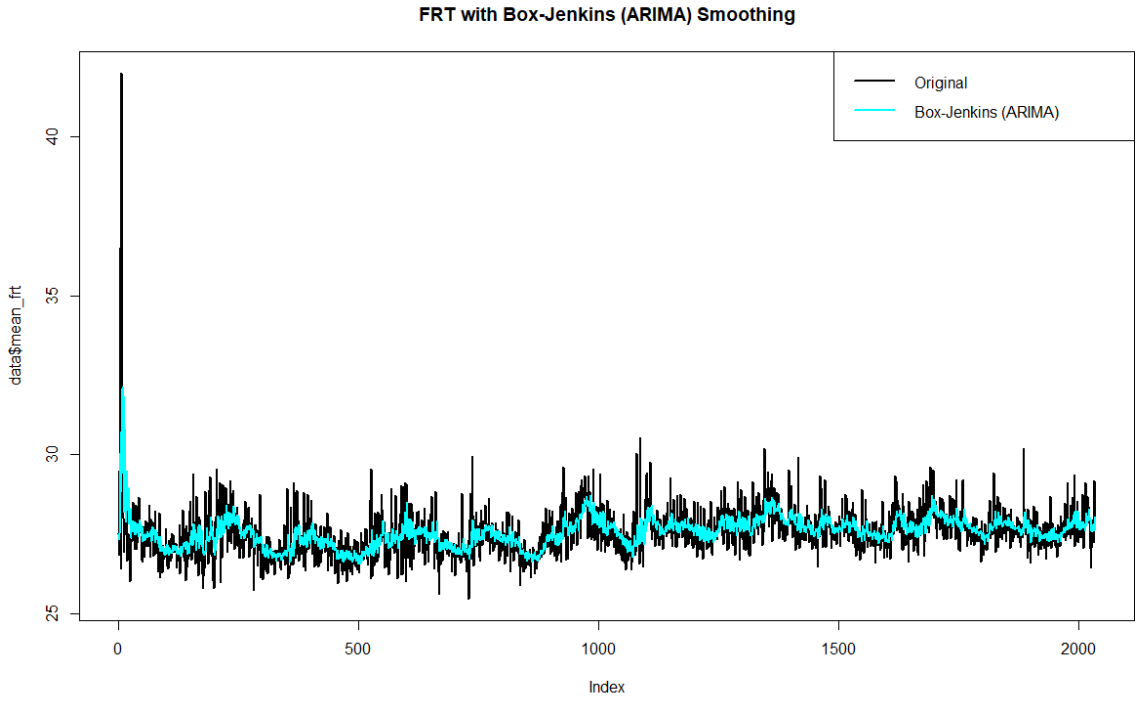


Figure 4. The smoothed mean of FRT by ARIMA.

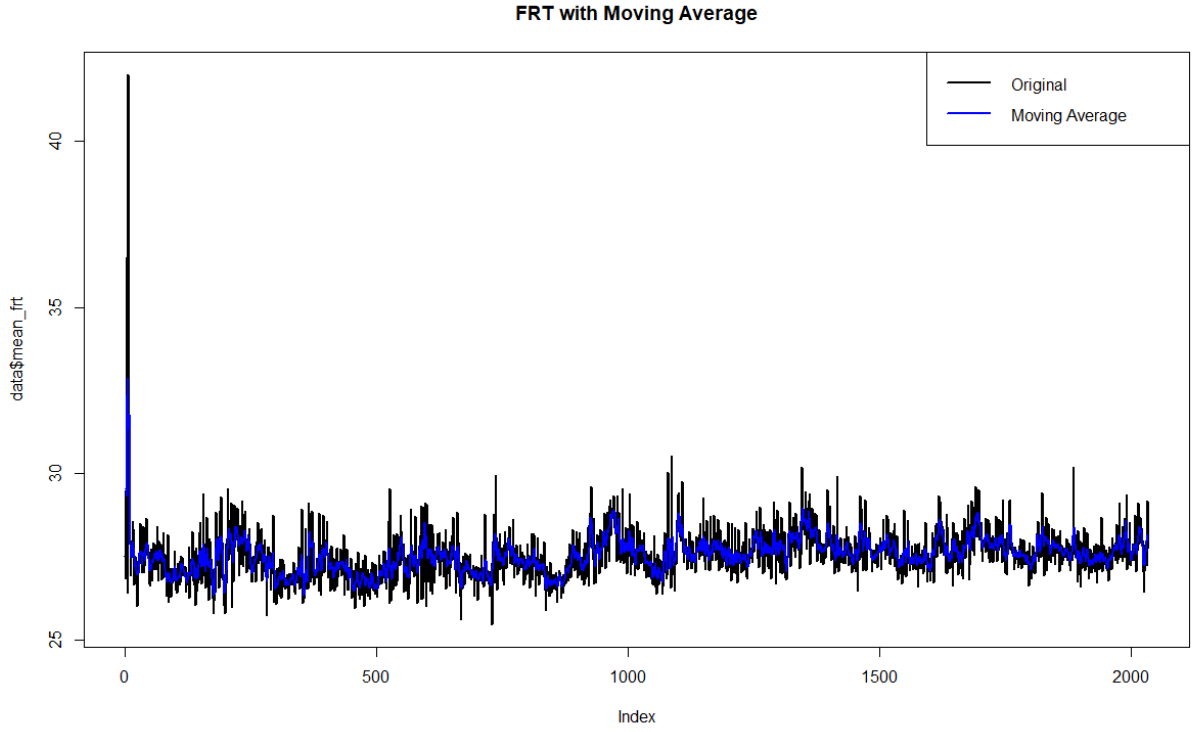


Figure 5. The smoothed mean of FRT by MA.

Figure 3 shows the FRT data over time, comparing the original raw data (black line) with using AR (orange line). The autoregressive smoothing effectively reduces the noise in the data while retaining the underlying structure. The orange line captures the general trend by leveraging past values to predict current ones, creating a smoother data representation. It maintains a closer alignment with the original data compared to

other smoothing techniques, such as exponential smoothing, which might average out more variability. The figure demonstrates how the AR model filters out short-term noise and emphasizes long-term trends in the FRT data. The smoothed line provides a more interpretable view of the underlying behavior of FRT while preserving key fluctuations that might carry meaningful information. This approach is advantageous when short-term dependencies or patterns in the data are of interest.

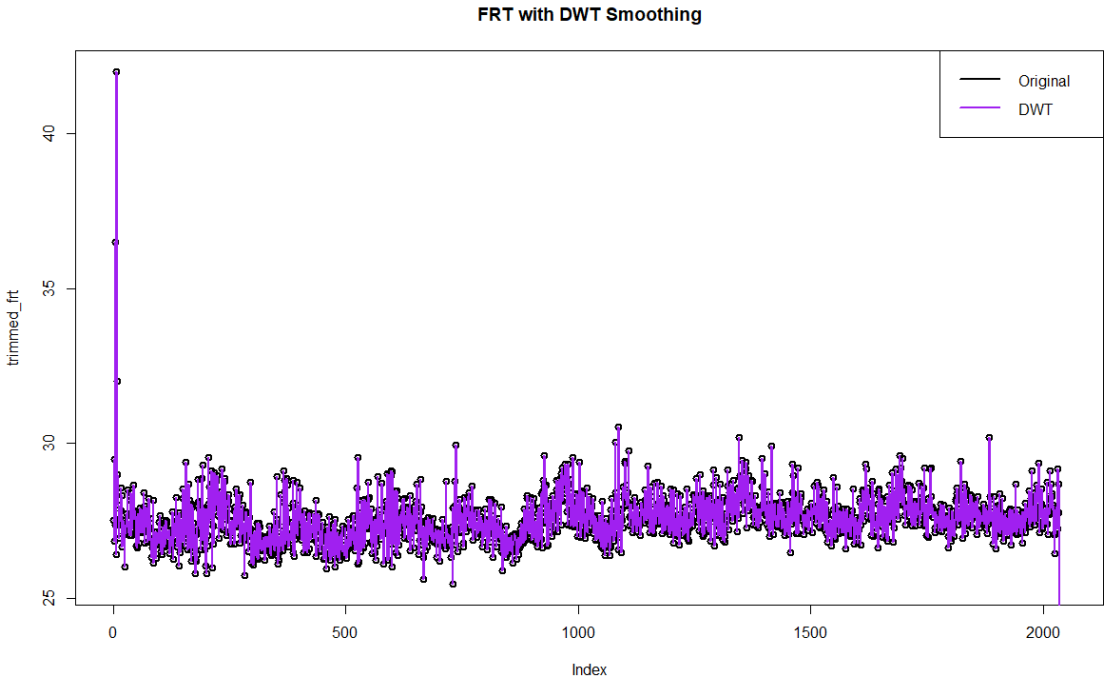


Figure 6. The smoothed mean of FRT by DWT.

