Article

A Functional Data Approach for Continuous-Time Analysis Subject to Modeling Discrepancy under Infill Asymptotics

Tao Chen [1,2,3,4,*,†], **Yixuan Li** [2,5,†] and **Renfang Tian** [2,6,†]

1 Department of Economics, Cross Appointed to Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada
2 Big Data Research Lab, University of Waterloo, Waterloo, ON N2L 3G1, Canada; liy257@macewan.ca (Y.L.); rtian2@uwo.ca (R.T.)
3 Labor and Worklife Program, Harvard University, Cambridge, MA 02138, USA
4 Ordered Number Technology Inc., Shanghai 200131, China
5 Department of Anthropology, Economics and Political Science, MacEwan University, Edmonton, AB T5J 4S2, Canada
6 School of Management, Economics, and Mathematics, King's University College at Western University, London, ON N6A 2M3, Canada
* Correspondence: t66chen@uwaterloo.ca
† These authors contributed equally to this work.

Abstract: Parametric continuous-time analysis often entails derivations of continuous-time models from predefined discrete formulations. However, undetermined convergence rates of frequency-dependent parameters can result in ill-defined continuous-time limits, leading to modeling discrepancy, which impairs the reliability of fitting and forecasting. To circumvent this issue, we propose a simple solution based on functional data analysis (FDA) and truncated Taylor series expansions. It is demonstrated through a simulation study that our proposed method is superior—compared with misspecified parametric methods—in fitting and forecasting continuous-time stochastic processes, while the parametric method slightly dominates under correct specification, with comparable forecast errors to the FDA-based method. Due to its generally consistent and more robust performance against possible misspecification, the proposed FDA-based method is recommended in the presence of modeling discrepancy. Further, we apply the proposed method to predict the future return of the S&P 500, utilizing observations extracted from a latent continuous-time process, and show the practical efficacy of our approach in accurately discerning the underlying dynamics.

1. Introduction

Much of finance and economics is about the study of dynamics over time, in which analysis using time-series data plays a vital part. Despite discrete observations in practice, many time-series data such as stock prices, interest rates, and GDP are essentially drawn from their continuous-time underlying processes, for which consistent estimation can be achieved from continuous-time analyses but not necessarily from their discrete-time counterparts. Thus, the former has become increasingly incorporated by modern time-series analysis (e.g., [1–4]).

In traditional parametric continuous-time analyses, the modeling routinely originates in a discrete-time setting and is then extended to a continuous-time formulation as the length of the time window shrinks towards zero [3]. However, such derivation often imposes strong restrictions while specifying the convergence rates of frequency-dependent (hereafter, f.d.) parameters, resulting in models that are unadaptable to real-world data and

thus forecasting failure [4–12]. For instance, with different sets of conditions regarding the converging speeds of the parameters, many GARCH-like processes have various diffusion processes (e.g., [6,9,11,13]). Moreover, through a testing procedure involving discrimination between the classes of deterministic and stochastic volatility models, ref. [7] discovered that, unlike the traditional stochastic view of short-term interest rates, the Japanese short-term interest rate indeed follows a deterministic pattern. Consequently, there arises discrepancy in continuous-time modeling, as it is debatable which assumption is the correct one (e.g., [14], pp. 176–178, [15]), and mistaking the assumption will lead to a misspecified limit and thereby unreliable analysis results. In such circumstances, nonparametric approaches appear to be appealing tools that can adapt to the true limits based on their data-driven nature and bypass this discrepancy in continuous-time analysis [16].

There is a large body of literature demonstrating the effectiveness of nonparametric methods for accurate estimation and forecasting in cases where parametric assumptions are deemed to be inadequate (e.g., [17–26]). Many studies incorporate nonparametric methods to approximate the density of the states in the absence of a closed-form expression, so that they could use maximum likelihood estimation (MLE) for continuous-time diffusion models (e.g., [22,27,28]). However, they still presume the parametric format of the underlying processes. Another stream of literature uses kernel-based methods to estimate the processes that are nonanticipative smooth functions with unknown structures (e.g., [23]) and make forecasts through conditional density estimation (e.g., [19,20,29,30]). It has been shown, however, that traditional kernel estimators can become inconsistent as the sampling density grows despite the underlying processes becoming further revealed [18,31], and the infill asymptotics (i.e., the asymptotic properties achieved as the sample becomes increasingly dense [32]) require the careful consideration of the dependence among observations, which is substantial work [33].

The main contribution of our paper is to propose a fitting and forecasting approach from the viewpoint of functional data analysis (FDA) to accommodate f.d. data structures and make good use of high-frequency data in order to achieve robust infill asymptotics. Indeed, the FDA fitting method, by design, deals with a collection of subsets of finite-dimensional parameter spaces, which becomes richer and denser with an enlarging sample, and the target function can be consistently estimated by optimizing an empirical criterion (e.g., [34–36]). For such a feature, in our method we employ the FDA-based approach to fit the underlying process using local polynomial bases and obtain the forecasts by extending the movement of the process based on the boundary derivatives of the functional fitting. FDA is renowned for its ability to uncover the dynamics of unknown continuous processes without necessitating stringent assumptions about the data structure [26,37,38]. However, to the best of our knowledge, its application in tracing the f.d. data structures and performing forecasting in continuous-time analysis remains underexplored in the existing literature. Thus, with a simulation study, this paper illustrates the procedure and properties of the proposed FDA-based method and emphasizes that this method excels when the convergence rate of f.d. parameters is undefined, effectively mitigating the common problem of misspecification prevalent in parametric estimations. In addition, we remark that functional approaches can be further developed in this context to achieve desired infill asymptotics with bounded or unbounded domains. To improve readability, all proofs are placed in Appendix A.

In addition to the simulation analysis, we substantiate the practical utility of the proposed method in forecasting S&P 500 prices, a topic of enduring significance that has consistently commanded the interest of financial practitioners and researchers within the field. Throughout the years, various statistical and mathematical methods have emerged to offer insights into stock price trends including but not limited to time-series analysis (e.g., [39–41]); machine learning algorithms (e.g., [22,42,43]); technical indicators (e.g., [44–46]); and fundamental analysis (e.g., [47–49]). However, to the best of our knowledge, the utilization of FDA for future stock price predictions appears to be infrequent. This paper adopts the FDA-based approach to predict the S&P 500 log-returns. Through the root mean squared forecast error

(RMSFE) among different sample sizes and frequencies, we examine the performance of the proposed method in forecasting the process in a given step ahead.

The rest of this paper is structured as follows. In Section 2, we explain the FDA-based method, followed by a discussion of the large-sample properties. Section 3 sets up a simulation study and takes strong GARCH(1,1) as a motivating example to show that in continuous-time analysis, while parametric methods—such as MLE—encounter discrepancy, the proposed FDA-based method can provide robust and reliable estimation and forecasting. We then present an application of the proposed method in predicting future S&P 500 prices in Section 4, and Section 5 concludes the paper.

2. Methods

In this section, we explain the functional fitting and forecasting methods, followed by a discussion of the large-sample properties of the estimators. The method concerns target functions that are continuous but not necessarily differentiable. While smoothness assumptions are often imposed in FDA to accommodate incomplete and noisy observations, as noted by [50], we construct the functional estimator through the Bernstein polynomial approximation, which induces the convergence that requires only the continuity of the underlying process.

2.1. Fitting and Forecasting

Consider a complete probability space $(\Omega, \mathcal{F}, P)$ and a re-scaled bounded time interval $[0, 1]$. The sample path of the stochastic process $\{X(t) : t \in [0, 1]\}$ for state $\omega \in \Omega$ over $[0, 1]$ is denoted as X_ω. Realizations are drawn at countably many time points with observational error $\epsilon_{\omega,t}$; as such,

$$X_{\omega,t} = X_\omega(t) + \epsilon_{\omega,t}. \tag{1}$$

In parametric modeling, $X_\omega(t)$ is the conditional mean with a specified functional form, whereas in FDA, $X_\omega(t)$ is the underlying sample path assumed to be continuous at any given point t. Thus, in the discussion below, we establish the properties of our fitting and forecasting quantities through an arbitrarily close approximation of $X_\omega(t)$ by the Bernstein polynomial $B_K(t, X_\omega)$, so that for each $K \in \mathbb{N}$,

$$B_K(t, X_\omega) := \sum_{k=0}^{K} X_\omega\left(\frac{k}{K}\right) \times \binom{K}{k} \times t^k (1-t)^{K-k}, \tag{2}$$

whence we have the following lemma.

Lemma 1. *Consider $B_K(t, X_\omega)$ as in Equation (2) with the continuous sample path X_ω for any given $\omega \in \Omega$. For any $\varepsilon > 0$, there exists some $\delta(\varepsilon) > 0$ that induces $|t_1 - t_2| \leq \delta(\varepsilon) \implies |X_\omega(t_1) - X_\omega(t_2)| \leq \varepsilon/2$ for all $t_1, t_2 \in [0, 1]$. Then, for all $K \geq 4\sup_{t\in[0,1]} |X_\omega(t)|/(\delta^2(\varepsilon)\varepsilon)$, the following holds:*

$$\sup_{t\in[0,1]} |B_K(t, X_\omega) - X_\omega(t)| \leq \varepsilon.$$

Lemma 1 suggests that, fixing everything else, for any given ε, a high-degree polynomial is needed for approximation if the continuity of X_ω requires a very small $\delta(\varepsilon)$, while a low-degree polynomial can achieve the desired approximation if a large $\delta(\varepsilon)$ suffices for the given ε (see, e.g., Theorem 5.14 [51]). This lemma also implies that while fitting any continuous process, the typical smoothness assumptions for functional data can be imposed on the Bernstein approximating polynomials $B_K(t, X_\omega)$, and the size of the remaining estimation error can be regulated by a high-degree $B_K(t, X_\omega)$. There are three remarks following Lemma 1. First, this lemma is based on the continuity of $X_\omega(t)$, which does not require the continuity of the observed series. Second, the lemma holds for any continuous sample path and thus can also be applicable when $X_\omega(t)$ itself is a diffusion

process of Hölder continuity. Last but not the least, there are various methods to obtain approximations to a continuous function with desirably small errors (Chapter 10 [52]), and here we adopt the Bernstein polynomial for the delivery of the discussion. Hence, the establishment of the asymptotics in the current study is achieved with the help of the parametrization of $B_K(t, X_\omega)$, while in practice, we do not directly estimate or determine the degree K in $B_K(t, X_\omega)$. Instead, the degree of the approximating polynomials will be controlled through the construction of the functional estimator, as explained below.