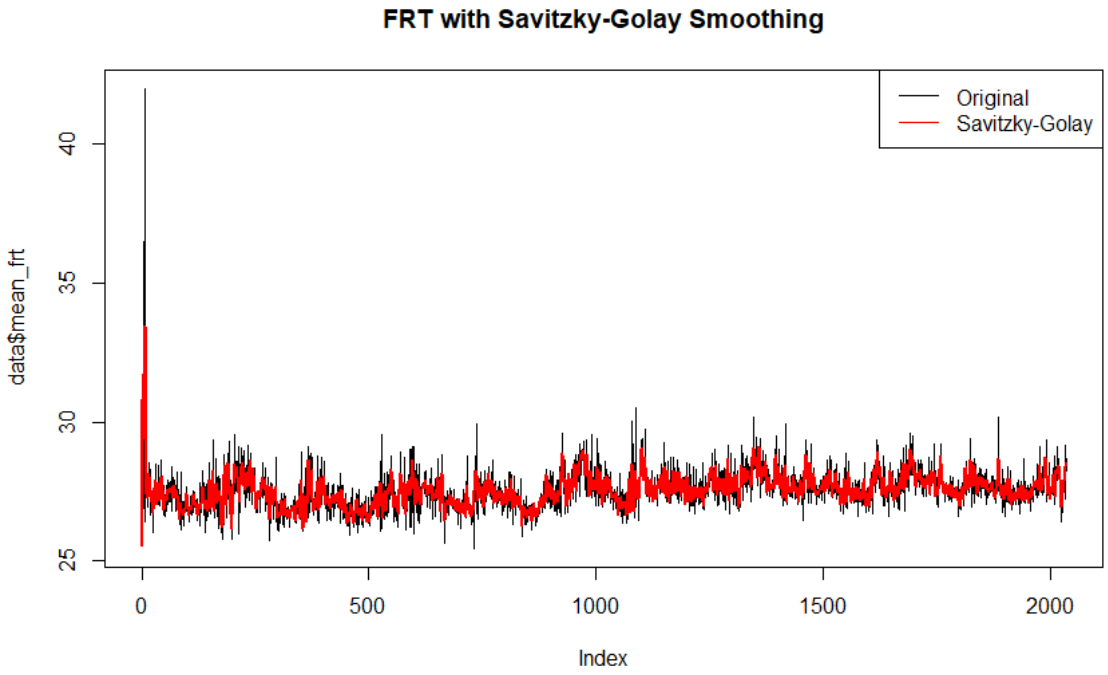


Figure 7. The smoothed mean of FRT by SG filter.

Figure 4 illustrates the FRT data over time, comparing the original raw data (black line) with the ARIMA model (cyan line). The ARIMA-smoothed data effectively reduce noise and highlight the underlying trends while maintaining the data's structure. Unlike simpler smoothing techniques, the ARIMA model is well suited for capturing both short-term

dependencies and long-term trends by modeling the temporal structure of the data. The cyan line closely follows the overall trajectory of the original data but smooths out sharp, irregular fluctuations, particularly after the initial spike. This visualization demonstrates the ARIMA model's ability to extract meaningful patterns from volatile data while preserving critical characteristics. It is beneficial for analyzing time series data where both short-term dynamics and long-term trends are interesting.

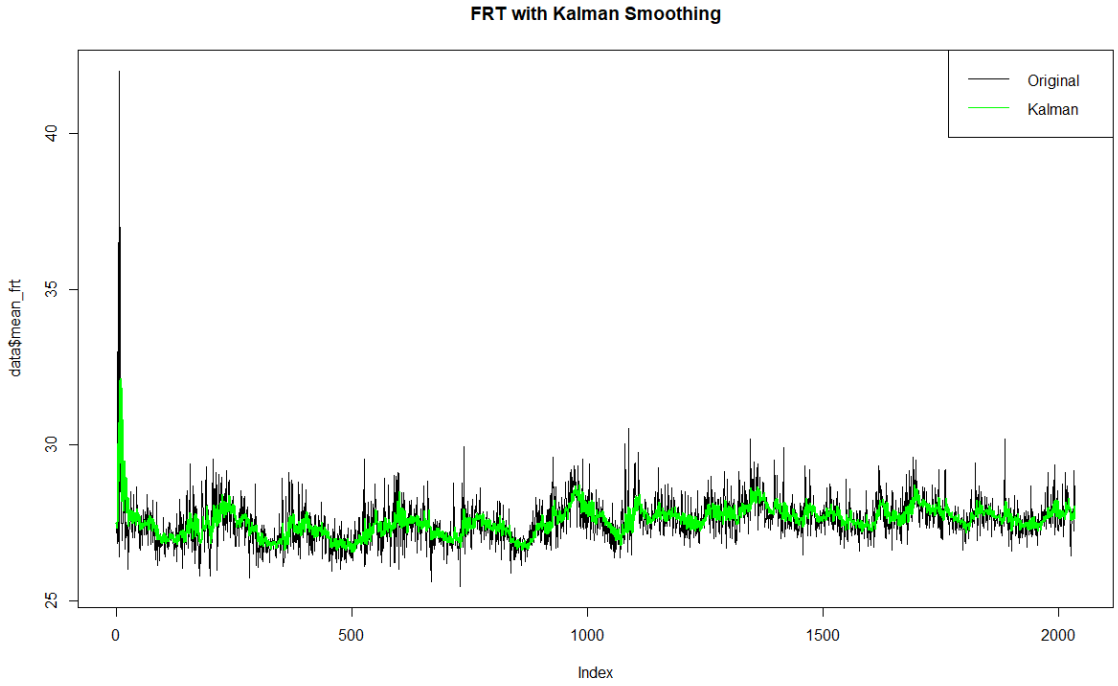


Figure 8. The smoothed mean of FRT by Kalman filter.

Figure 5 shows FRT data over time, comparing the original raw data (black line) with MA (blue line). The MA-smoothed data (blue line) reduce noise by averaging values over a fixed window. This approach smooths the sharp fluctuations in the original data, revealing the broader trend while eliminating much of the day-to-day variability. The MA closely follows the general trajectory of the original data but removes irregular spikes, particularly after the initial period. This visualization highlights the effectiveness of the MA technique in smoothing volatile time series data.

Figure 6 presents FRT data over time, comparing the original raw data (black line) with the DWT (purple line). The DWT-smoothed data (purple line) reduce noise by decomposing the original data into wavelet components and filtering out high-frequency variations while retaining the underlying trends. This method effectively captures the data's long-term patterns and key features while smoothing short-term fluctuations. DWT smoothing maintains a closer alignment with the original data than other techniques, preserving essential details while reducing random noise. This graph highlights the ability of DWT smoothing to balance the trade-off between noise reduction and the retention of meaningful features. It provides a clear view of the underlying trends in FRT while keeping essential variations intact, making it useful for identifying patterns in time series data with complex dynamics.

Figure 7 illustrates FRT data over time, comparing the original raw data (black line) with the smoothed version obtained using Savitzky-Golay smoothing (red line). The Savitzky-Golay smoothed data (red line) effectively reduce noise while preserving the overall shape and finer details of the data. This smoothing method is particularly well suited for data with a lot of local variability, as it applies polynomial fitting within a sliding window to minimize distortion of the original signal. The smoothed line closely follows

the general trajectory of the original data while filtering out random noise and maintaining sharper transitions compared to other smoothing techniques. This graph demonstrates how the Savitzky-Golay smoothing technique balances noise reduction and the retention of essential features in the data.

Figure 8 presents the FRT data over time, comparing the original raw data (black line) with the smoothed version using Kalman smoothing (green line). The Kalman smoothed data (green line) effectively filter out noise while retaining the general structure and trends of the original data. Kalman smoothing, a recursive algorithm, combines the observed data with an underlying state-space model to estimate the actual state of the system. This technique produces a smooth trajectory resembling the original data's patterns while significantly reducing random fluctuations. The Kalman smoothed line captures the broader trends and periodic fluctuations in the data while minimizing noise and erratic movements. Unlike simpler smoothing methods, Kalman smoothing adapts dynamically to the data's characteristics, making it well suited for time series with both rapid variations and stable regions.

Similarly, taking all seven smoothing techniques, Figure 9 shows the smoothed standard deviation of FRT by ES compared to the raw standard deviation of FRT; Figure 10 shows the smoothed standard deviation of FRT by AR compared to the raw standard deviation of FRT; Figure 11 shows the smoothed standard deviation of FRT by ARIMA compared to the raw standard deviation of FRT; Figure 12 shows the smoothed standard deviation of FRT by MA compared to the raw standard deviation of FRT; Figure 13 shows the smoothed standard deviation of FRT by DWT compared to the raw standard deviation of FRT; Figure 14 shows the smoothed standard deviation of FRT by the SG filter compared to the raw standard deviation of FRT; and Figure 15 shows the smoothed standard deviation of FRT by the Kalman filter compared to the raw standard deviation of FRT.

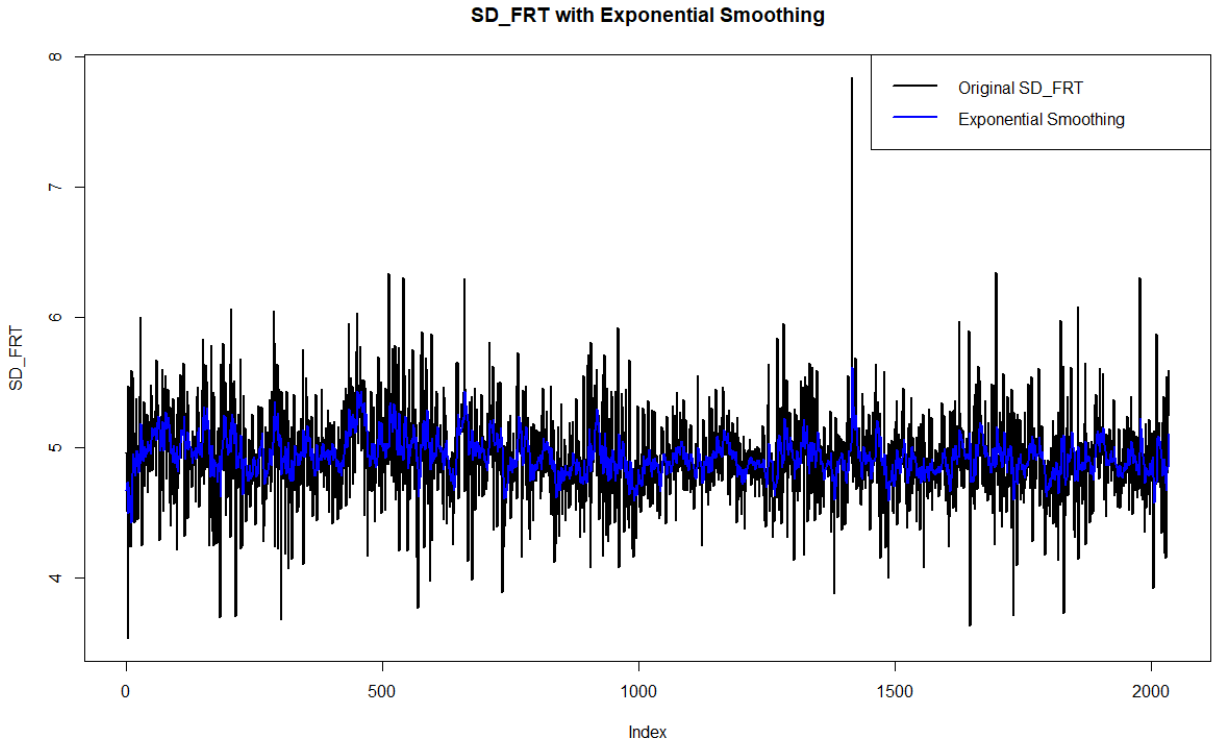


Figure 9. The smoothed standard deviation of FRT by ES.