In the spirit of FDA [38], one can construct a B-spline representation of a certain degree for the polynomial approximation $B_K(t, X_\omega)$, such that $B_K(t, X_\omega) = \mathbf{\Phi}^\top(\cdot)\mathbf{C}$, where $\mathbf{\Phi}$ is a B-spline basis that contains a vector of Q basis functions, $\mathbf{C}_\omega$ is a vector of Q corresponding basis coefficients, and the superscript "⊤" indicates the transpose operation. Then, Equation (1) can be rewritten as

$$X_{\omega,t} \approx \mathbf{\Phi}^\top(t)\mathbf{C}_\omega + \epsilon_{\omega,t}. \tag{3}$$

With the predetermined basis functions $\mathbf{\Phi}$ and the J observations $\{X_{t_j}\}_{j=1}^J := \{X_{\omega,t_j}\}_{j=1}^J$ at $\{t_j\}_{j=1}^J \subseteq [0, 1-\Delta]$ for some $0 < \Delta < 1$, the coefficients $\mathbf{C}_\omega$ are estimated by minimizing the penalized squares $J^{-1}\sum_{j=1}^J [X_{t_j} - \tilde{X}(t_j)]^2 + \lambda \int_0^1 \tilde{X}^{(2)}(t)dt$ with a tuning parameter λ and some fitted function $\tilde{X}(\cdot)$, which yields

$$\tilde{\mathbf{C}}_\omega := \left[\frac{1}{J}\sum_{j=1}^J \mathbf{\Phi}(t_j)\mathbf{\Phi}^\top(t_j) + \lambda \int_0^1 \mathbf{\Phi}^{(2)}(t)\left\{\mathbf{\Phi}^{(2)}(t)\right\}^\top dt\right]^{-1} \frac{1}{J}\sum_{j=1}^J \mathbf{\Phi}(t_j)X_{\omega,t_j}, \tag{4}$$

where the superscript "(2)" indicates the second-order derivative of a function (the second-order derivative of the fitted function is one option for the roughness penalty—other penalties can also be applied depending on the situation). In practice, Q and λ can be selected through certain data-driven algorithms (e.g., [38]). Henceforth, the fitted process of $X_\omega(t)$ is defined as

$$\tilde{X}_\omega(t) := \mathbf{\Phi}^\top(t)\tilde{\mathbf{C}}_\omega, \ \forall t. \tag{5}$$

By choosing a B-spline basis of a proper degree larger than some value R for the fitted process $\tilde{X}_\omega(t)$, the Δ-step-ahead predictor at any given t, $\hat{X}_\omega(t + \Delta)$ is defined as a truncated Taylor expansion with the first R derivatives of the fitted functions, such that

$$\hat{X}_\omega(t + \Delta) := \sum_{r=0}^R \frac{1}{r!}\Delta^r \tilde{X}_\omega^{(r)}(t) = \sum_{r=0}^R \frac{1}{r!}\Delta^r \tilde{\mathbf{C}}_\omega^\top \mathbf{\Phi}^{(r)}(t) \quad \forall t \in (0, 1-\Delta]. \tag{6}$$

The validity of the above Taylor expansion requires the fitted function $\tilde{X}_\omega(t)$ to be at least R times continuously differentiable, while the underlying process $X_\omega(t)$ only needs to be continuous. We want the relationship between the order R of continuous differentiability and the degree K of the approximated function $B_K(t, X_\omega)$ to be such that $R = K + 1$, as the Taylor polynomial approximation will have a non-zero residual if $B_K(t, X_\omega)$ is of a higher degree (i.e., if $R - 1 < K$), while a higher degree of the functional estimator than necessary (i.e., $R - 1 > K$) will bring extra volatility to the fitting and forecasting values. In practice, R can also be selected through certain data-driven methods, such as that explained in the simulation and empirical studies below. Due to the features of the B-spline basis employed, one can adopt the order of $R + 3$ to ensure the desired differentiability of $\tilde{X}_\omega(t)$. The half-open interval $(0, 1-\Delta]$ indicates that, technically, the forecast cannot be achieved without any observation, and the forecast can be obtained as far as $t + \Delta = 1$. Furthermore, although the derivatives $\tilde{X}_\omega^{(r)}(t)$ are applied in the approximation, we do not actually need the convergence of these derivatives to, even if they exist, the derivatives of the truth—only the convergence of the forecasting function in the level is of interest.

2.2. Large-Sample Properties

　　We now discuss the consistency and the asymptotic normality of our functional predictors based on the following assumptions.

Assumption 1. *Consider a continuous sample path $X_\omega(t)$ where for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that $|t_1 - t_2| \le \delta(\varepsilon)$ implies $|X_\omega(t_1) - X_\omega(t_2)| \le \varepsilon/2$. Then, we have the following:*

(a)　　*The sample path $X_\omega(t)$ is observed on a set of evenly spaced time points.*

(b)　　*The observational error ϵ_{ω,t_j} is uncorrelated across $\{t_j\}_{j=1}^J \subseteq [0,1]$, with $\mathbb{E}[\epsilon_{\omega,t} \mid X_\omega(s)] = 0$ for $s < t$ and $Var[\epsilon_{\omega,t}] < c < \infty$ for all $t \in [0,1]$ and some constant c.*

(c)　　*$Q \sim J^{\alpha_1}$ and $\lambda \sim J^{\alpha_2}$ with $0 < \alpha_1 < 1$ and $\alpha_2 < 0$.*

　　For simplicity, we assume in Assumption 1(a) that the observations are equally spaced without a loss of generality. A less restrictive version would be to assume that the ratio between the lengths of the largest and the smallest time windows is bounded above and away from zero as in [53]. Part (b) states a sufficient condition to achieve the consistency of the functional estimator $\tilde{X}_\omega$ as in Equation (5), while a proper "low correlation" assumption for this error term is also sufficient. Finally, part (c) indicates that the dimension of the basis expansion shall increase with the sample size to allow for a consistent functional estimator $\tilde{X}_\omega$, and, meanwhile, the tuning parameter shall be kept to $o(1)$ so that the estimation error introduced by the roughness penalty dies down as the sample size grows towards infinity.

　　For any given $0 < \Delta < 1$, we consider the convergence of the forecasting process $\hat{X}_\omega$ over $(\Delta, 1)$ or, equivalently, over $t \in (0, 1 - \Delta)$. Note that even though the forecasting process can be defined at the point $t + \Delta = 1$ as in Equation (6), the asymptotics exclude the point of $t = 1 - \Delta$ due to the limited boundary performance.

Theorem 1 (Convergence of $\hat{X}_\omega$ given Δ). *Consider a continuous sample path X_ω on $[0,1]$ such that for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$, which induces $|t_1 - t_2| \le \delta(\varepsilon) \implies |X_\omega(t_1) - X_\omega(t_2)| \le \varepsilon/2$.*

　　Let $\mathcal{F}_K$ be the collection of all the (local) polynomials over $[0, 1 - \Delta]$ that are at least $K + 1$ times continuously differentiable, and $F_K : \mathcal{F}_K \to \mathbb{R}$ be the linear functional such that $F_K(\psi) := \langle \hat{X}_\omega(\cdot + \Delta) - B_K(\cdot + \Delta, X_\omega), \psi \rangle$ for $\psi \in \mathcal{F}_K$.

　　If Assumption 1 holds, then for any given K and $R = K + 1$, $F_K(\psi) \xrightarrow{p} 0$ as $J \to \infty$ for all $\psi \in \mathcal{F}_K$, whence under $K \to \infty$, $\sup_{t \in (0, 1-\Delta)} |\hat{X}_\omega(t + \Delta) - X_\omega(t + \Delta)| \xrightarrow{p} 0$ as $J \to \infty$.

　　Theorem 1 establishes the convergence of the forecasting process $\hat{X}_\omega(\cdot + \Delta)$ through its weak convergence to the Bernstein approximating polynomial $B_K(\cdot + \Delta, X_\omega)$ and the uniform convergence of $B_K(\cdot, X_\omega)$ to X_ω. Specifically, for any given K and $R = K + 1$, $\hat{X}_\omega(\cdot + \Delta)$ weakly converges to the approximating polynomial $B_K(\cdot + \Delta, X_\omega)$ with the enlarging sample size J, while $B_K(\cdot, X_\omega)$ uniformly converges to the true path X_ω with an enlarging K and regardless of the sample size J. Henceforth, $\hat{X}_\omega(\cdot + \Delta)$ converges uniformly to the true path X_ω with an enlarging K and J. For $R = K + 1$ with $K \ge 4 \sup_{t \in [0,1]} |X_\omega(t)| / (\delta^2(\varepsilon)\varepsilon)$, it is worth noting that for a B-spline basis of order $R + 3$ to span the function space, the number of basis functions Q needs to be such that $Q \ge 2(R + 2) + 1$, while Q grows towards infinity at a lower speed than J; hence, we also have $K = o(J)$ as $K, J \to \infty$.

　　Now we explore the asymptotic distribution of the functional prediction $\hat{X}_\omega$. However, before we state any further assumptions or the resulting asymptotic properties, it is important to note that the realizations are subject to observational error $\epsilon_{\omega,t}$, the behavior of which can impact the asymptotic distribution of $\hat{X}_\omega$. For example, when a continuous sample path X_ω is observed without noise, it is more plausible to consider the functional fitting as an interpolation—in this case, we do not necessarily have asymptotic normality for $\hat{X}_\omega$. On the other hand, if X_ω is observed with noise, we can establish pointwise asymptotic normality under the following sufficient conditions:

Assumption 2. *(a)*   *The noise ϵ_{t_j} is independent across all $\{t_j\}_{j=1}^{J} \subseteq [0,1]$.*

(b)   *$Q \sim J^{\alpha_1}$ and $\lambda \sim J^{\alpha_2}$ with $\alpha_1 > 0$ and $\alpha_2 < 0$, such that $Q^2\lambda = o(J^{-1/2})$ and $J^{-1}Q^2 = O(J^{-1/2})$.*

Assumption 2(a) states a sufficient condition for generating the asymptotic normality by the Lyapunov CLT, which allows for non-identically distributed variables. If one has mixing random processes, a different version of this assumption can be specified by imposing identical distribution, where a different CLT can be applied without affecting the result regarding asymptotic normality. Part (b) assumes a stronger condition on the orders of the two parameters Q and λ, so that the non-normal part of the estimation error is dominated and will not be inflated while deriving the asymptotic distribution.

Theorem 2. *Consider a continuous sample path X_ω on $[0,1]$ where for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that $|t_1 - t_2| \leq \delta(\varepsilon)$ implies $|X_\omega(t_1) - X_\omega(t_2)| \leq \varepsilon/2$.*

Under Assumptions 1 and 2, the following asymptotic normality holds for all $t \in (0, 1 - \Delta]$, $R \to \infty$, and $J \to \infty$:

$$[V(t; \Delta, R, \mathbf{\Phi})]^{-1/2}[\hat{X}_\omega(t + \Delta) - X_\omega(t + \Delta)] \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\sigma_j^2 := Var\left[\epsilon_{t_j}\right]$ for all j, and

$$V(t; \Delta, R, \mathbf{\Phi}) := \frac{1}{J^2} \sum_{j=1}^{J} \sigma_j^2 A_{t_j}^2(t; \Delta, R, \mathbf{\Phi}),$$

$$A_{t_j}(t; \Delta, R, \mathbf{\Phi}) := \sum_{r=0}^{R} \frac{1}{r!} \Delta^r \left\{\mathbf{\Phi}^{(r)}(t)\right\}^{\top} \left\{\int_0^1 \mathbf{\Phi}(s)\mathbf{\Phi}^{\top}(s)ds\right\}^{-1} \mathbf{\Phi}(t_j).$$

Theorem 2 constructs the pointwise asymptotic normality of the functional predictor $\hat{X}_\omega$ through the asymptotically normal prediction error from $\hat{X}_\omega$ to the Bernstein approximating polynomial $B_K(t, X_\omega)$ and the convergence of $B_K(t, X_\omega)$ to X_ω as K increases towards infinity. Specifically, $\hat{X}_\omega(t + \Delta) - X_\omega(t + \Delta)$ can be decomposed into $\hat{X}_\omega(t + \Delta) - B_K(t + \Delta, X_\omega)$ and $B_K(t + \Delta, X_\omega) - X_\omega(t + \Delta)$, where the former is inflated by $V(t; \Delta, R, \mathbf{\Phi})$ to an asymptotic normal, and the latter is dominated because a desirably small approximation error $B_K(t + \Delta, X_\omega) - X_\omega(t + \Delta)$ can be achieved by a sufficiently large K (and R). Meanwhile, since the inverse factorial series $\sum_{r=0}^{\infty}(r!)^{-1}$ converges, when $R \to \infty$ we have $\sum_{r=0}^{R}(r!)^{-1}\Delta^r < C < \infty$ for some small Δ and some fixed $C \in \mathbb{R}$, whence $A_{t_j}(t; \Delta, R, \mathbf{\Phi})$, and thus $A_{t_j}^2(t; \Delta, R, \mathbf{\Phi})$ is bounded. Then, by Assumptions 1(b) and 2(a), the term $V(t; \Delta, R, \mathbf{\Phi})$—and so is the variance $\hat{X}_\omega(t + \Delta) - X_\omega(t + \Delta)$—is of order J^{-1}. As a result, a $\sqrt{J}$-asymptotic normality can be achieved.

3. Simulation

We now explain the numerical procedure of the proposed methods through a simulation study and compare its performance to that of a parametric approach under both correct specification and misspecification, thereby revealing FDA's superiority in tracking the f.d. data structures of stochastic processes. The simulation study was conducted using the software R (version 4.0.5) and MATLAB (version R2023a), and the computer code is available in the Supplementary Materials.