Figure 9 displays the standard deviation of FRT over time, comparing the original data (black line) with the exponentially smoothed version (blue line). The original data exhibit

frequent fluctuations, indicating variability in daily FRT measurements. Midway through the series, a noticeable spike occurs, suggesting a temporary increase in dispersion. The ES effectively reduces short-term volatility, providing a clearer view of overall trends in the variability of FRT. Unlike the original line, the smoothed version filters out noise while retaining key shifts, making identifying long-term patterns in the data's dispersion easier.

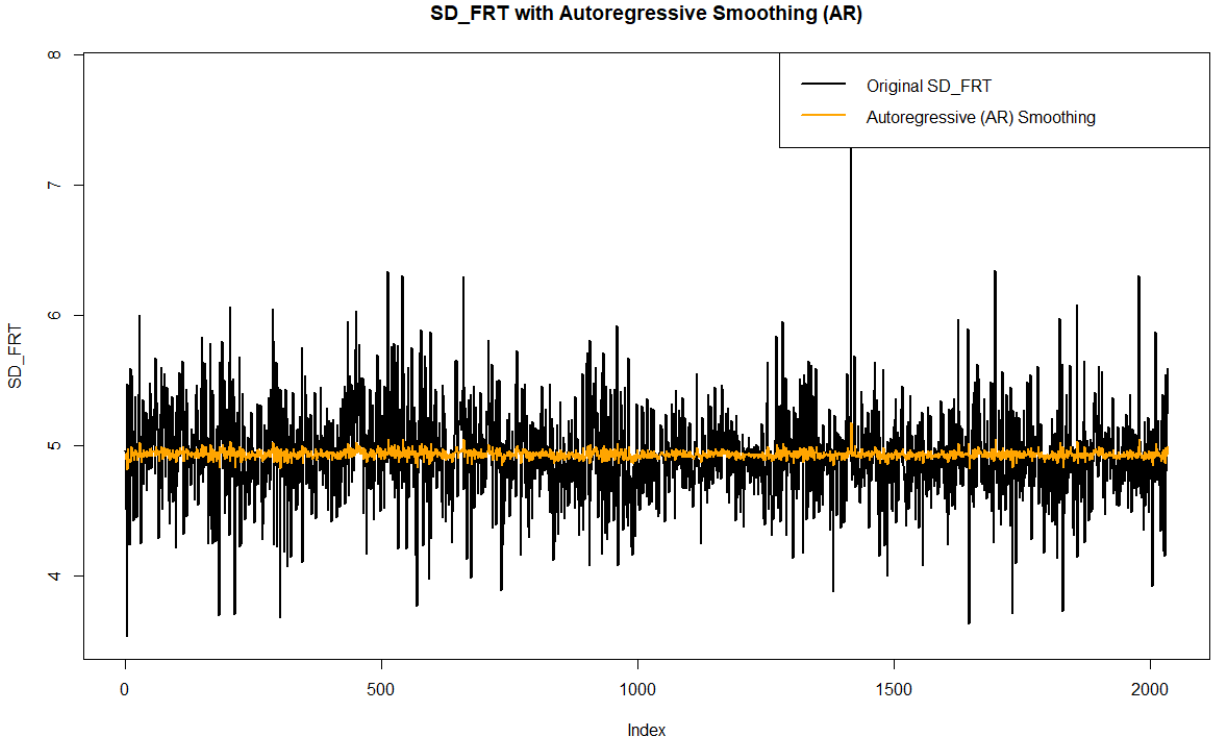


Figure 10. The smoothed standard deviation of FRT by AR.

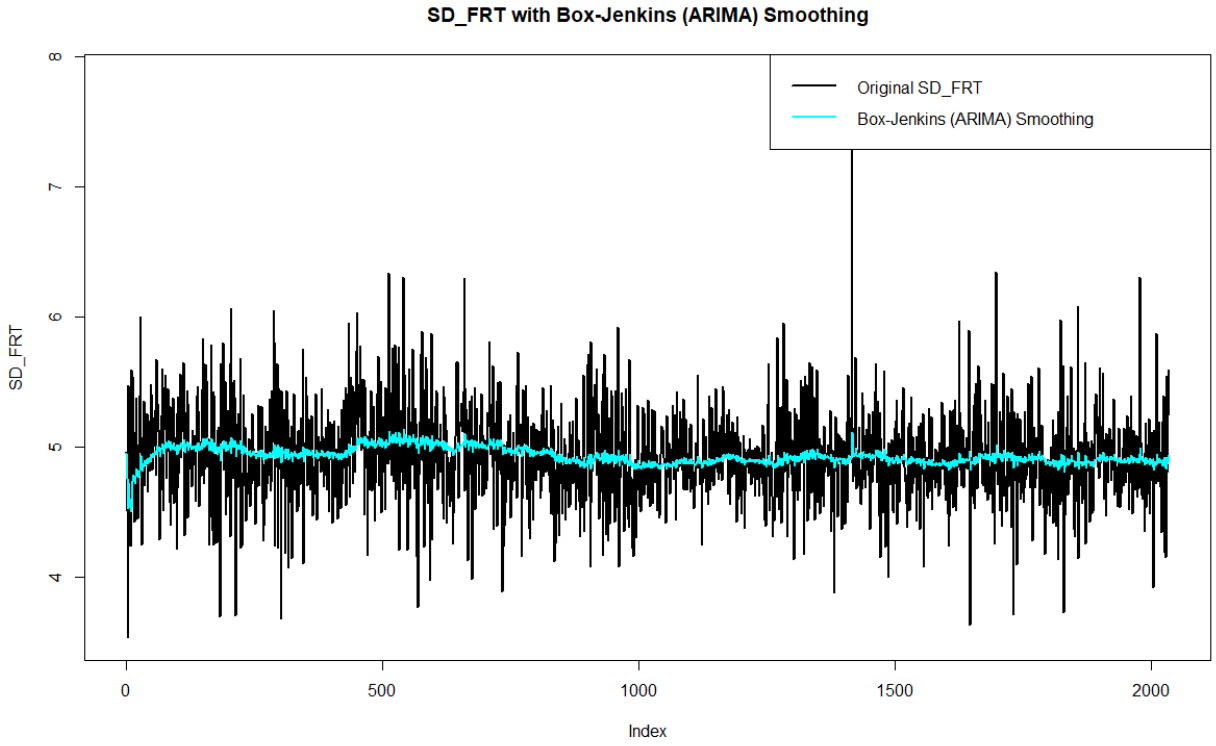


Figure 11. The smoothed standard deviation of FRT by ARIMA.

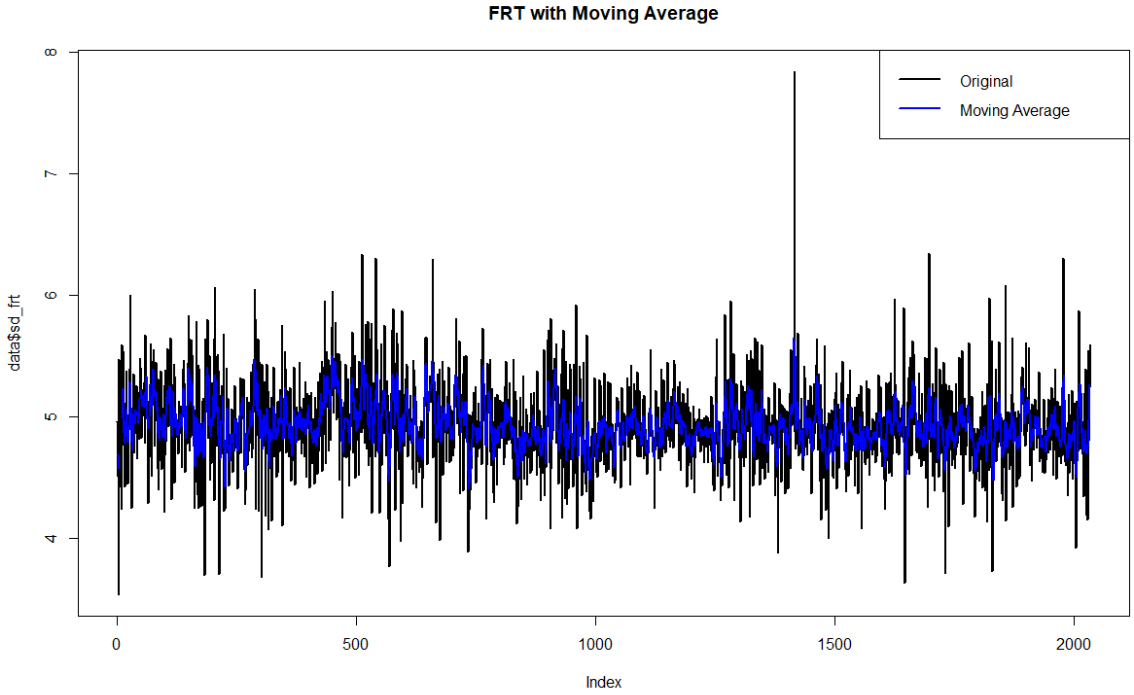


Figure 12. The smoothed standard deviation of FRT by MA.