3.1. The Data-Generating Process

The data-generating process is based on the strong GARCH (1,1), where $y_k = (S_k - S_{k-1})/S_{k-1}$, $k = 1, 2, \ldots$ is the arithmetic return on a financial asset with the price S_k. Let h be the time window, and recall that a strong GARCH (1,1) process is represented by $y_k = \mu + \epsilon_k$

with $\epsilon_k \sim \mathcal{N}(0, V_k)$ given the σ-algebra generated by ϵ_{k-1} and $V_k = \omega_h + \xi_h \epsilon_{k-1}^2 + \gamma_h V_{k-1}$ for the f.d. parameters $\omega_h, \xi_h, \gamma_h > 0$ and $\xi_h + \gamma_h < 1$ [54,55].

As h shrinks, (S_{kh}, V_{kh}) weakly converges to its continuous-time limit (S_t, V_t) [6]. This is a simple but effective example, since given different assumptions as to the f.d. parameters' convergence rates (i.e., the convergence of ξ_h can have rate $\sqrt{h}$ or rate h), (S_t, V_t) has been shown to be a diffusion process solution to either a stochastic volatility (SV) model [13,56]

$$\begin{bmatrix} ds_t \\ dv_t \end{bmatrix} = \begin{bmatrix} a \\ (\beta - \frac{1}{2}\sigma^2) + \alpha \exp(-v_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{1-\rho^2}\exp(\frac{v_t}{2}) & \rho\exp(\frac{v_t}{2}) \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} dW_t^1 \\ dW_t^2 \end{bmatrix}, \tag{7}$$

where $s_t := \log(S_t)$, $v_t := \log(V_t)$, and W_t^1 and W_t^2 are independent standard Brownian motions, or a deterministic volatility (DV) model [6]

$$\begin{bmatrix} ds_t \\ dv_t \end{bmatrix} = \begin{bmatrix} a \\ \beta + \alpha \exp(-v_t) \end{bmatrix} dt + \begin{bmatrix} \exp(\frac{v_t}{2}) \\ 0 \end{bmatrix} dW_t^1. \tag{8}$$

Indeed, despite the fact that the ARCH-type diffusion models under both SV and DV have been frequently applied in the literature to estimate continuous-time processes (e.g., [8,9,22,27,57,58]), there has been considerable debate as to the choice of the convergence rate assumptions, which makes parametric analysis such as MLE tailored for either limit questionable (e.g., [14,59]). More generally, if the f.d. data structures have different limiting processes under different convergence conditions, it would be impractical for parametric continuous-time generalizations to exhaust all possible limits to choose the correct one.

To demonstrate how the FDA-based method can bypass this issue, we generate pseudo-continuous return and volatility processes, as shown in Equations (7) and (8), over the interval of $[0,1]$, resembling the "continuous-time" processes of five equally spaced trading points per day for one year of 252 trading days (i.e., 5 time points per day × 252 days per year × 1 year = 1260 data points in total). The true values of the parameters in Tables 1 and 2 are used to generate $N = 1000$ pseudo-continuous log-return and log-volatility trajectories from the SV limit, denoted respectively by $s_S(t)$ and $v_S(t)$, and those trajectories from the DV limit, denoted respectively by $s_D(t)$ and $v_D(t)$. It should be noted that the volatility is not always observed in practice, and discussions addressing parametric analysis with such unobservability have been provided in the literature (e.g., [22,27,60]). However, the FDA-based method can handle the return and volatility processes separately; hence, in what follows, we assume that volatility is accessible without harming the analysis of the return. Should we be interested in the unobserved volatility, the same FDA-based method could be applied to suitable proxies. We take the pseudo-continuous-time processes as the observations, denoted as $s_S(t_j)$, $v_S(t_j)$, $s_D(t_j)$, and $v_D(t_j)$ for $j = 1, \ldots, J$, to numerically mimic the scenario where the sampling time window becomes arbitrarily small, and the estimation is performed based on one-month rolling windows (i.e., $J = 1260/12 = 105$) and eight-month rolling windows (i.e., $J = 1260/12 * 8 = 840$), respectively. Cases with an evenly spaced "daily" sampling density are also included to demonstrate the performance of the FDA-based method under different sample sizes and frequencies.

3.2. Fitting and Forecasting with FDA

Now we present the procedures for fitting and forecasting with FDA in steps (a) to (c) below, and the forecast evaluation is discussed in steps (d) and (e). Note that only the implementations in terms of $s_S(t)$ and $s_D(t)$ are explained here, and similar procedures can be applied to the $v_S(t)$ and $v_D(t)$ processes.

(a)　Provide a candidate pool for the vector (λ, R) of the tuning parameter and the Taylor expansion order. For each of the 1000 replications, take the length $J - 10$ rolling windows (except for the one-month daily sample, where we take the length $J - 3$ rolling windows) from the first J observed data points and estimate the underlying processes $s_S(t)$

and $s_D(t)$ as in Equations (4) and (5) using the B-spline basis of order $R + 3$ and the number of basis functions $\min\{2(R + 2) + 1, J + R + 3\}$. We then obtain $\tilde{s}_S(t)$ and $\tilde{s}_D(t)$, respectively, for each rolling window with each pair of (λ, R) candidates.

(b)　For the first J observed data points from each of the 1000 replicates, compute the forecasting values $\hat{s}_S(t + \Delta)$ and $\hat{s}_D(t + \Delta)$ for each of the length $J - 10$ rolling windows with each pair of (λ, R) candidates using Equation (6). The (λ, R) pair that minimizes the RMSFE over the ten rolling windows across all 1000 replications is selected and denoted by $(\hat{\lambda}, \hat{R})$ to be used for later fitting and forecasting.

(c)　For each of the 1000 replications, perform the fitting and forecasting on the length J rolling windows (either one-month or eight-month) with the given observation frequency using the selected parameters $(\hat{\lambda}, \hat{R})$, the basis order of $R + 3$, and the number of basis functions $\min\{2(R + 2) + 1, J + R + 3\}$, yielding the rolling-window forecasting values $\hat{s}_S(t + \Delta)$ and $\hat{s}_D(t + \Delta)$. It is worth noting that the forecasting step of this method can be flexible, as Δ is technically a continuous quantity. In this simulation study, we only consider the forecasting step that is the same as the sampling time window, which we later refer to as the "one-step-ahead forecast", while in general, for example, one can make a one-hour-ahead forecast with daily data or a one-day-ahead forecast with hourly data.

(d)　Implement the Kolmogorov-Smirnov (*K-S*) test on the pairs "$\hat{s}_S(t + \Delta)$ and $S_S(t + \Delta)$", "$\hat{s}_S(t + \Delta)$ and $S_D(t + \Delta)$", "$\hat{s}_D(t + \Delta)$ and $S_S(t + \Delta)$", and "$\hat{s}_D(t + \Delta)$ and $S_D(t + \Delta)$" for all time points t to check whether the FDA-based predictors can correctly distinguish the true underlying processes in terms of distributions.

(e)　Calculate the RMSFE according to [61,62] between the pairs "$\hat{s}_S(t + \Delta)$ and $s_S(t + \Delta)$", "$\hat{s}_S(t + \Delta)$ and $s_D(t + \Delta)$", "$\hat{s}_D(t + \Delta)$ and $s_S(t + \Delta)$", and "$\hat{s}_D(t + \Delta)$ and $s_D(t + \Delta)$" for the out-of-sample performance evaluation.

In practice, the relevant parameters can be selected more exhaustively for each forecast and each replication as needed; however, developing the asymptotic properties of such data-driven methods is not the focus of the current paper, so we simplify the search in step (a) by having only three possible Rs and three possible λs for each search, forming nine pairs of (λ, R) candidates, and the resulting $(\hat{\lambda}, \hat{R})$ is used for all the replicates in all the later forecasts. Also, the upper bound $J + R + 3$ of the number of basis functions Q is used to simplify the search by first fixing Q at the sample size J while extending the fitting function beyond the fitting domain by $R + 3$ equally spaced basis functions to capture the smoothness on $[1 - \Delta, 1]$ and to avoid the boundary noise induced by the last $R + 3$ (i.e., the order of the basis in our setting) basis functions.

3.3. Comparison with Parametric Methods

We adopted MLE as a benchmark parametric method for a performance comparison with our FDA approach as it is a commonly used approach for GARCH-like models (see e.g., [22,63]). It is imperative to underscore that within this context, MLE was employed as the benchmark parametric approach for the estimation of a sequence of discrete observations drawing from an unknown underlying diffusion process. While alternative parametric methodologies, like the general method of moments (GMM), are available for application, it is noteworthy that they all necessitate certain parametric assumptions about the data distribution, which could potentially result in erroneous estimation results and forecasts when dealing with f.d. data structures under infill asymptotics. Based on the stochastic differential equations of the two limits in Equations (7) and (8), the likelihood functions were obtained utilizing the (joint) normality of dW_t^1 and dW_t^2, and the differentials ds and dv were approximated by the corresponding first difference of the discrete observations. In favor of the parametric methods, we took the eight-month pseudo-continuous rolling windows as the observed series, and the comparison between the two approaches was conducted in terms of RMSFEs.

Thus, based on Equations (7) and (8), the forecasts based on conditional expectation are such that

$$\begin{cases} \hat{s}_{t+\Delta} := \hat{\mathbb{E}}[s_{t+\Delta}|s_t] = s_t + \hat{a}\Delta, \\ \hat{v}_{t+\Delta} := \hat{\mathbb{E}}[v_{t+\Delta}|v_t] = v_t + \left[\left(\hat{\beta} - \frac{1}{2}\hat{\sigma}^2\right) + \hat{\alpha}\exp(-v_t)\right]\Delta \end{cases} \tag{9}$$

for the SV limit and

$$\begin{cases} \hat{s}_{t+\Delta} := \hat{\mathbb{E}}[s_{t+\Delta}|s_t] = s_t + \hat{a}\Delta, \\ \hat{v}_{t+\Delta} := \hat{\mathbb{E}}[v_{t+\Delta}|v_t] = v_t + \left[\hat{\beta} + \hat{\alpha}\exp(-v_t)\right]\Delta \end{cases} \tag{10}$$

for the DV limit, where $\hat{a}, \hat{\beta}, \hat{\alpha}$, and $\hat{\sigma}$ are the estimated parameters from MLE, and Δ is the forecasting step as previously defined.

3.4. Results

Figures 1 and 2 present the forecasting results using the FDA-based method, where the underlying processes are indicated by gray lines, the one-step-ahead rolling forecasts are indicated by black lines, and the selected order of Taylor expansion is noted in the labels of the plots. The figures reveal that the proposed functional approach accurately traced the movements of both the underlying return and volatility processes.
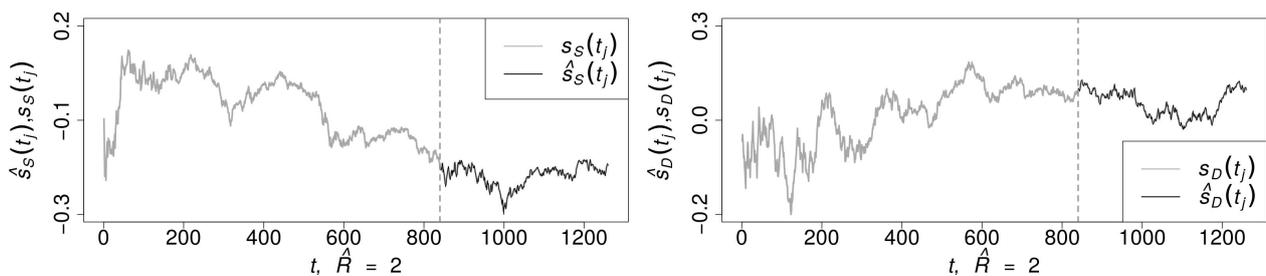
Figure 1. Functional data prediction, eight-month rolling window, continuous returns.
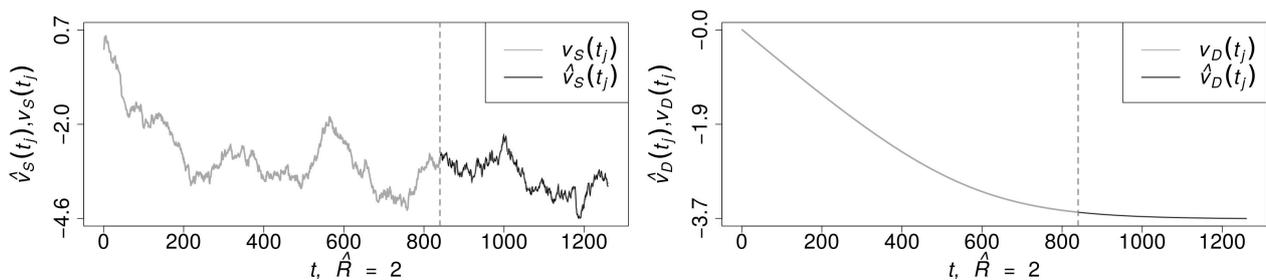
Figure 2. Functional data prediction, eight-month rolling window, continuous volatility.