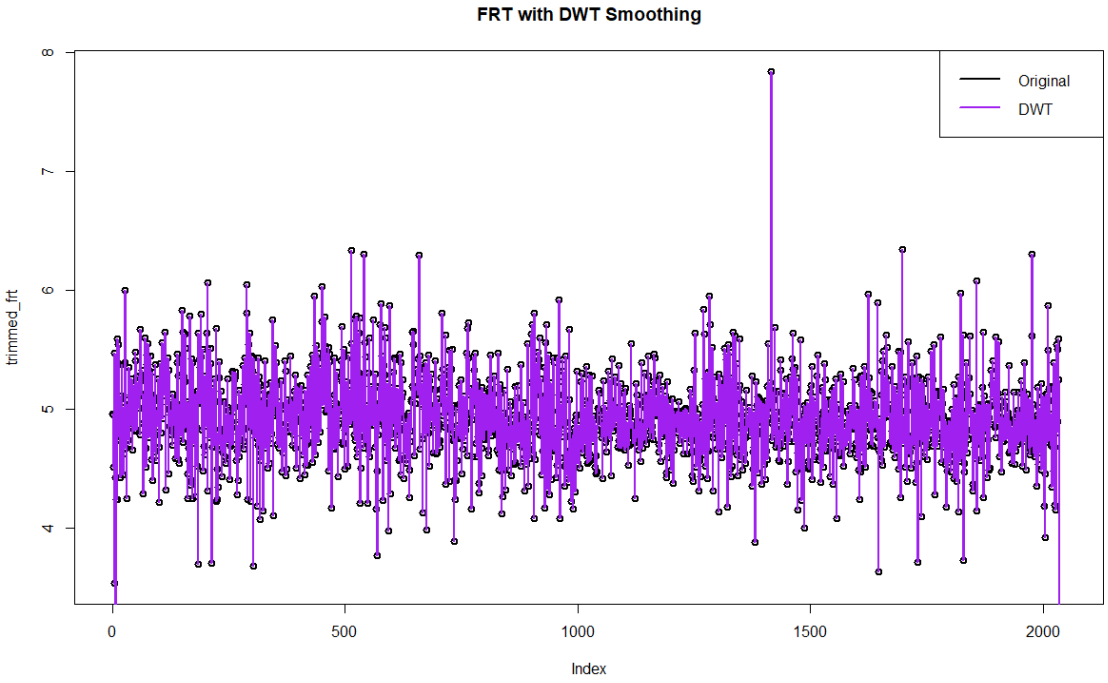


Figure 13. The smoothed standard deviation of FRT by DWT.

Figure 10 shows the standard deviation of FRT over time, comparing the original data (black line) with the smoothed data using the AR model (orange line). The original data exhibit high variability with frequent sharp spikes, indicating fluctuations in the dispersion of FRT measurements. The AR-smoothed data significantly reduce noise and emphasize a stable underlying pattern. Unlike the raw data, the smoothed line captures a consistent trend in variability while filtering out extreme fluctuations.

Figure 11 depicts the standard deviation of FRT over time, comparing the original data (black line) with the smoothed data using the Box-Jenkins ARIMA (cyan line). The original data show frequent sharp fluctuations, reflecting high variability and noise in

the dispersion of FRT. Spikes are particularly noticeable, making it challenging to discern underlying trends. The ARIMA-smoothed data effectively reduce noise while capturing the broader patterns in the variability. The smoothed line highlights a consistent trend over time, filtering out extreme short-term fluctuations while preserving significant shifts in dispersion.

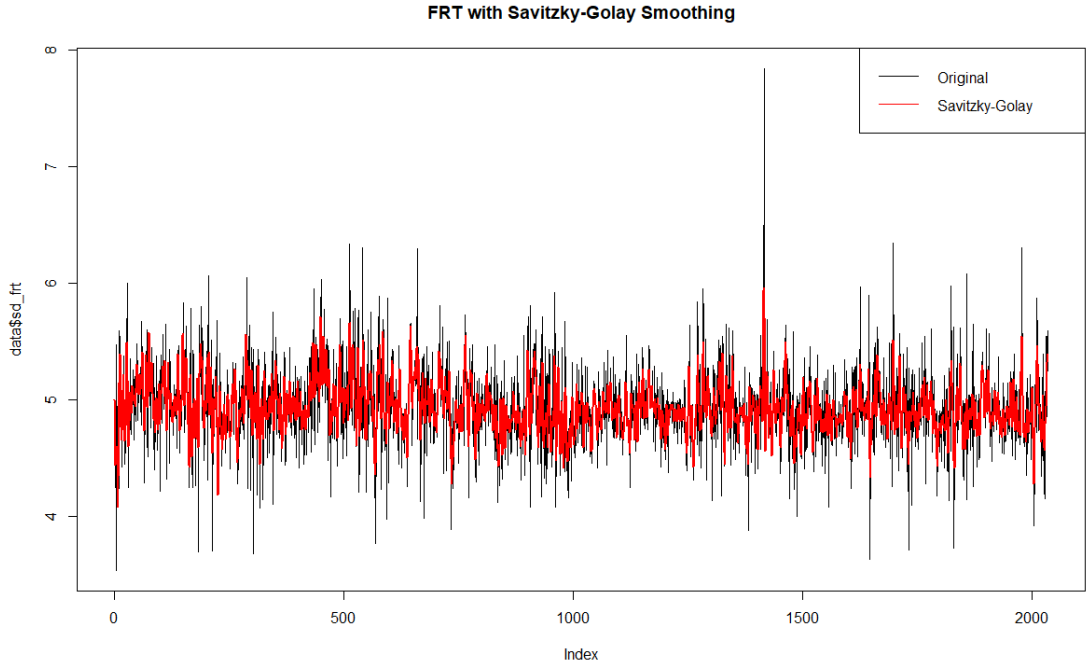


Figure 14. The smoothed standard deviation of FRT by SG filter.

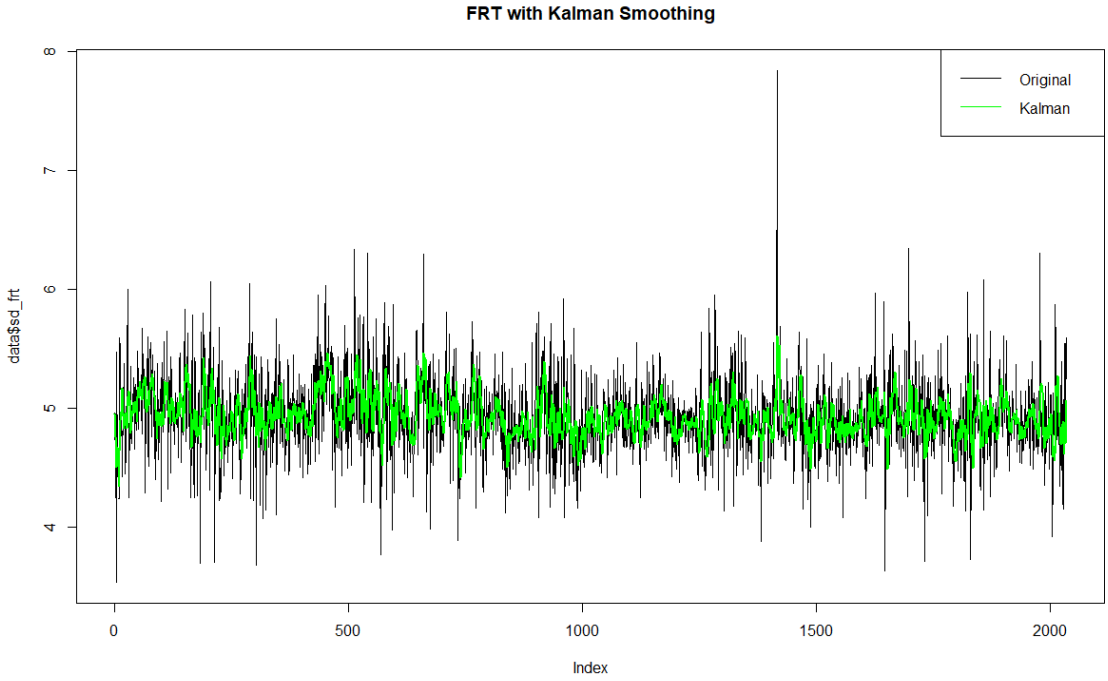


Figure 15. The smoothed standard deviation of FRT by Kalman filter.

Figure 12 compares the standard deviation of FRT over time using the original data (black line) and the MA-smoothed data (blue line). The original data show frequent fluctuations and sharp spikes, indicating substantial short-term variability in the dispersion of FRT. One substantial spike stands out, highlighting a potential anomaly or event-driven

increase in variability. The MA-smoothed data (blue line) effectively reduce the short-term noise, providing a clearer view of the overall trend in variability. The smoothing minimizes sharp deviations while still capturing the general pattern in the data.

Figure 13 compares the standard deviation of FRT over time between the original data (black line) and the smoothed data using DWT (purple line). The DWT-smoothed data (purple line) effectively reduce noise from the original data while retaining the essential characteristics of the original data. It smooths out the rapid fluctuations and highlights the broader trends in the data. The smoothed line adapts well to the structure of the original data while mitigating the influence of outliers, including the significant spike.

Figure 14 compares the standard deviation of FRT over time between the original data (black line) and the smoothed data using Savitzky-Golay smoothing (red line). The Savitzky-Golay smoothed data (red line) effectively reduce noise while preserving the data's overall structure and finer details. Unlike some other smoothing methods, Savitzky-Golay applies polynomial fitting within a sliding window, which minimizes distortion and retains local features. It smooths out the rapid, erratic movements in the original data while still responding to meaningful shifts, including capturing the prominent spike with reduced intensity.

Figure 15 compares the standard deviation of FRT over time between the original data (black line) and the smoothed data using Kalman smoothing (green line). The Kalman smoothed data (green line) effectively reduce the noise and clearly represent the underlying trend. Kalman smoothing dynamically adjusts to the data's structure, allowing it to filter out random fluctuations while preserving significant patterns, including the spike in the middle with reduced prominence.

As shown in Figures 2–8, the average FRT shows no visual difference in trends between the original average and the smoothed average, except for a reduction in the range. However, in Figures 9–15, smoothing the SD of FRT clearly removes outliers, resulting in a notable difference between the raw SD and the smoothed SD. These characteristics are well illustrated in Table 2. The FRT average shows little difference when moving from the raw value to the smooth value, with only a slight reduction in the range. On the other hand, the FRT SD displays a notable effect when transitioning from raw to smooth. While the average SD remains almost unchanged, the SD of the FRT is significantly reduced, indicating that the smoothed SD has effectively removed much of the variability.

Table 2. Comparison between raw FRT and smoothed FRT (n = 1394 days).

	Raw FRT		**Smoothed FRT**	
	Mean	**SD**	**Mean**	**SD**
Daily average	27.46	0.57	27.56	0.41
Daily SD	4.95	0.30	4.93	0.10

3.2. Market Indicators' Stationary Process

Table 3 shows the results of the stationary test (ADF test and Phillips-Perron test). The analysis identified that the time series data for SPY were non-stationary ($p > 0.05$). To solve the original data's non-stationary, log transformation was useful for stabilizing the variance in some cases. However, it failed to ensure stationarity for this SPY dataset. To address this, first-order differencing was applied, successfully achieving stationarity.

Table 3. Stationary test results of market indicators.

	ADF Test	Phillips-Perron Test	Stationary Original	Stationary Log Transformation	Stationary Differencing
	p-value	Z-tau			
SPY	0.3882	−0.9132	Non-stationary	Non-stationary	Stationary
VIXCLS	0.01	−4.7627	Stationary	Stationary	N/A

On the other hand, the VIXCLS series was confirmed to be stationary through both the ADF test and the Phillips–Perron test. Furthermore, even after applying log transformation, the stationary properties of the VIXCLS series were preserved. The omission of differencing was based on the results of a stationarity test. During the stationary testing process, the ADF test is conducted as a priority. If the test determines that original format or log transformation is sufficient, the algorithm proceeds without performing additional differencing. The ADF test confirmed that the log transformation was already stationary, making additional differencing unnecessary. By achieving stationarity through log transformation alone, the data were prepared for further analysis without introducing unnecessary transformations.

Overall, differencing proved to be an effective method for achieving stationarity in the SPY, while original and log transformation did not guarantee stationarity. In the case of VIXCLS, the series is inherently stationary; however, log transformation was applied in this study to enable a more robust analysis.

Figure 16 compares two visualizations of the VIXCLS time series: the original and log-transformed series. The plot on the left shows the raw VIXCLS values over time. The series exhibits significant fluctuations and extreme spikes, with notable peaks around certain time periods, indicating periods of high market volatility. Despite the variability, the original series is stationary, as its statistical properties, such as mean and variance, remain stable over time.

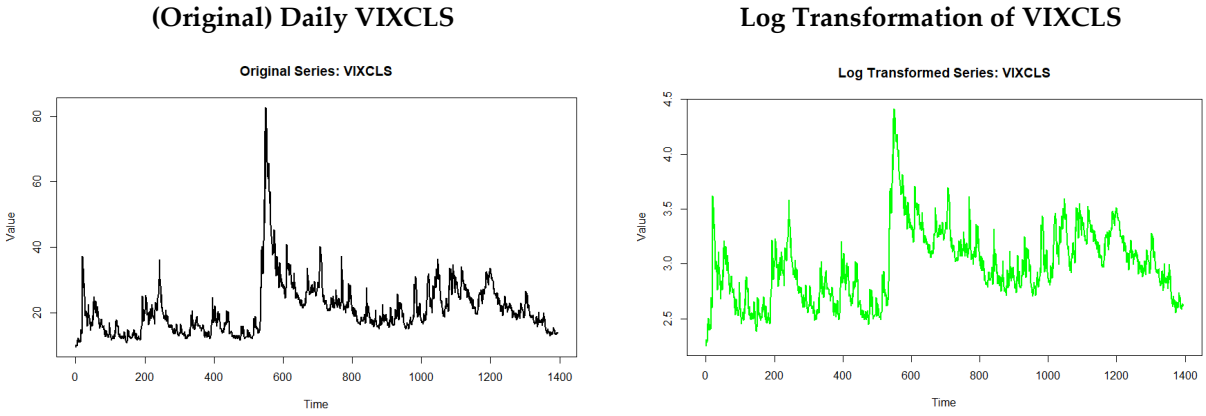


Figure 16. Augmented Dickey-Fuller test and Phillips-Perron test of VIXCLS. (*Left*) The original daily VIXCLS time series, showing significant fluctuations with notable spikes. (*Right*) The log-transformed VIXCLS series, illustrating how the log transformation compresses high values and stabilizes the variance. Augmented Dickey-Fuller (ADF) and Phillips-Perron tests confirm that both the original and log-transformed series exhibit stationarity, although the log transform aids interpretability by reducing the impact of extreme peaks.

The plot on the right displays the log-transformed VIXCLS series. While the log transformation is not required for stationarity, it is applied to stabilize the variance, compressing large spikes and reducing the impact of extreme values. The transformation makes the fluctuations more uniform and manageable for further statistical analysis. This log transformation enhances the interpretability of the VIXCLS data and improves compatibility with

modeling methods that benefit from stabilized variance, even though the original series is already stationary.

Figure 17 illustrates three visualizations related to the daily close price of SPY: the original series (top left), the differenced series (top right), and the log-transformed series (bottom left). First, the plots in the top left show SPY's raw daily closing prices over time. The data exhibit an upward trend with periods of noticeable fluctuations and volatility. This indicates that the price series is non-stationary and has a clear trend component.

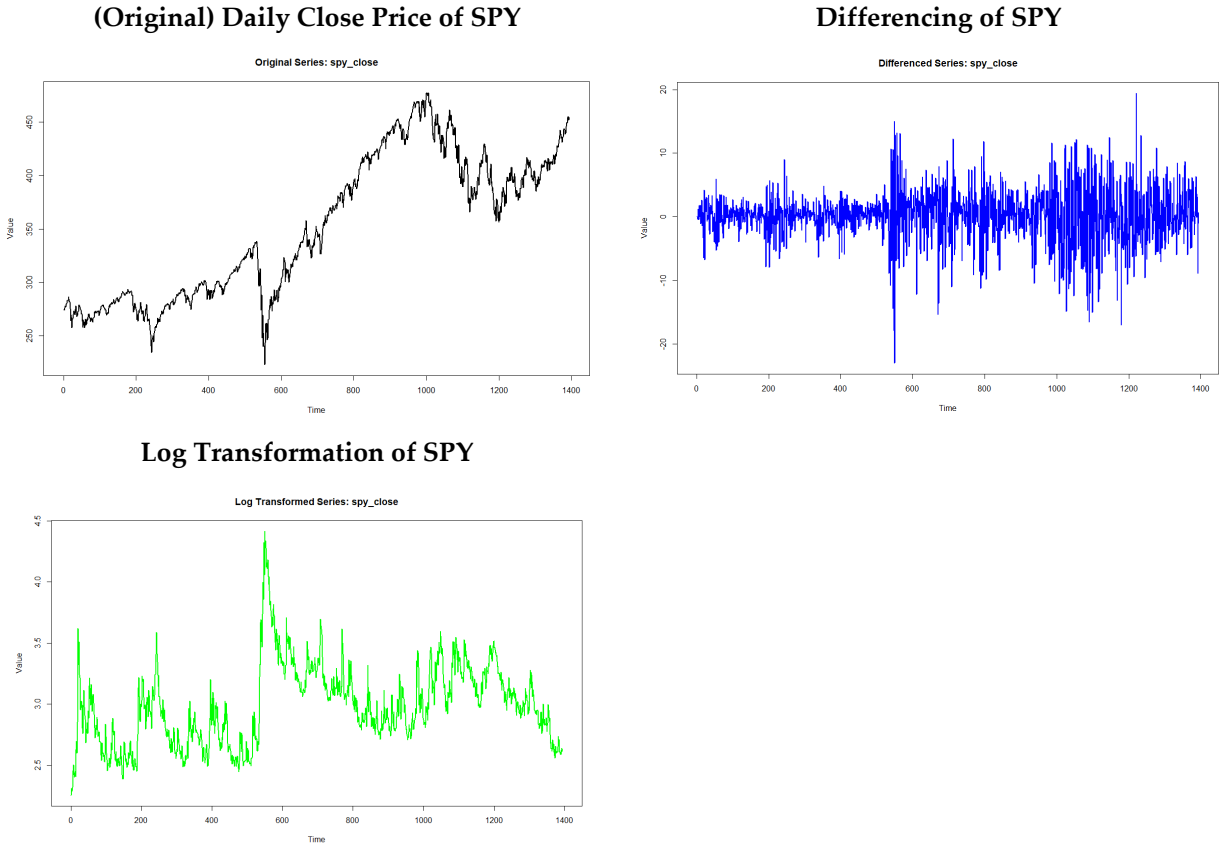


Figure 17. Augmented Dickey-Fuller test and Phillips-Perron test of SPY. (*Top left*) The original daily SPY closing prices, which show a clear upward trend and non-stationary behavior. (*Top right*) The differenced SPY series, indicating that first differencing removes the trend and yields a mean stationary process. (*Bottom left*) The log-transformed SPY series, which stabilizes variance but still retains an underlying trend. The ADF and Phillips-Perron tests confirm that the differenced SPY series meets stationarity assumptions, making it suitable for the predictive models described in this study.