Then, we used the *K-S* test to compare the predicted and the underlying processes. The *p*-values of the test on the null of distribution equality over the entire time domain are presented in Figures 3 and 4, where the dashed lines indicate the 5% significance level. As depicted in the two plots from the first column of Figure 3, the true return and volatility underlying processes generated from the SV and DV models exhibited significantly disparate distributions at almost every fixed observation time t_j. The second and third columns of Figure 3 suggest that the null hypothesis of distribution equality was rejected when the underlying processes showed significant differences in the cross-comparison, while the FDA forecast shared the same distribution as its true underlying process at almost every t_j, as evident in Figure 4. The results imply that the one-step-ahead rolling forecast using the FDA-based method could preserve the distributions of the true underlying processes pointwisely and could distinguish the true underlying processes from a falsely assumed continuous-time limit in terms of their distributions.
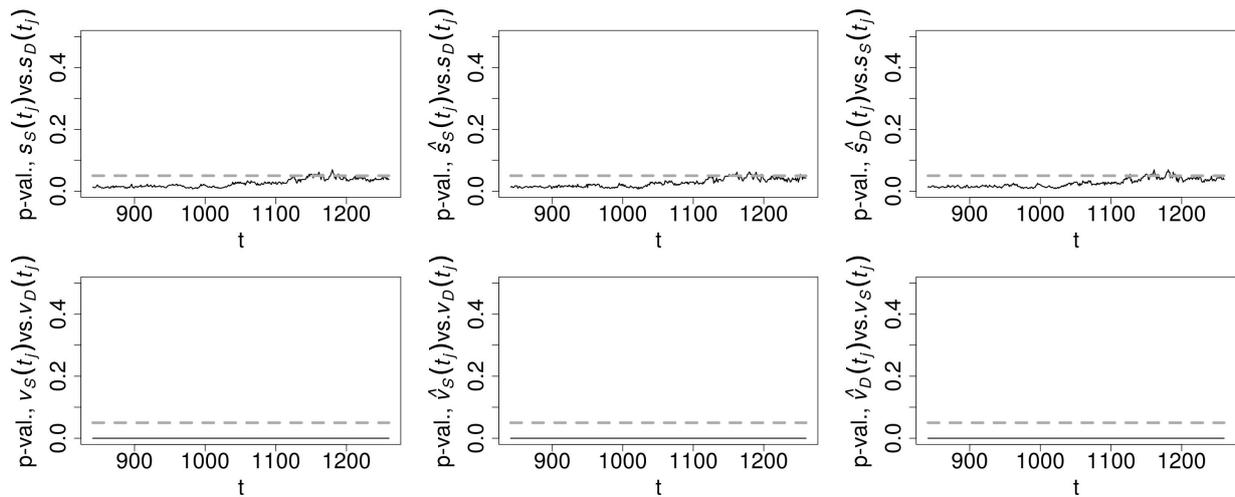
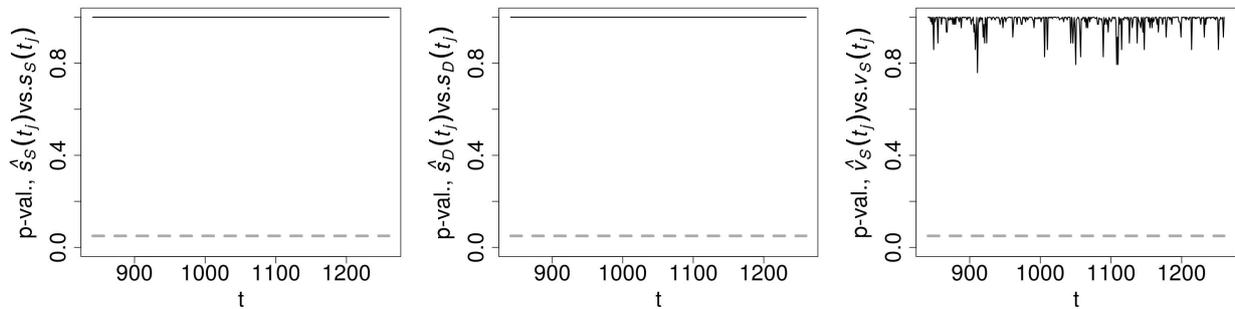**Figure 3.** *K-S* test, eight-month rolling window, cross-comparison.

Figure 4. *K-S* test, eight-month rolling window, comparison with the true underlying process.

The proposed method was further applied to a lower observation frequency and a smaller sample size, where it exhibited decent performance with short and sparse observed series; the corresponding results are illustrated in Appendix B.1. A summary of comparisons across different sample sizes and frequencies in terms of RMSFE is presented in Figure 5, where "1 Md" denotes the one-month rolling window with daily observations, "8 Md" denotes the eight-month rolling window with daily observations, "1 Mps" denotes the one-month rolling window with pseudo-continuous observations, and "8 Mps" denotes the eight-month rolling window with pseudo-continuous observations. The small yet discernible differences in the RMSFEs among the different domain spans and different sampling densities indicates that the FDA-based method benefited from both the infill sample growth and the increasing-domain sample growth, though more from the former than from the latter.

(a) $s_S(t)$ **(b)** $v_S(t)$ **(c)** $s_D(t)$ **(d)** $v_D(t)$
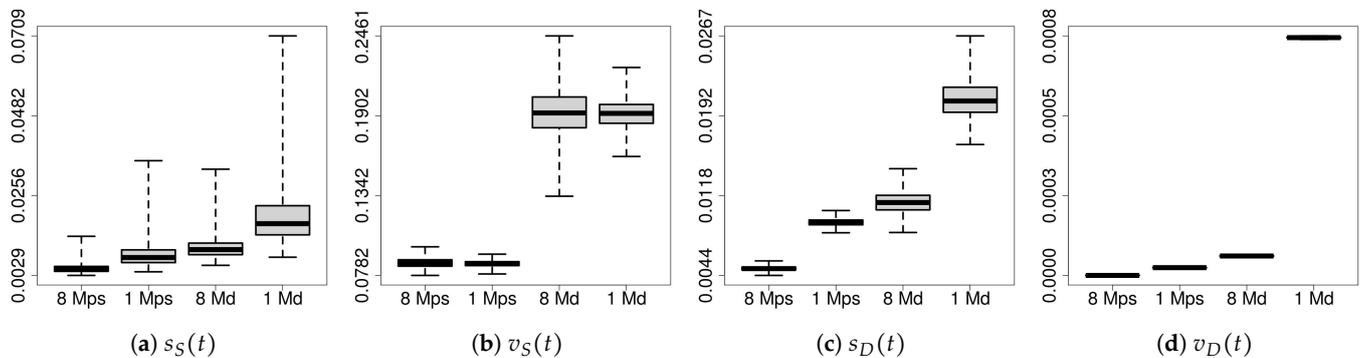
Figure 5. RMSFE distributions for different underlying processes.

The results for the first rolling window of the MLE are shown in Tables 1 and 2 to provide a snapshot of the performance of MLE, while the results for all 420 rolling-window

estimations are summarized by boxplots in Appendix B.2. Unsurprisingly, with the misspecified models, the estimates of the affected parameters mostly departed from the true values with statistical significance, while MLE showed good asymptotic performance under the correct specification. As noted in Tables 1 and 2, the rejection rates for some parameters slightly exceeded the reasonable range of errors induced by a binomial distribution with N = 1000 trials and a probability of success at 5% even when the model was correctly specified, which could be attributed to the insufficient number of observations. Nevertheless, in general, methods that can circumvent such modeling discrepancy will be useful to avoid misleading estimation results in continuous-time analysis.

Table 1. Estimating SV parameters with MLE.

Parameter	a	α	β	σ	ρ
True value	0.1	0.2	−8.5	2.7	−0.8
SV process fitted by an SV model (correct specification)					
Estimate	0.004	0.289	−9.546	2.699	−0.800
Bias	−0.096	0.089	−1.046	−0.001	0.000
Rejection rate	0.066	0.080	0.047	0.056	0.062
SV process fitted by a DV model (misspecification)					
Estimate	0.101	232.294	−259.989	–	–
Bias	0.001	232.094	−251.489	–	–
Rejection rate	0.060	0.789	0.902	–	–

Note: a 1000-time simulation allows for a "$\pm\sqrt{0.95 * 0.05/1000} * 2 = 0.014$" error on the 5% rejection rate under correct specification.

Table 2. Estimating DV parameters with MLE.

Parameter	a	α	β	σ	ρ
True value	0.1	0.2	−8.5	0	0
DV process fitted by a DV model (correct specification)					
Estimate	0.107	0.200	−8.531	–	–
Bias	0.007	0.000	−0.031	–	–
Rejection rate	0.048	0.051	0.053	–	–
DV process fitted by an SV model (misspecification)					
Estimate	1.033	0.200	−8.500	0.000	0.154
Bias	0.933	0.000	0.000	0.000	0.154
Rejection rate	0.346	1.000	1.000	1.000	0.877

Note: a 1000-time simulation allows for a "$\pm\sqrt{0.95 * 0.05/1000} * 2 = 0.014$" error on the 5% rejection rate under correct specification.

Finally, to compare the performance of the FDA and MLE methods, we applied the MLE rolling-window estimation to perform one-step-ahead out-of-sample forecasting and compared the distributions of the relative RMSFEs (rRMSFEs) calculated as the ratio of the RMSFEs between FDA and MLE—a ratio below one signified that the FDA method provided a more precise forecast than the MLE method. The results for cases in which the MLE forecast was affected by the misspecification are shown in Figure 6, and those for cases in which the MLE forecast was unaffected are summarized in Figure A13 in Appendix B.2.

First, note that misspecifying an SV process and estimating it with a DV model did not cause significant bias in the estimation of the parameter a, which is the only parameter needed for the forecast of the return process, according to Equations (9) and (10). Thus, the estimation of s_S was not significantly affected by misspecifying it as s_D, and the distribution of the rRMSFEs is summarized in Figure A13. However, the estimations for the parameters α and β were affected by such a misspecification, which in turn caused inaccuracy and instability in the forecast of the volatility process, as shown in the first plot of Figure 6. Specifically, 47.9% of the rRMSFEs were lower than one, and around 44% of the rRMSFEs were lower than 0.01, while when MLE performed better, the rRMSFEs of the two methods

were still within a comparable range. Hence, the FDA-based method appeared to be more stable and generally reliable. On the contrary, when a DV process was misspecified and estimated by an SV model, the estimation of all the parameters was affected statistically, but for the parameters α, β, and σ, the biases were negligibly small. Hence, according to Equations (9) and (10) with $\hat{\sigma} \approx 0$, the forecast for the volatility process would not be substantially affected, as summarized in Figure A13. For the forecast of the return process that relied on $\hat{a}$, the estimates significantly deviated from the true value. However, since the forecast error was determined by the re-scaled estimation error $(\hat{a} - a)\delta$ and the noise, the sizes of which were of the same order numerically, the final forecast error was not notably influenced by the error in $\hat{a}$, and the forecast exhibited a similar performance to the FDA-based method, as shown in the second plot of Figure 6. However, in practice, when the step Δ became larger, the error in $\hat{a}$ began to show a further impact on the forecast. Finally, the FDA-based method could still outperform the parametric method even in a correctly specified case, as shown in the last plot of Figure 6.
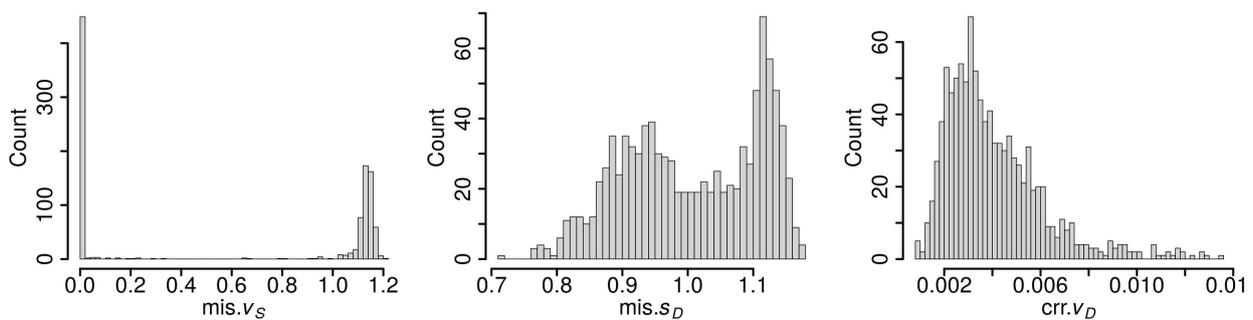
Figure 6. rRMSFEs of FDA vs. MLE.

Above all, the FDA-based method showed general consistency and robustness against modeling discrepancy, while the parametric methods could demonstrate unreliable performance due to misspecification, and even when the parametric method excelled under correct specification, the forecast errors of the two methods still lay within a comparable range, as shown in Figures 6 and A13. When DV was misspecified as SV, the MLE forecast for the volatility process surpassed FDA the furthest in terms of the RMSFE. However, for the forecast of this non-random function, both methods performed well, as shown by the RMSFEs in Figure 5. Hence, the proposed FDA-based method is recommended in the presence modeling discrepancy.

4. Empirical Study

This section illustrates the application of our FDA-based method in stock price prediction, through which we obtained the closing prices of the S&P 500 Index from the Bloomberg terminal. Predicting volatility entails similar processes to forecasting future price movements, for which historical volatility indices like the VIX could be incorporated into our methodology. In accordance with the procedure utilized in our prior simulation study, we proceeded to assess the predictive efficacy of the FDA method utilizing observations spanning a duration of one year. Specifically, one-month and eight-month rolling windows of hourly data (every hour from 9:30 to 15:30) and daily data (every day at 12:30) were employed, and the fitting and forecasting with parameter selection followed the procedures described in steps (a) to (c) from Section 3.

Figures 7 and 8 depict the functional predictions across various sampling frequencies and rolling windows. Figures 9 and 10 present the distribution of the ratio of absolute forecast errors to log-returns in the four corresponding scenarios. The percentage of the forecast errors was consistently small, suggesting a close alignment between the forecast and the actual log-returns, which underscored the robustness and reliability of the proposed FDA-based method.
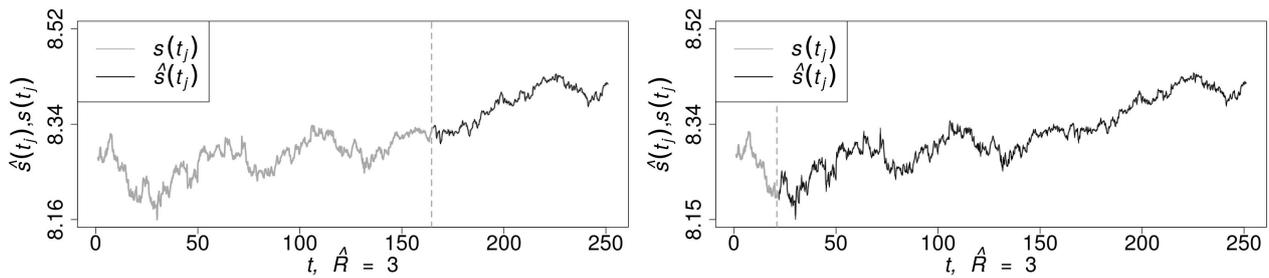
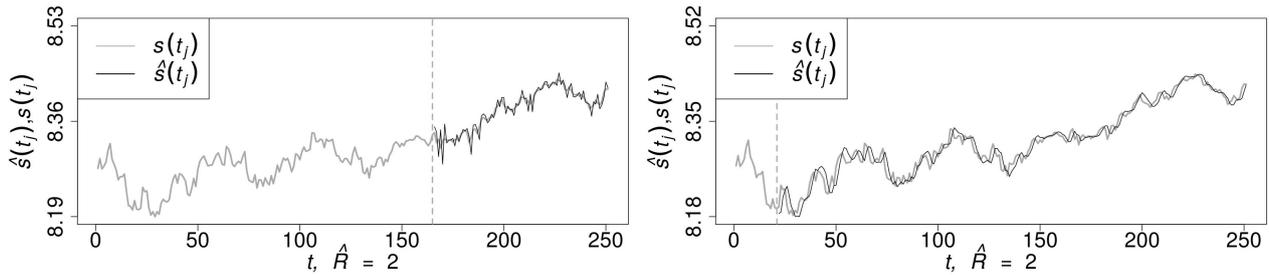**Figure 7.** Functional data prediction, hourly S&P 500 log-returns.

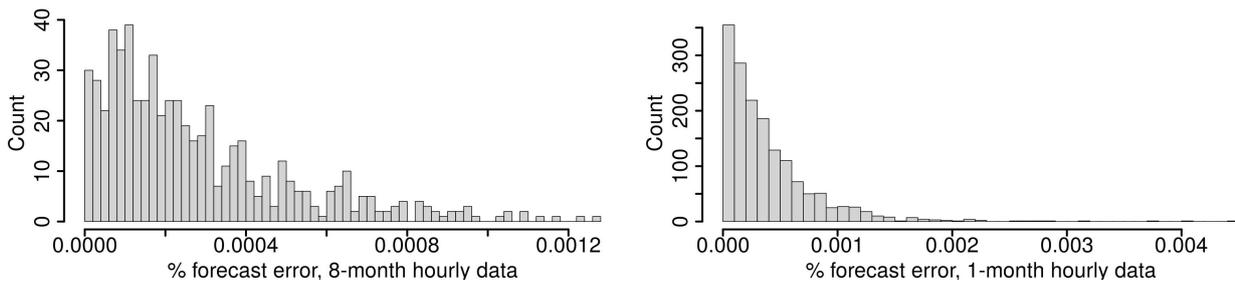**Figure 8.** Functional data prediction, daily S&P 500 log-returns.

Figure 9. Functional data prediction, hourly S&P 500 log-returns, distribution of % forecast error.
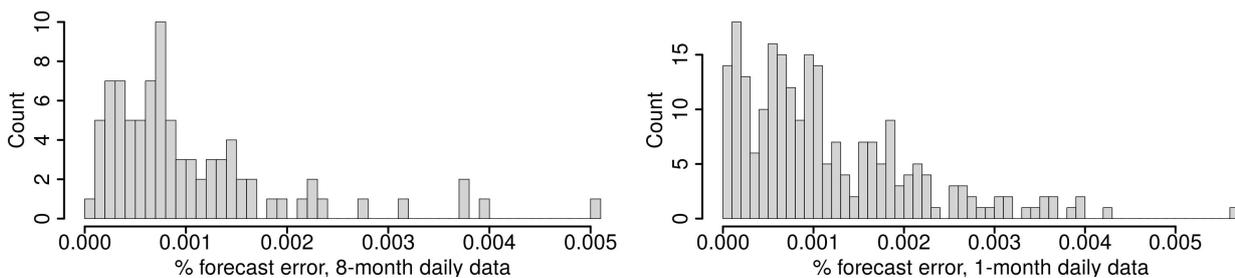

Figure 10. Functional data prediction, daily S&P 500 log-returns, distribution of % forecast error.

5. Conclusions

In continuous-time modeling, parametric methods fail to provide reliable analysis when there is discrepancy due to the existence of multiple limits. This paper adopted FDA to uncover the true continuous-time underlying processes subject to f.d. data structures under infill asymptotics and suggested a forecasting method by integrating FDA with Taylor series expansion, also exploring an application of FDA in out-of-sample prediction.

Our theorems demonstrate that the FDA-based method only requires the continuity of the underlying sample path and the low correlation of the observation errors, and with proper basis expansions, enlarging samples with an increasing sampling density ensure that the functional estimator converges to a unique and well-defined limit. The simulation analysis showed that the FDA-based method is capable of distinguishing processes with different continuous-time limits in out-of-sample prediction. Furthermore, although the parametric method dominated under correct specification, the forecast errors of the two

methods were still within a comparable range, while the FDA-based method showed a generally consistent and much more robust performance against possible misspecification when the parametric method was subject to modeling discrepancy, which makes the functional method a preferable tool for fitting and forecasting when there is uncertainty in modeling the underlying process. Further, we validated the practical applicability of the proposed method in predicting future S&P 500 prices. Our findings demonstrated that the functional approach yielded dependable predictions even in cases where the underlying process was unknown. It is important to note that, as shown in the empirical study, the suggested FDA-based method is not limited to any particular continuous-time model; instead, it is able to uncover the underlying process of many discretely observed stochastic processes, including equity returns.

Lastly, the proposed FDA-based method shows limitations in certain circumstances. First, as the most natural way to obtain the optimal values of the parameters determining the functional estimators and forecasts, e.g., the number and the order of the basis function, the tuning parameter λ, and the order of the Taylor expansion, is by data-driven algorithms, the search within such an optimization process can induce computationally expensive tasks. Additionally, in the presence of a continuous but volatile underlying process, our approach also necessitates increased computational resources for determining the high-order derivatives of the fitted functions. Moreover, the nature of the method implies that, with everything else fixed, it generally works more comfortably and easily for smoother processes than volatile ones and is more suited for capturing long-term patterns than short-term shocks. Therefore, the performance of the method may heavily rely on high-frequency data while fitting and forecasting highly volatile processes. For future research, more efficient methods to obtain the optimal values of the parameters are to be developed, and the effects of the estimated parameters on the overall performance of the fitting and forecasting are to be explored. Furthermore, the method's performance with volatile underlying processes and/or sparse observations is to be further studied and improved.

Abbreviations

The following abbreviations are used in this manuscript:

f.d.	frequency-dependent
GARCH	generalized autoregressive conditional heteroskedasticity
MLE	maximum likelihood estimation
FDA	functional data analysis
SV	stochastic volatility
DV	deterministic volatility
MSFE	mean squared forecast error
RMSFEs	relative MSFEs

Appendix A

Additional notations and lemmas (proofs of the lemmas are given further below).

Let $N_0(\zeta)$ and $N_{1-\Delta}(\zeta)$ denote the ζ-neighborhood of $t = 0$ and $t = 1 - \Delta$, respectively, for some small $\zeta > 0$. Then, we define $\mathcal{T}_{\Delta,\zeta} := [0, 1 - \Delta] \cup N_0(\zeta) \cup N_{1-\Delta}(\zeta)$. Also, let $\tilde{X}_{\omega,J}$ denote the fitted function for X_ω under the sample size J.

Lemma A1. *Consider a continuous sample path $X_\omega(t)$ on $[0, 1]$ where for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that $|t_1 - t_2| \leq \delta(\varepsilon)$ implies $|X_\omega(t_1) - X_\omega(t_2)| \leq \varepsilon/2$. Suppose Assumption 1 holds. Then, for $B_K(t, X_\omega)$, as in Equation (2), and $K \geq 4 \sup_{t \in [0,1]} |X_\omega(t)|/(\delta^2(\varepsilon)\varepsilon)$, we have the following:*

(a)   *For any $t \in \mathcal{T}_{\Delta,\zeta}$ and $\omega \in \Omega$, $\tilde{X}_{\omega,J}(t) - B_K(t, X_\omega) \xrightarrow{p} 0$ as $J \to \infty$.*

(b)   *For every $\varrho, \eta > 0$ there exists a $\zeta > 0$ such that $\{\tilde{X}_{\omega,J}\}$ is asymptotically stochastically equicontinuous on $\mathcal{T}_{\Delta,\zeta}$ in that $\limsup_{J \to \infty} \mathbb{P}\{\sup_{t_1,t_2 \in \mathcal{T}_{\Delta,\zeta}, |t_1-t_2|<\zeta} |\tilde{X}_{\omega,J}(t_1) - \tilde{X}_{\omega,J}(t_2)| > \varrho\} < \eta$.*

(c)   *For any $t \in \mathcal{T}_{\Delta,\zeta}$, $\omega \in \Omega$, and $r = 1, \ldots, K$, $\tilde{X}_{\omega,J_1}^{(r)}(t) - \tilde{X}_{\omega,J_2}^{(r)}(t) \xrightarrow{p} 0$ for any $J_1, J_2 \to \infty$.*

(d)   *For every $\varrho, \eta > 0$ there exists a $\zeta > 0$ such that for all $r = 1, \ldots, K$,*
$\limsup_{J \to \infty} \mathbb{P}\{\sup_{t_1,t_2 \in N_0(\zeta)} |\tilde{X}_{\omega,J}^{(r)}(t_1) - \tilde{X}_{\omega,J}^{(r)}(t_2)| > \varepsilon\} < \eta$, *and*
$\limsup_{J \to \infty} \mathbb{P}\{\sup_{t_1,t_2 \in N_{1-\Delta}(\zeta)} |\tilde{X}_{\omega,J}^{(r)}(t_1) - \tilde{X}_{\omega,J}^{(r)}(t_2)| > \varepsilon\} < \eta$.