On the other hand, the differenced series at the top right is shown to remove the trend and stabilize the mean of the original series. This process transforms the non-stationary series into a stationary one, making it suitable for time series analysis methods that require stationarity. The differenced data fluctuate around zero, indicating that the trend has been effectively removed.

Finally, the log transformation of SPY at the bottom left displays the log-transformed daily closing prices of SPY. Log transformation is applied to stabilize the variance and reduce the impact of large price movements. The transformed data retain the overall trend of the original series but compress large fluctuations, making the data more suitable for further analysis.

The original series is non-stationary and exhibits both a trend and varying volatility. The differencing process removes the trend, producing a stationary series, while the log

transformation stabilizes variance. These transformations are essential steps in preparing time series data for modeling and forecasting.

4. Empirical Testing

The smoothed FRT scores (mean and standard deviation) can be tested with ML methods: four types of SVMs and ten settings with NNs. Following a general machine learning process, the data were split into a 50:50 ratio, with one-half used for training to develop the model and the other half used for testing to evaluate the predictive accuracy of the trained model. To assess the model performance, the criteria used to compare each model here are root mean squared error (RMSE) and mean absolute error (MAE). Both are widely used metrics for evaluating machine learning algorithms [51].

For the smoothing techniques, the parameters (e.g., α in exponential smoothing, $p/d/q$ for ARIMA, wavelet levels for DWT, and polynomial degree/window size for the Savitzky-Golay filter) were set based on a combination of established best practices and, where relevant, a grid search over candidate values. The Kalman filter parameters (initial state, process noise, and measurement noise covariances) were refined by iteratively minimizing the mean squared error on a holdout portion of the FRT data. The SVMs' kernel coefficients and regularization terms were identified for the ML models through a stepwise grid search employing cross-validation. Similarly, the neural networks' hyperparameters (e.g., number of hidden neurons, learning rate, and number of epochs) were selected by monitoring validation loss to avoid overfitting.

4.1. Prediction Comparison with Market Indicators by SVM

Table 4 compares SVM prediction performance with ln(VIXCLS) using the original FRT and smoothed FRT datasets. The results show that the RBF kernel consistently delivers the best performance across all metrics compared to other kernels (linear, polynomial, and sigmoid). For the original FRT dataset, the RBF kernel achieves the lowest RMSE (0.315) and MAE (0.243) during the training phase and continues to perform best during the testing phase, with an RMSE of 0.335 and MAE of 0.257. Although the Linear kernel performs reasonably well, it is slightly outperformed by the RBF kernel. In contrast, the polynomial and sigmoid kernels exhibit significantly higher error metrics, indicating poor prediction accuracy.

Table 4. SVM prediction comparison with ln(VIXCLS).

	Original FRT				**Smoothed FRT**			
	Training		**Testing**		**Training**		**Testing**	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Linear	0.473	0.382	0.472	0.384	0.399	0.321	0.404	0.324
Polynomial	0.645	0.518	0.675	0.534	0.483	0.385	0.515	0.407
RBF	**0.315**	**0.243**	**0.335**	**0.257**	**0.285**	**0.233**	**0.299**	**0.245**
Sigmoid	8.910	7.663	7.863	7.863	6.962	5.787	6.898	5.785

Note. The model uses the 1-month T-Bill as a control variable for market inflation.

The smoothed FRT dataset further enhances the RBF kernel's performance. During the training phase, the RBF kernel achieves an RMSE of 0.285 and MAE of 0.233, showing a marked improvement compared to the original FRT dataset. In the testing phase, the RBF kernel maintains its superiority, with the lowest RMSE (0.299) and MAE (0.245). The smoothed dataset demonstrates similar trends for other kernels, but the RBF kernel consistently outperforms them.

Overall, the results indicate that the smoothed FRT dataset improves predictive accuracy, particularly when paired with the RBF kernel. The testing RMSE and MAE for

the RBF kernel decrease significantly when using the smoothed dataset, highlighting the benefits of smoothing in enhancing the relationship between FRT and ln(VIXCLS). These findings confirm that the combination of the smoothed FRT dataset and the RBF kernel provides the most robust and accurate predictions.

Table 5 compares the SVM prediction performance with the differencing of SPY (D(SPY)) using both the original FRT and smoothed FRT datasets. For the original FRT dataset, the Linear kernel consistently performs the best among all kernels. During the training phase, it achieves an RMSE of 5.979 and an MAE of 4.972, while in the testing phase, it maintains superior performance with an RMSE of 6.085 and an MAE of 5.096. The RBF kernel follows in performance but is less accurate than the Linear kernel. The polynomial and sigmoid kernels perform significantly worse, with much higher RMSE and MAE values, indicating their unsuitability for predicting D(SPY) in this context.

Table 5. SVM prediction comparison with D(SPY).

	Original FRT				**Smoothed FRT**			
	Training		**Testing**		**Training**		**Testing**	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Linear	**5.979**	**4.972**	**6.085**	**5.096**	**5.765**	**4.685**	**5.941**	**4.837**
Polynomial	17.245	16.180	17.168	16.119	15.731	14.565	15.355	14.080
RBF	6.295	5.305	6.436	5.441	6.207	5.215	6.284	5.283
Sigmoid	10.652	8.912	10.569	8.927	12.593	10.560	12.664	10.675

With the smoothed FRT dataset, the Linear kernel continues to deliver the best results, further improving its accuracy. In the training phase, it achieves the lowest RMSE (5.765) and MAE (4.685), and in the testing phase, it maintains its strong performance with an RMSE of 5.941 and MAE of 4.837. Although the RBF kernel improves slightly with the smoothed dataset, it still underperforms compared to the Linear kernel. Similarly to the original dataset, the polynomial and sigmoid kernels exhibit poor predictive accuracy with significantly higher error values.

The results indicate that the Linear kernel is the most effective for predicting D(SPY) across both datasets, achieving the lowest RMSE and MAE values. Additionally, the smoothed FRT dataset enhances the prediction accuracy of all kernels, particularly the Linear kernel, demonstrating that smoothing improves the predictive relationship between FRT and D(SPY). Overall, the combination of the smoothed FRT dataset and the Linear kernel is the preferred approach for accurate predictions.

4.2. Prediction Comparison with Market Indicators by NN

Table 6 presents the performance of NN models with varying numbers of neurons for predicting ln(VIXCLS) using both the original FRT and smoothed FRT datasets. For the original FRT dataset, the NN model with 90 neurons achieves the best results, with the lowest RMSE (0.270) and MAE (0.211) during the training phase, and an RMSE of 0.302 and MAE of 0.241 during the testing phase. Performance improves as the number of neurons increases, particularly beyond 30 neurons, where the RMSE and MAE values significantly decrease. Smaller models, such as those with 10 neurons, exhibit higher error metrics, indicating lower predictive accuracy.

When using the smoothed FRT dataset, the NN model with 90 neurons also performs the best. It achieves an RMSE of 0.261 and MAE of 0.207 during the training phase, showing a slight improvement over the original FRT dataset. During the testing phase, the 90-neuron model maintains its strong performance, with an RMSE of 0.286 and MAE of 0.225. Similarly

to the original dataset, the performance improves as the number of neurons increases, while smaller models show relatively higher error metrics.

Table 6. NN prediction comparison with ln(VIXCLS).

	Original FRT				**Smoothed FRT**			
	Training		**Testing**		**Training**		**Testing**	
No. of Neurons	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
10	0.782	0.640	0.845	0.691	0.754	0.605	0.786	0.634
20	0.381	0.298	0.414	0.315	0.387	0.304	0.404	0.316
30	0.460	0.358	0.508	0.393	0.471	0.371	0.509	0.398
40	0.329	0.261	0.367	0.282	0.318	0.251	0.339	0.266
50	0.307	0.238	0.338	0.261	0.317	0.251	0.332	0.263
60	0.307	0.238	0.340	0.265	0.301	0.237	0.315	0.249
70	0.300	0.235	0.326	0.255	0.293	0.234	0.313	0.248
80	0.300	0.235	0.328	0.255	0.295	0.237	0.309	0.248
90	**0.270**	**0.211**	**0.302**	**0.241**	**0.261**	**0.207**	**0.286**	**0.225**
100	0.287	0.225	0.321	0.252	0.286	0.225	0.303	0.239

Note. The model uses the 1-month T-Bill as a control variable for market inflation.

Overall, the smoothed FRT dataset enhances the model's performance, particularly during training, as evidenced by slightly lower RMSE and MAE values. The improvements during testing are less pronounced, with the smoothed dataset delivering comparable or slightly better results. However, the smoothed FRT still showed better performance than the original FRT. Across both datasets, the 90-neuron NN model consistently demonstrates superior predictive performance, making it the most effective configuration for capturing the relationship between FRT and ln(VIXCLS). These findings highlight the importance of both dataset preparation and model complexity in achieving optimal predictions.