Lemma A2. *Suppose Assumption 1 and Lemma A1 hold. Then, for $r = 1, \ldots, K$,*

(a) $\text{plim}_{J \to \infty} \lim_{t \to 0^+} \tilde{X}_{\omega,J}^{(r)}(t) = \lim_{t \to 0^+} \text{plim}_{J \to \infty} \tilde{X}_{\omega,J}^{(r)}(t)$ *and*
$\text{plim}_{J \to \infty} \lim_{t \to (1-\Delta)^-} \tilde{X}_{\omega,J}^{(r)}(t) = \lim_{t \to (1-\Delta)^-} \text{plim}_{J \to \infty} \tilde{X}_{\omega,J}^{(r)}(t)$.

(b) $\sup_{t \in [0,1-\Delta]} |\tilde{X}_{\omega,J}(t) - B_K(t, X_\omega)| \xrightarrow{p} 0$ *as $J \to \infty$.*

(c) $\tilde{X}_{\omega,J}^{(r)}(0) - B_K^{(r)}(0, X_\omega) \xrightarrow{p} 0$ *and* $\tilde{X}_{\omega,J}^{(r)}(1 - \Delta) - B_K^{(r)}(1 - \Delta, X_\omega) \xrightarrow{p} 0$ *as $J \to \infty$.*

Lemma A3. *Let f be a function that maps a squared matrix to a real value; then, for full-rank Q-by-Q squared matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{Z}$,*

$$f(\boldsymbol{A}) = f(\boldsymbol{B}) + tr\left[\left\{\frac{\partial f(\boldsymbol{Z})}{\partial \boldsymbol{Z}}\right\}^\top (\boldsymbol{A} - \boldsymbol{B})\right],$$

where $\min\{a_{ij}, b_{ij}\} < z_{ij} < \max\{a_{ij}, b_{ij}\}$ for all elements a_{ij}, b_{ij}, and z_{ij} of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{Z}$, respectively, with $i, j = 1, \ldots, Q$.

Appendix A.1. Proof of Theorem 1

Proof of Theorem 1. First, applying integration by parts given any positive integer $H \leq K + 1$, we have that for all $\psi \in \mathcal{F}_K$,

$$\sum_{r=1}^{H} (-1)^{r-1} \tilde{X}_{\omega,J}^{(r-1)}(1 - \Delta) \psi^{(H-r)}(1 - \Delta) = \sum_{r=1}^{H} (-1)^{r-1} \tilde{X}_{\omega,J}^{(r-1)}(0) \psi^{(H-r)}(0) +$$
$$\int_0^{1-\Delta} \tilde{X}_{\omega,J}(t) \psi^{(H)}(t) dt - (-1)^H \int_0^{1-\Delta} \tilde{X}_{\omega,J}^{(H)}(t) \psi(t) dt,$$

$$\sum_{r=1}^{H} (-1)^{r-1} B_K^{(r-1)}(1 - \Delta, X_\omega) \psi^{(H-r)}(1 - \Delta) = \sum_{r=1}^{H} (-1)^{r-1} B_K^{(r-1)}(0, X_\omega) \psi^{(H-r)}(0) +$$
$$\int_0^{1-\Delta} B_K(t, X_\omega) \psi^{(H)}(t) dt - (-1)^H \int_0^{1-\Delta} B_K^{(H)}(t, X_\omega) \psi(t) dt,$$

which implies that

$$\left| \int_0^{1-\Delta} \tilde{X}_{\omega,J}^{(H)}(t)\psi(t)dt - \int_0^{1-\Delta} B_K^{(H)}(t, X_\omega)\psi(t)dt \right|$$

$$\leq \left| \sum_{r=1}^{H} (-1)^{r-1} \left\{ \tilde{X}_{\omega,J}^{(r-1)}(1-\Delta) - B_K^{(r-1)}(1-\Delta, X_\omega) \right\} \psi^{(H-r)}(1-\Delta) \right|$$

$$+ \left| \sum_{r=1}^{H} (-1)^{r-1} \left\{ \tilde{X}_{\omega,J}^{(r-1)}(0) - B_K^{(r-1)}(0, X_\omega) \right\} \psi^{(H-r)}(0) \right|$$

$$+ \left| \int_0^{1-\Delta} \tilde{X}_{\omega,J}(t)\psi^{(H)}(t)dt - \int_0^{1-\Delta} B_K(t, X_\omega)\psi^{(H)}(t)dt \right|.$$

Lemma A2(c) indicates that both $\left| \sum_{r=1}^{H} (-1)^{r-1} \left\{ \tilde{X}_{\omega,J}^{(r-1)}(1-\Delta) - B_K^{(r-1)}(1-\Delta, X_\omega) \right\} \psi^{(H-r)} \right.$
$\left. (1-\Delta) \right|$ and $\left| \sum_{r=1}^{H} (-1)^{r-1} \left\{ \tilde{X}_{\omega,J}^{(r-1)}(0) - B_K^{(r-1)}(0, X_\omega) \right\} \psi^{(H-r)}(0) \right|$ are $o_p(1)$. Meanwhile, Lemma A2(b) implies that $\left| \int_0^{1-\Delta} \tilde{X}_{\omega,J}(t)\psi^{(H)}(t)dt - \int_0^{1-\Delta} B_K(t, X_\omega)\psi^{(H)}(t)dt \right| = o_p(1)$. Hence, given any positive integer $H \leq K+1$,

$$\left| \int_0^{1-\Delta} \tilde{X}_{\omega,J}^{(H)}(t)\psi(t)dt - \int_0^{1-\Delta} B_K^{(H)}(t, X_\omega)\psi(t)dt \right| = \left| \left\langle \tilde{X}_{\omega,J}^{(H)}, \psi \right\rangle - \left\langle B_K^{(H)}(\cdot, X_\omega), \psi \right\rangle \right| = o_p(1), \ \forall \psi \in \mathcal{F}_K,$$

and for $R = K+1$, applying the properties of inner products, one can obtain

$$\left\langle \hat{X}_\omega(\cdot + \Delta), \psi \right\rangle = \left\langle \sum_{r=0}^{R} \frac{1}{r!} (\Delta)^r \tilde{X}_{\omega,J}^{(r)}, \psi \right\rangle = \sum_{r=0}^{R} \frac{1}{r!} (\Delta)^r \left\langle \tilde{X}_{\omega,J}^{(r)}, \psi \right\rangle$$

$$= \sum_{r=0}^{R} \frac{1}{r!} (\Delta)^r \left\langle B_K^{(r)}(\cdot, X_\omega), \psi \right\rangle + o_p(1) = \left\langle \sum_{r=0}^{R} \frac{1}{r!} (\Delta)^r B_K^{(r)}(\cdot, X_\omega), \psi \right\rangle + o_p(1)$$

$$= \left\langle B_K(\cdot + \Delta, X_\omega), \psi \right\rangle + o_p(1).$$

Then, under Lemma 1, the desired results follow. $\square$

Appendix A.2. Proof of Theorem 2

Proof of Theorem 2. To improve readability, we define the following:

$$\mathbf{Y}_\Delta(t) := \sum_{r=0}^{R} \frac{1}{r!} \Delta^r \left[\mathbf{\Phi}^{(r)}(t) \right]^\top, \quad \mathbf{\Omega} := \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) \mathbf{\Phi}^\top(t_j), \quad \mathbf{\Gamma} := \int_0^1 \mathbf{\Phi}^{(2)}(t) \left\{ \mathbf{\Phi}^{(2)}(t) \right\}^\top dt.$$

Given the order of Q relative to J, we impose that $Q < J$ without a loss of generality. Then, applying Equation (3), Lemma 1, and Lemma A3 with $f(\mathbf{M}) = \mathbf{a}^\top \mathbf{M}^{-1} \mathbf{b}$ for square matrix $\mathbf{M}$, vectors $\mathbf{a} := \mathbf{Y}_\Delta(t)$ and $\mathbf{b} := \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) X_{t_j}$, and matrices $\mathbf{A} := \mathbf{\Omega} + \lambda \mathbf{\Gamma}$ and $\mathbf{B} := \mathbf{\Omega}$ indicates that

$$\hat{X}_\omega(t + \Delta) = \mathbf{Y}_\Delta(t) \tilde{\mathbf{C}}_\omega = \mathbf{Y}_\Delta(t) (\mathbf{\Omega} + \lambda \mathbf{\Gamma})^{-1} \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) X_{t_j} = \mathbf{a}^\top \mathbf{A}^{-1} \mathbf{b} = \mathbf{a}^\top \mathbf{B}^{-1} \mathbf{b} + \mathrm{tr}\left[\left\{ \frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}} \right\}^\top (\mathbf{A} - \mathbf{B}) \right]$$

$$= \mathbf{Y}_\Delta(t) \mathbf{\Omega}^{-1} \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) X_{t_j} - \mathrm{tr}\left[\mathbf{Z}^{-1} \mathbf{b} \mathbf{a}^\top \mathbf{Z}^{-1} \lambda \mathbf{\Gamma} \right]$$

$$= \mathbf{Y}_\Delta(t) \mathbf{\Omega}^{-1} \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) \mathbf{\Phi}^\top(t_j) \mathbf{C} + \mathbf{Y}_\Delta(t) \mathbf{\Omega}^{-1} \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) \epsilon_{t_j} + \mathcal{O}\left(Q^2 \lambda \right)$$

$$= \mathbf{Y}_\Delta(t) \mathbf{C} + \mathbf{Y}_\Delta(t) \mathbf{\Omega}^{-1} \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) \epsilon_{t_j} + \mathcal{O}\left(Q^2 \lambda \right),$$

and thus

$$\hat{X}_\omega(t + \Delta) - B_K(t + \Delta, X_\omega) = \mathbf{Y}_\Delta(t) \mathbf{\Omega}^{-1} \frac{1}{J} \sum_{j=1}^{J} \mathbf{\Phi}(t_j) \epsilon_{t_j} + \mathcal{O}\left(Q^2 \lambda \right).$$

By Assumption 2(b) and [64], $\Omega = \frac{1}{J} \sum_{j=1}^{J} \boldsymbol{\Phi}(t_j) \boldsymbol{\Phi}^\top(t_j) = \int_0^1 \boldsymbol{\Phi}(t) \boldsymbol{\Phi}^\top(t) dt + o(J^{-1})$. Then, again by Lemma A3, we have

$$\hat{X}_\omega(t+\Delta) - B_K(t+\Delta, X_\omega) = \frac{1}{J} \sum_{j=1}^{J} \sum_{r=0}^{R} \frac{1}{r!} \Delta^r \left\{ \boldsymbol{\Phi}^{(r)}(t) \right\}^\top \left\{ \int_0^1 \boldsymbol{\Phi}(s) \boldsymbol{\Phi}^\top(s) ds \right\}^{-1} \boldsymbol{\Phi}(t_j) \epsilon_{t_j} + o(J^{-1} Q^2),$$

whence with $A_{t_j}(t; \Delta, R, \boldsymbol{\Phi}) := \sum_{r=0}^{R} \frac{1}{r!} \Delta^r \left\{ \boldsymbol{\Phi}^{(r)}(t) \right\}^\top \left\{ \int_0^1 \boldsymbol{\Phi}(s) \boldsymbol{\Phi}^\top(s) ds \right\}^{-1} \boldsymbol{\Phi}(t_j)$, $\sigma_j^2 := \mathrm{Var}(\epsilon_{t_j})$ and $V(t; \Delta, R, \boldsymbol{\Phi}) := J^{-2} \sum_{j=1}^{J} \sigma_j^2 A_{t_j}^2(t; \Delta, R, \boldsymbol{\Phi})$ for all j, by Assumption 2(a) and the Lyapunov CLT, it follows that

$$[V(t; \Delta, R, \boldsymbol{\Phi})]^{-1/2} \left\{ \hat{X}_\omega(t+\Delta) - B_K(t+\Delta, X_\omega) \right\}$$

$$= \left[\sum_{j=1}^{J} \sigma_j^2 A_{t_j}^2(t; \Delta, R, \boldsymbol{\Phi}) \right]^{-1/2} \sum_{j=1}^{J} A_{t_j}(t; \Delta, R, \boldsymbol{\Phi}) \epsilon_{t_j} + o_p(1)$$

$$\xrightarrow{d} \mathcal{N}(0, 1).$$

Hence, under Lemma 1, the desired results follow with $K, R \to \infty$. □

Appendix A.3. Proof of Lemma 1

Proof of Lemma 1. For any given state $\omega \in \Omega$, the continuous sample path X_ω is bounded on the compact set $[0, 1]$, where we have $M := \sup_{t \in [0,1]} |X_\omega(t)| < \infty$. Then, the result can be justified by the proof of Theorem 5.14 from [51]. □

Appendix A.4. Proof of Lemma A1

Proof of Lemma A1. Lemma A1(a) can be justified by the results from previous studies: for example, under Assumptions 1 to 3 in [53], which state conditions for the choice of basis functions and the distribution of the sampling points.

For part (b), with properly selected basis functions that are continuously differentiable up to a desired order, applying the mean-value theorem indicates the Lipschitz condition such that for all $t_1 < t_2 \in \mathcal{T}_{\Delta, \zeta}$ with $t_2 - t_1 < \zeta$, $\left| \tilde{X}_{\omega, J}(t_1) - \tilde{X}_{\omega, J}(t_2) \right| \leq \sup_{s \in \mathcal{T}_{\Delta, \zeta}} \left| \tilde{X}_{\omega, J}^{(1)}(s) \right| (t_2 - t_1)$, where $\sup_{s \in \mathcal{T}_{\Delta, \zeta}} \left| \tilde{X}_{\omega, J}^{(1)}(s) \right| = \mathcal{O}_p(1)$. Then, based on the fact that $\lim_{\zeta \to 0} \sup_{|t_1 - t_2| < \zeta} |t_1 - t_2| = 0$ for all $t_1, t_2 \in \mathcal{T}_{\Delta, \zeta}$, the asymptotic stochastic equicontinuity of $\{ \tilde{X}_{\omega, J} \}$ follows on $\mathcal{T}_{\Delta, \zeta}$.

For part (c), note that the convergence of the functional estimators is achieved through the convergence of the estimated basis coefficients, and the derivatives of these functional estimators are obtained through the derivatives of the non-stochastic basis functions; hence, the convergence of the higher-order derivatives of the estimated functions can be easily justified by choosing the proper basis functions that are continuously differentiable up to a desired order.

For part (d), similarly to the justification for (b), with properly selected basis functions, one can determine that for all $t_1 < t_2 \in N_0(\zeta)$, $\left| \tilde{X}_{\omega, J}^{(r)}(t_1) - \tilde{X}_{\omega, J}^{(r)}(t_2) \right| \leq \sup_{s \in [t_1, t_2]} \left| \tilde{X}_{\omega, J}^{(r+1)}(s) \right| (t_2 - t_1)$, where $\sup_{s \in [t_1, t_2]} \left| \tilde{X}_{\omega, J}^{(r+1)}(s) \right| = \mathcal{O}_p(1)$ and $\lim_{\zeta \to 0} \sup_{t_1 < t_2 \in N_0(\zeta)} |t_2 - t_1| = 0$. Therefore, the asymptotic stochastic equicontinuity of $\{ \tilde{X}_{\omega, J}^{(r)} \}$ on $N_0(\zeta)$ for $r = 1, \ldots, K$ follows. The same justification holds for $N_{1-\Delta}(\zeta)$. □

Appendix A.5. Proof of Lemma A2

Proof of Lemma A2. To verify Lemma A2(a), note that

$$\left| \lim_{t \to 0} \tilde{X}_{\omega, J}^{(r)}(t) - B_K^{(r)}(0, X_\omega) \right| \leq \left| \lim_{t \to 0} \tilde{X}_{\omega, J}^{(r)}(t) - \tilde{X}_{\omega, J}^{(r)}(t) \right| + \left| \tilde{X}_{\omega, J}^{(r)}(t) - B_K^{(r)}(t, X_\omega) \right| + \left| B_K^{(r)}(t, X_\omega) - B_K^{(r)}(0, X_\omega) \right|.$$

Given any $\varrho > 0$, there exists an $\eta > 0$ such that one can find a t for which there is a $\bar{J}$ such that

$$\mathbb{P}\left\{\max\left\{\left|\lim_{t\to 0}\tilde{X}^{(r)}_{\omega,J}(t) - \tilde{X}^{(r)}_{\omega,J}(t)\right|, \left|\tilde{X}^{(r)}_{\omega,J}(t) - B^{(r)}_K(t, X_\omega)\right|, \left|B^{(r)}_K(t, X_\omega) - B^{(r)}_K(0, X_\omega)\right|\right\} < \varrho\right\} > 1 - \eta, \ \forall J > \bar{J},$$

implying that $\mathbb{P}\left\{\left|\lim_{t\to 0}\tilde{X}^{(r)}_{\omega,J}(t) - B^{(r)}_K(0, X_\omega)\right| < 3\varrho\right\} > 1 - \eta$. Therefore, with Lemma A1(c), we can state that $\text{plim}_{J\to\infty}\lim_{t\to 0}\tilde{X}^{(r)}_{\omega,J}(t) = \lim_{t\to 0}B^{(r)}_K(t, X_\omega) = \lim_{t\to 0}\text{plim}_{J\to\infty}\tilde{X}^{(r)}_{\omega,J}(t)$.

For Lemma A2(b), applying Theorem 21.9 from [65], with $\tilde{X}_{\omega,J}(t) - B_K(t, X_\omega) \xrightarrow{p} 0$ for each $t \in \mathcal{T}_{\Delta,\zeta}$ according to Lemma A1(a), as well as the asymptotic stochastic equicontinuity of $\{\tilde{X}_{\omega,J}(t)\}$ on $t \in \mathcal{T}_{\Delta,\zeta}$ according to Lemma A1(b), the uniform convergence in probability of $\tilde{X}_{\omega,J}$ on $\mathcal{T}_{\Delta,\zeta}$ such that $\sup_{t\in\mathcal{T}_{\Delta,\zeta}}\left|\tilde{X}_{\omega,J}(t) - B_K(t, X_\omega)\right| \xrightarrow{p} 0$ follows. Hence, Lemma A2(b) is verified.

For Lemma A2(c), since the proofs for the two convergences follow the same idea, we only focus on $t \to 0^+$ and omit the proof under $t \to (1-\Delta)^-$. Let $B_\zeta = N_0(\zeta) \cap [0, 1-\Delta]$. Then, Lemma A2(c) can be proved by induction—we show that $\text{plim}_{J\to\infty}\tilde{X}^{(1)}_{\omega,J}(0) = B^{(1)}_K(0, X_\omega)$, and we justify that $\text{plim}_{J\to\infty}\tilde{X}^{(r)}_{\omega,J}(0) = B^{(r)}_K(0, X_\omega)$ implies $\text{plim}_{J\to\infty}\tilde{X}^{(r+1)}_{\omega,J}(0) = B^{(r+1)}_K(0, X_\omega)$ for $r = 1, \ldots, K-1$.

Similarly to the verification for Lemma A2(b) based on pointwise convergence and asymptotic stochastic equicontinuity, by Lemmas A1(c) and (d), as well as Theorem 21.9 from [65], we can show that $\tilde{X}^{(1)}_{\omega,J}$ converges uniformly in probability on B_ζ, such that given any $\varrho > 0$, there exists an $\eta > 0$ for which one can find a $\bar{J}$ such that

$$\mathbb{P}\left\{\sup_{t\in B_\zeta}\left|\tilde{X}^{(1)}_{J_1}(t) - \tilde{X}^{(1)}_{J_2}(t)\right| < \varrho\right\} > 1 - \eta, \ \forall J_1, J_2 > \bar{J}. \tag{A1}$$

Under the same ϱ, η, J_1, and J_2, for all $\tau \neq 0 \in B_\zeta$, applying the mean-value theorem yields

$$\left|\frac{\tilde{X}_{J_1}(\tau) - \tilde{X}_{J_2}(\tau) - \tilde{X}_{J_1}(0) + \tilde{X}_{J_2}(0)}{\tau - 0}\right| \leq \sup_{t\in[0,\tau]\subset B_\zeta}\left|\tilde{X}^{(1)}_{J_1}(t) - \tilde{X}^{(1)}_{J_2}(t)\right| \leq \sup_{t\in B_\zeta}\left|\tilde{X}^{(1)}_{J_1}(t) - \tilde{X}^{(1)}_{J_2}(t)\right|,$$

and with (A1), we have

$$\mathbb{P}\left\{\sup_{t\neq 0\in B_\zeta}\left|\frac{\tilde{X}_{J_1}(t) - \tilde{X}_{J_2}(t) - \tilde{X}_{J_1}(0) + \tilde{X}_{J_2}(0)}{t}\right| < \varrho\right\} > 1 - \eta. \tag{A2}$$

We define the following two functions for $t \neq 0 \in B_\zeta$:

$$g_J(t) = \frac{\tilde{X}_{\omega,J}(t) - \tilde{X}_{\omega,J}(0)}{t} \quad \text{and} \quad g(t) = \frac{B_K(t, X_\omega) - B_K(0, X_\omega)}{t};$$

then, (A2) implies that g_J converges uniformly in probability on $B_\zeta \backslash \{0\}$. Since $\tilde{X}_{\omega,J}$ converges uniformly to $B_K(\cdot, X_\omega)$ in probability on B_ζ, it follows that

$$\text{plim}_{J\to\infty}g_J(t) = g(t), \quad \forall t \neq 0 \in B_\zeta. \tag{A3}$$

Meanwhile, given the differentiability of $\tilde{X}_{\omega,J}(t)$ and $B_K(t, X_\omega)$, we have

$$\lim_{t\to 0}g_J(t) = \tilde{X}^{(1)}_{\omega,J}(0) \quad \text{and} \quad \lim_{t\to 0}g(t) = B^{(1)}_K(0, X_\omega). \tag{A4}$$

Then, applying Lemma A2(a) on (A3) and (A4) indicates that $\text{plim}_{J\to\infty}\tilde{X}^{(1)}_{\omega,J}(0) = B^{(1)}_K(0, X_\omega)$.