Table 7 compares the performance of NN models with varying numbers of neurons when predicting D(SPY) using both the original FRT and smoothed FRT datasets. For the original FRT dataset, the NN model with 10 neurons performs the best during testing, achieving the lowest RMSE (4.467) and MAE (3.026). The model maintains competitive results during training as well, with an RMSE of 4.287 and MAE of 3.039. As the number of neurons increases, the RMSE and MAE slightly increase during testing, suggesting that smaller models are better suited for this dataset in predicting D(SPY).

Table 7. NN prediction comparison with D(SPY).

	Original FRT				**Smoothed FRT**			
	Training		**Testing**		**Training**		**Testing**	
No. of Neurons	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
10	**4.287**	**3.039**	**4.467**	**3.026**	**4.302**	**3.048**	**4.457**	**3.016**
20	4.280	3.028	4.473	3.028	4.295	3.047	4.463	3.026
30	4.260	3.022	4.484	3.042	4.288	3.042	4.467	3.031
40	4.254	3.019	4.502	3.047	4.275	3.032	4.480	3.041
50	4.256	3.022	4.491	3.045	4.276	3.033	4.481	3.039
60	4.253	3.016	4.498	3.052	4.272	3.025	4.482	3.039
70	4.250	3.016	4.496	3.052	4.277	3.034	4.478	3.036
80	4.242	3.009	4.506	3.062	4.271	3.033	4.493	3.050
90	4.243	3.001	4.508	3.061	4.269	3.028	4.484	3.045
100	4.231	3.000	4.517	3.074	4.269	3.031	4.488	3.047

For the smoothed FRT dataset, the NN model with 10 neurons also performs the best during testing, with an RMSE of 4.457 and MAE of 3.016. In the training phase, the same model achieves an RMSE of 4.302 and MAE of 3.048, showing slightly higher error values compared to the original dataset. Like the original FRT dataset, larger models (e.g., those with 60 neurons or more) exhibit slightly higher RMSE and MAE values, indicating diminishing returns as model complexity increases.

Across both datasets of the original FRT and smoothed FRT, the NN model with 10 neurons consistently delivers the best predictive performance, achieving the lowest error metrics in both the training and testing phases. The smoothed FRT dataset shows slightly better performance during testing, as seen with the marginally lower RMSE and MAE values compared to the original FRT dataset. This suggests that smoothing improves the predictive relationship between FRT and D(SPY), particularly for simpler models.

As shown in the results, the NN model with 10 neurons is the most effective configuration for predicting D(SPY) across both datasets, with the smoothed FRT dataset offering an advantage in reducing testing errors. These results demonstrate that simpler NN models paired with smoothed datasets can effectively balance prediction accuracy and model efficiency.

One crucial concern raised by the reviewer is that smoothing the FRT data may inadvertently eliminate high-frequency noise and critical signals related to genuine market shifts. To assess this possibility, predictive performances were checked in each ML model. The results compared the performance of our machine learning models—using standard metrics such as RMSE and MAE—both before and after smoothing. It would have expected weaker predictive accuracy if smoothing had compromised crucial information. However, the results indicated that smoothed data generally improved predictive power. It demonstrates that smoothing primarily filters out random fluctuations while retaining the key patterns underlying investor sentiment. At the same time, there is a legitimate risk of excessive smoothing, where abrupt and meaningful changes could be lost. To mitigate this, multiple smoothing techniques (e.g., exponential smoothing, Kalman filter, Savitzky-Golay) aggregated the results to balance any method's potential biases. This multi-pronged approach allowed us to retain the most relevant FRT dynamics while minimizing noise, ultimately enhancing, rather than undermining, the interpretive and predictive utility of the FRT dataset.

5. Discussion, Implication, and Future Studies

This study demonstrates the ability to differentiate between normal and abnormal patterns by smoothing data with a relatively narrow range of fluctuations, such as FRT. Despite the limited daily fluctuations in FRT averages, the results show that smoothing the data reveals important dynamics that are obscured in the raw data. By applying smoothing techniques to the FRT mean and standard deviation, this study uncovers significant relationships with market volatility, measured by VIXCLS and SPY.

The results indicate that the smoothed FRT mean and standard deviation predicted VIXCLS and SPY better than raw FRT. This suggests that the smoothing process filters out noise, enhancing the signal and making the data more suitable for explaining market volatility. The key takeaway is that smoothing helps to emphasize the underlying relationships between investor risk tolerance and market behavior, providing a clearer and more accurate depiction of market trends. These findings are consistent with the concept of noise reduction, as smoothing eliminates short-term volatility that may mask long-term patterns.

Although smoothing can reduce certain high-frequency or short-term fluctuations, the resulting dataset still retains mid- and long-term patterns that may exhibit significant nonlinearity. By applying support vector machines and neural networks to these smoothed

data, this study leverages methods that are well suited to uncover these more complex relationships. Rather than focusing on minute temporal dependencies, the ML approach emphasizes broader structural patterns that persist even after noise reduction. In this manner, smoothing and machine learning can function together: the smoothing process clarifies salient signals, and the subsequent ML modeling uses those signals to capture nuanced, nonlinear interactions that simple time series forecasting might overlook.

The implications of this research are twofold. First, from a methodological standpoint, smoothing techniques such as those applied in this study can significantly improve the explanatory power of FRT data in models related to financial markets. The improvement in the prediction by smoothed FRT demonstrates that noise reduction allows for a more meaningful understanding of FRT's relationship with VIXCLS and SPY. This offers potential for further research in financial risk tolerance and market dynamics, suggesting that smoothed FRT data might be more effective than the raw FRT for predictive modeling in finance.

Second, the findings contribute to a broader understanding of how behavioral finance metrics such as FRT interact with market volatility and sensitivity. The significant relationship between smoothed FRT (mean and standard deviation), VIXCLS, and SPY underscores the potential for using FRT as a valuable indicator for market trends, provided that proper smoothing techniques are applied. However, the concern of overfitting remains, and further research is needed to ensure the application of these findings across different market contexts. Future studies should explore the broader validity of smoothed FRT's relationship with other market indicators, such as the S&P 500 or investor behaviors, to evaluate whether the improvement observed in this study would fit across various financial contexts.

Beyond theoretical contributions, this study's findings have direct implications for real-world financial applications. The smoothed FRT data could be instrumental in portfolio optimization, offering insights into how risk tolerance aligns with market volatility and enabling investors to allocate assets more effectively. Similarly, the integration of smoothed FRT metrics into risk assessment tools could improve the accuracy and reliability of financial institutions' evaluations, particularly in stress-testing scenarios or risk mitigation strategies.

Furthermore, this study highlights the potential for leveraging smoothed FRT data in investor profiling. By reducing noise, these metrics can help financial advisors and institutions better understand their clients' risk preferences and behavioral tendencies. This could lead to the development of tailored investment products and strategies that align individual risk tolerances while accounting for market dynamics.

Finally, emphasizing how noise reduction improves decision-making for both investors and financial institutions enhances the practical relevance of the findings. By offering a clearer picture of market trends and reducing the risk of misinterpretation due to noise, smoothed FRT data can become a cornerstone for developing more robust and adaptive financial tools. Future studies should delve deeper into these applications, exploring how the integration of smoothed FRT data into real-world decision-making processes could drive more informed and effective financial strategies.

Although the Kalman filter was initially employed for smoothing, the filter also supports real-time state estimation. In a real-world setting, it can be updated iteratively as new FRT observations are received, thereby adapting to rapidly changing market conditions. Future analyses aim to implement this update mechanism to capture evolving investor sentiment more dynamically.

While this study primarily employs smoothing techniques (e.g., exponential smoothing, Kalman filtering) to reduce high-frequency noise, the authors acknowledge the recommendation to incorporate time series decomposition, such as Seasonal-Trend Decomposition using Loess (STL), for further isolating long-term trends and seasonal components.

Implementing STL requires a consistent pattern of seasonality within the data, which may not be as pronounced in daily FRT observations due to the nature of participant recruitment and response timing. Additionally, preliminary tests did not reveal strong seasonal effects. Nonetheless, future expansions of this research may revisit STL or alternative decomposition techniques if data collection protocols change to capture clearer cyclical patterns. By then, decomposing the time series into seasonal and residual components could offer a deeper understanding of investor behavior and risk tolerance dynamics.

This study evaluates predictive accuracy using conventional metrics (RMSE, MAE) and does not incorporate advanced feature importance methods such as Shapley Additive Explanations (SHAP). Although SHAP could illuminate how each smoothed FRT measure contributes to market indicator predictions, implementing SHAP for multiple machine learning models requires substantial computational resources and a refined methodological design. Future expansions of this research may adopt SHAP or similar approaches to deepen interpretability and offer clearer insights into how investor sentiment measures interact with broader market dynamics.

Author Contributions: Conceptualization, W.H. and E.K.; Methodology, W.H.; Software, W.H.; Validation, E.K.; Formal analysis, W.H.; Investigation, E.K.; Resources, W.H.; Data curation, W.H.; Writing—original draft, W.H.; Writing—review & editing, E.K.; Visualization, W.H. and E.K.; Supervision, W.H.; Project administration, E.K.; Funding acquisition, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the research fund of Hanyang University (HY-202400000003279).

Data Availability Statement: The research dataset can be obtained upon a proper request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grable, J.; Lytton, R.H. Financial Risk Tolerance Revisited: The Development of a Risk Assessment Instrument. *FSR* **1999**, *8*, 163–181. [CrossRef]
2. Heo, W.; Grable, J.E.; Rabbani, A.G. A Test of the Relevant Association between Utility Theory and Subjective Risk Tolerance: Introducing the Profit-to-Willingness Ratio. *J. Behav. Exp. Financ.* **2018**, *19*, 84–88. [CrossRef]
3. Bali, T.G.; Zhou, H. Risk, Uncertainty, and Expected Returns. *J. Financ. Quant. Anal.* **2016**, *51*, 707–735. [CrossRef]
4. Kamstra, M.J.; Kramer, L.A.; Levi, M.D.; Wermers, R. Seasonal Asset Allocation: Evidence from Mutual Fund Flows. *J. Financ. Quant. Anal.* **2017**, *52*, 71–109. [CrossRef]
5. Grable, J.E.; Lyons, A.C.; Heo, W. A Test of Traditional and Psychometric Relative Risk Tolerance Measures on Household Financial Risk Taking. *Financ. Res. Lett.* **2019**, *30*, 8–13. [CrossRef]
6. Hoffmann, A.O.I.; Post, T.; Pennings, J.M.E. Individual Investor Perceptions and Behavior during the Financial Crisis. *J. Bank. Financ.* **2013**, *37*, 60–74. [CrossRef]
7. Michaelides, A.; Zhang, Y. Stock Market Mean Reversion and Portfolio Choice over the Life Cycle. *J. Financ. Quant. Anal.* **2017**, *52*, 1183–1209. [CrossRef]
8. Heo, W.; Rabbani, A.; Grable, J.E. An Evaluation of the Effect of the COVID-19 Pandemic on the Risk Tolerance of Financial Decision Makers. *Financ. Res. Lett.* **2021**, *41*, 101842. [CrossRef] [PubMed]
9. Hansen, P.R.; Lunde, A.; Nason, J.M. Choosing the Best Volatility Models: The Model Confidence Set Approach. *Oxf. Bull. Econ. Stat.* **2003**, *65*, 839–861. [CrossRef]
10. Tsay, R.S. Analysis of Financial Time Series. In *Wiley Series in Probability and Statistics*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2010; ISBN 978-1-118-01709-8.
11. Woo, G.; Liu, C.; Sahoo, D.; Kumar, A.; Hoi, S. ETSformer: Exponential Smoothing Transformers for Time-Series Forecasting. *arXiv* **2022**, arXiv:2202.01381.
12. Sirisha, U.M.; Belavagi, M.C.; Attigeri, G. Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison. *IEEE Access* **2022**, *10*, 124715–124727. [CrossRef]
13. De Oliveira, E.M.; Cyrino Oliveira, F.L. Forecasting Mid-Long Term Electric Energy Consumption through Bagging ARIMA and Exponential Smoothing Methods. *Energy* **2018**, *144*, 776–788. [CrossRef]

14. Smyl, S. A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [CrossRef]
15. Huang, F.; Qin, T.; Wang, L.; Wan, H.; Ren, J. An ECG Signal Prediction Method Based on ARIMA Model and DWT. In Proceedings of the 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 20–22 December 2019; pp. 1298–1304.
16. Särkkä, S.; Svensson, L. *Bayesian Filtering and Smoothing*; Cambridge University Press: Cambridge, UK, 2023; ISBN 978-1-108-91230-3.
17. Ameriks, J.; Kézdi, G.; Lee, M.; Shapiro, M.D. Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares: The Attenuation Puzzle. *J. Bus. Econ. Stat.* **2020**, *38*, 633–646. [CrossRef]
18. Park, J. Bayesian Filtering and Smoothing, 2nd ed. *J. Am. Stat. Assoc.* **2023**, *118*, 2943–2945. [CrossRef]
19. Liu, Y.; Dang, B.; Li, Y.; Lin, H.; Ma, H. Applications of Savitzky-Golay Filter for Seismic Random Noise Reduction. *Acta Geophys.* **2016**, *64*, 101–124. [CrossRef]
20. Pataky, T.C.; Robinson, M.A.; Vanrenterghem, J.; Challis, J.H. Smoothing Can Systematically Bias Small Samples of One-Dimensional Biomechanical Continua. *J. Biomech.* **2019**, *82*, 330–336. [CrossRef]
21. Shen, C.; He, Y.; Qin, J. Robust Multi-Dimensional Time Series Forecasting. *Entropy* **2024**, *26*, 92. [CrossRef]
22. Cboe VIX FAQ. Available online: https://www.cboe.com/tradable_products/vix/faqs/ (accessed on 11 October 2024).
23. About Us. Available online: https://www.spglobal.com/spdji/en/about-us/ (accessed on 19 January 2025).
24. Zhu, X.; Bao, S. Multifractality, Efficiency and Cross-Correlations Analysis of the American ETF Market: Evidence from SPY, DIA and QQQ. *Phys. A Stat. Mech. Its Appl.* **2019**, *533*, 121942. [CrossRef]
25. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer Texts in Statistics; Springer International Publishing: Cham, Switzerland, 2016; ISBN 978-3-319-29852-8.
26. Trasberg, T.; Soundararaj, B.; Cheshire, J. Using Wi-Fi Probe Requests from Mobile Phones to Quantify the Impact of Pedestrian Flows on Retail Turnover. *Comput. Environ. Urban Syst.* **2021**, *87*, 101601. [CrossRef]
27. Yadav, S.K.; Akhter, Y. Statistical Modeling for the Prediction of Infectious Disease Dissemination with Special Reference to COVID-19 Spread. *Front. Public Health* **2021**, *9*, 645405. [CrossRef]
28. Chiarella, C.; He, X.-Z.; Hommes, C. A Dynamic Analysis of Moving Average Rules. *J. Econ. Dyn. Control.* **2006**, *30*, 1729–1753. [CrossRef]
29. Tu, Y.; Ball, M.O.; Jank, W.S. Estimating Flight Departure Delay Distributions-A Statistical Approach with Long-Term Trend and Short-Term Pattern. *J. Am. Stat. Assoc.* **2008**, *103*, 112–125. [CrossRef]
30. Zhang, D. Wavelet Transform. In *Fundamentals of Image Data Mining*; Texts in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; pp. 35–44. ISBN 978-3-030-17988-5.
31. Luo, J.; Ying, K.; He, P.; Bai, J. Properties of Savitzky–Golay Digital Differentiators. *Digit. Signal Process.* **2005**, *15*, 122–136. [CrossRef]
32. Krishnan, S.R.; Seelamantula, C.S. On the Selection of Optimum Savitzky-Golay Filters. *IEEE Trans. Signal Process.* **2013**, *61*, 380–391. [CrossRef]
33. Khodarahmi, M.; Maihami, V. A Review on Kalman Filter Models. *Arch. Comput. Methods Eng.* **2023**, *30*, 727–747. [CrossRef]
34. Liang, Y.; Wen, H.; Nie, Y.; Jiang, Y.; Jin, M.; Song, D.; Pan, S.; Wen, Q. Foundation Models for Time Series Analysis: A Tutorial and Survey. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Barcelona, Spain, 25 August 2024; pp. 6555–6565.
35. Li, L.; Su, X.; Zhang, Y.; Lin, Y.; Li, Z. Trend Modeling for Traffic Time Series Analysis: An Integrated Study. *IEEE Trans. Intell. Transport. Syst.* **2015**, *16*, 3430–3439. [CrossRef]
36. Shi, J.; Wang, S.; Qu, P.; Shao, J. Time Series Prediction Model Using LSTM-Transformer Neural Network for Mine Water Inflow. *Sci. Rep.* **2024**, *14*, 18284. [CrossRef] [PubMed]
37. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; Otexts: Melbourne, Australia, 2018.
38. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. Time Series Analysis: Forecasting and Control. In *Wiley Series in Probability and Statistics*, 4th ed.; John Wiley & Sons, Inc: Hoboken, NJ, USA, 2016; ISBN 978-1-118-67502-1.
39. Qu, L.; Lyu, J.; Li, W.; Ma, D.; Fan, H. Features Injected Recurrent Neural Networks for Short-Term Traffic Speed Prediction. *Neurocomputing* **2021**, *451*, 290–304. [CrossRef]
40. Chen, G.; Wang, W.; Huo, C. Dynamic Modelling of Ship Using Gaussian Processes and Sg Filter. *J. Phys. Conf. Ser.* **2021**, *1861*, 012057. [CrossRef]
41. Paparoditis, E.; Politis, D.N. The Asymptotic Size and Power of the Augmented Dickey–Fuller Test for a Unit Root. *Econom. Rev.* **2018**, *37*, 955–973. [CrossRef]
42. Vogelsang, T.J.; Wagner, M. A Fixed- *b* Perspective on the Phillips–Perron Unit Root Tests. *Econom. Theory* **2013**, *29*, 609–628. [CrossRef]
43. Pisner, D.A.; Schnyer, D.M. Support Vector Machine. In *Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 101–121. ISBN 978-0-12-815739-8.

44. Diederich, J. Rule Extraction from Support Vector Machines: An Introduction. In *Rule Extraction from Support Vector Machines*; Diederich, J., Ed.; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2008; Volume 80, pp. 3–31. ISBN 978-3-540-75389-6.
45. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon* **2018**, *4*, e00938. [CrossRef] [PubMed]
46. Hasnat, A.; Salimullah, M. Causality of Interest Rate, Exchange Rate and Stock Prices in The Chicago Board Options and Exchange. *J. Econ. Financ.* **2020**, *11*, 38–52.
47. Chu, Q.C.; Hsieh, W.G. Pricing Efficiency of the S&P 500 Index Market: Evidence from the Standard & Poor's Depositary Receipts. *J. Futures Mark.* **2002**, *22*, 877–900.
48. Investment Risk Tolerance Assessment//College of Agriculture, Food and Natural Resources. Available online: https://cafnr.missouri.edu/divisions/division-of-applied-social-sciences/research/investment-risk-tolerance-assessment/ (accessed on 19 January 2025).
49. Consumer Price Index (CPI) Databases. Available online: https://www.bls.gov/cpi/data.htm (accessed on 15 February 2025).
50. Resource Center. Available online: https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve (accessed on 15 February 2025).
51. Chicco, D.; Warrens, M.J.; Jurman, G. The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef] [PubMed]