Now suppose for a given r, where $r = 1, \ldots, K-1$, we have $\text{plim}_{J\to\infty}\tilde{X}^{(r)}_{\omega,J}(0) = B^{(r)}_K(0, X_\omega)$. Then, Lemmas A1(c) and (d) as well as Theorem 21.9 from [65] imply that

$\tilde{X}^{(r+1)}_{\omega,J}$ converges uniformly in probability on B_ζ, such that given any $\varrho > 0$, there exists an $\eta > 0$ for which one can find a $\bar{J}$ such that $\mathbb{P}\left\{\sup_{t \in B_\zeta} \left| \tilde{X}^{(r+1)}_{J_1}(t) - \tilde{X}^{(r+1)}_{J_2}(t) \right| < \varrho\right\} > 1 - \eta$, for all $J_1, J_2 > \bar{J}$. Similarly to the previous proof, one can obtain

$$\mathbb{P}\left\{ \sup_{t \neq 0 \in B_\zeta} \left| \frac{\tilde{X}^{(r)}_{J_1}(t) - \tilde{X}^{(r)}_{J_2}(t) - \tilde{X}^{(r)}_{J_1}(0) + \tilde{X}^{(r)}_{J_2}(0)}{t} \right| < \varrho \right\} > 1 - \eta.$$

Note that the pointwise consistency of $\tilde{X}^{(1)}_{\omega,J}$ on B_ζ can be shown by the same means as for $\text{plim}_{J \to \infty} \tilde{X}^{(1)}_{\omega,J}(0) = B^{(1)}_K(0, X_\omega)$, switching m with any point in the domain. Then, we re-define the following two functions for $t \neq 0 \in B_\zeta$:

$$g_J(t) = \frac{\tilde{X}^{(r)}_{\omega,J}(t) - \tilde{X}^{(r)}_{\omega,J}(0)}{t} \quad \text{and} \quad g(t) = \frac{B^{(r)}_K(t, X_\omega) - B^{(r)}_K(0, X_\omega)}{t}.$$

It is implied by the uniform convergence and the pointwise consistency that $\tilde{X}^{(r)}_{\omega,J}$ converges uniformly to $B^{(r)}_K(\cdot, X_\omega)$ in probability on B_ζ. It follows that

$$\text{plim}_{J \to \infty} g_J(t) = g(t), \ \forall t \neq 0 \in B_\zeta.$$

Meanwhile, given the differentiability of $\tilde{X}^{(r+1)}_{\omega,J}(t)$ and $B^{(r+1)}_K(t, X_\omega)$, we have

$$\lim_{t \to 0} g_J(t) = \tilde{X}^{(r+1)}_{\omega,J}(0) \text{ and } \lim_{t \to 0} g(t) = B^{(r+1)}_K(0, X_\omega).$$

Then, again applying Lemma A2(a) indicates that $\text{plim}_{J \to \infty} \tilde{X}^{(r+1)}_{\omega,J}(0) = B^{(r+1)}_K(0, X_\omega)$. $\square$

Appendix A.6. Proof of Lemma A3

Proof of Lemma A3. First, let $\psi(q) := f(B + q(A - B))$ for $q \in [0,1]$. Then, taking the first-order derivative of $\psi(q)$ with respect to q through the matrix argument of the function f yields

$$\psi^{(1)}(q) = \text{tr}\left[\left\{ \frac{\partial f(B + q(A - B))}{\partial(B + q(A - B))} \right\}^\top \left\{ \frac{\partial(B + q(A - B))}{\partial q} \right\} \right] = \text{tr}\left[\left\{ \frac{\partial f(B + q(A - B))}{\partial(B + q(A - B))} \right\}^\top (A - B) \right].$$

By the mean-value theorem, there exists some $q \in [0,1]$ such that $\psi(1) - \psi(0) = \psi^{(1)}(q)$, which is equivalent to

$$f(A) - f(B) = \text{tr}\left[\left\{ \frac{\partial f(Z)}{\partial Z} \right\}^\top (A - B) \right].$$

$\square$

Appendix B

Appendix B.1. FDA Results

This appendix presents the functional data predictions of return and volatility, the corresponding *K-S* test results with different sample sizes, and the distribution of the comparisons in relative MSFE (RMSFE). In particular, we adjusted our simulation by either altering the observation frequency to daily or changing the rolling window to one month. The prediction results are graphed in Figures A1, A2, A5, A6, A9 and A10. All the plots of the one-step-ahead predictions are consistent with the scenario presented in the main text, where our forecast results could correctly trace the trends of the underlying processes. Further, Figures A4, A8 and A12 show that the predictions shared the same distribution as the true underlying process at (almost) every time point, while Figures A3, A7 and A11 imply that the null hypothesis of distribution equality is rejected where the underlying

processes show significant differences in the cross-comparison. In summary, the *K-S* test results indicate that the FDA-based method appeared to correctly distinguish between the processes with different limits in out-of-sample prediction and various sample sizes.
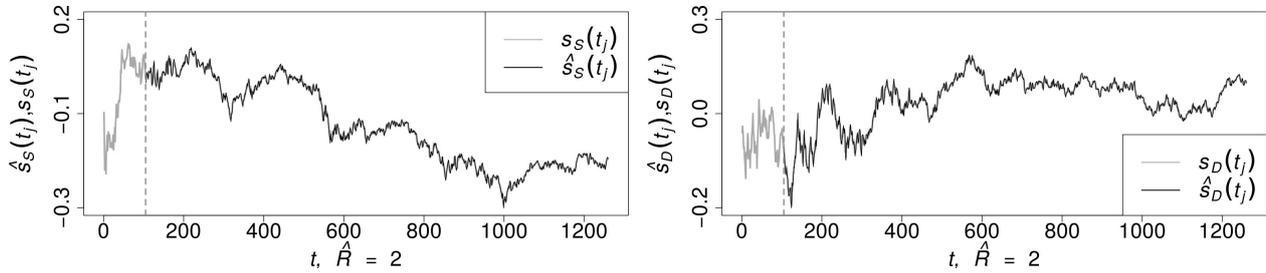

Figure A1. Functional data prediction, one-month rolling window, continuous returns.
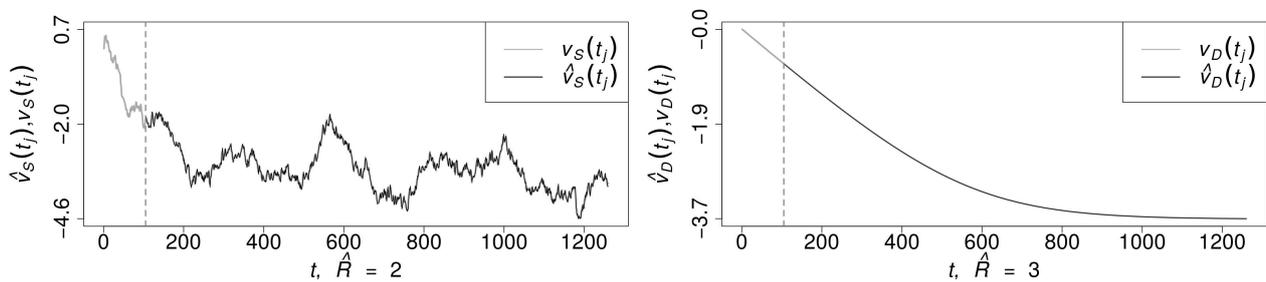

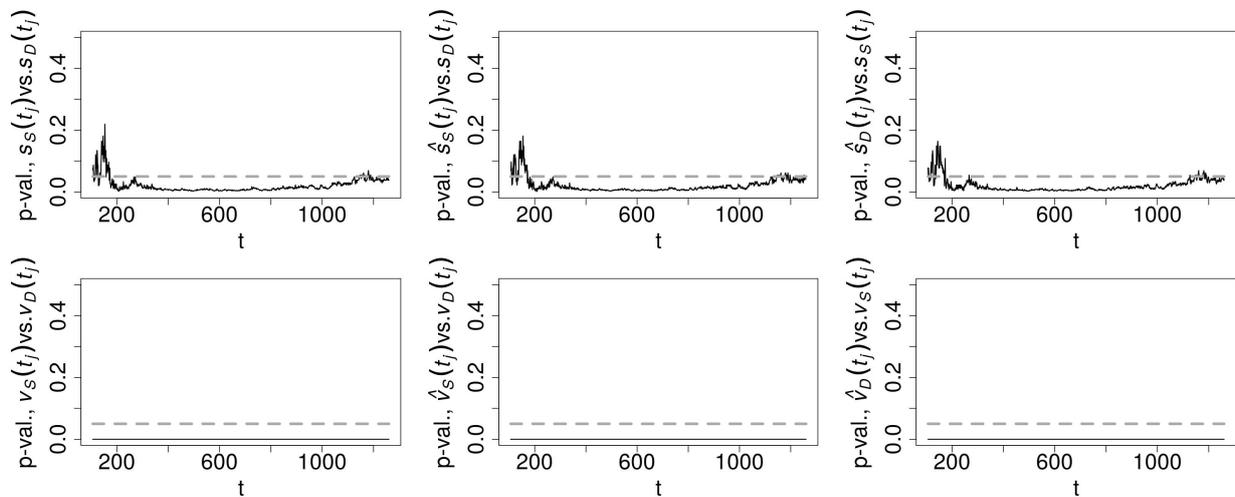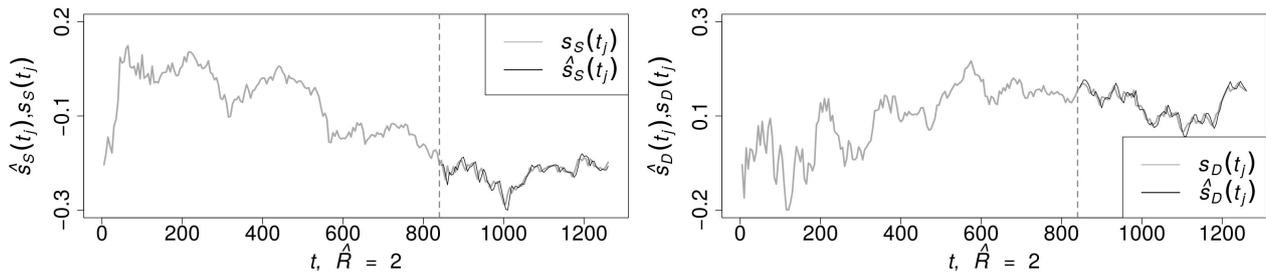**Figure A2.** Functional data prediction, one-month rolling window, continuous volatility.

Figure A3. *K-S* test, one-month rolling window, continuous observations, cross-comparison.
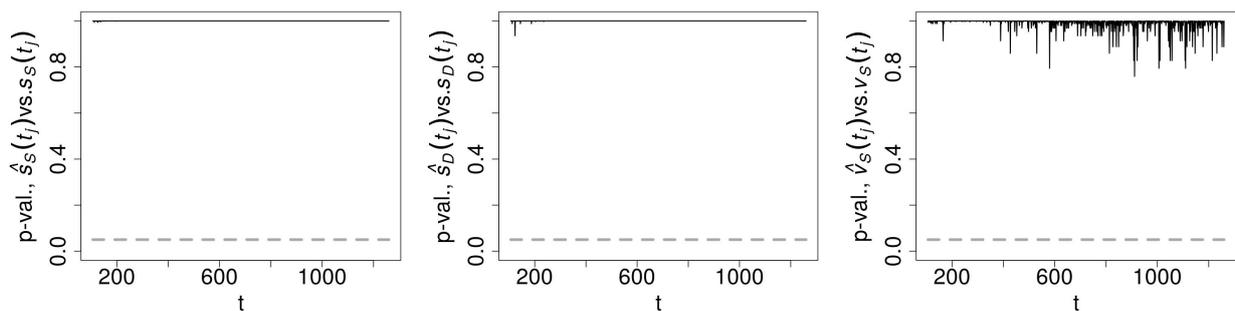
Figure A4. *K-S* test, one-month rolling window, continuous observations, comparison with the true underlying process.

Figure A5. Functional data prediction, eight-month rolling window, daily returns.
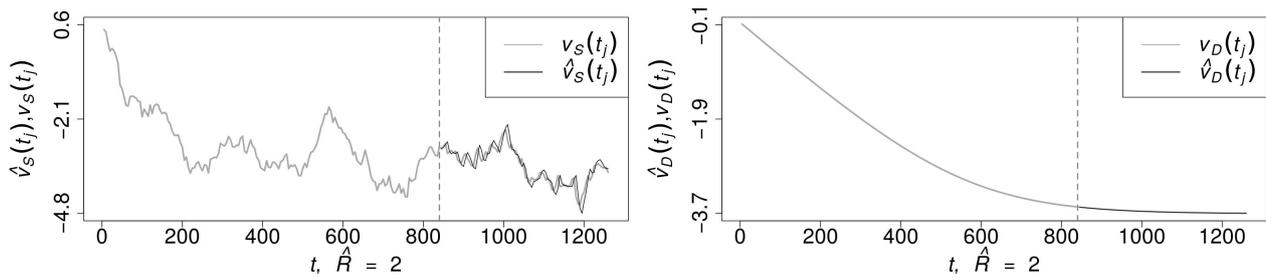
Figure A6. Functional data prediction, eight-month rolling window, daily volatility.
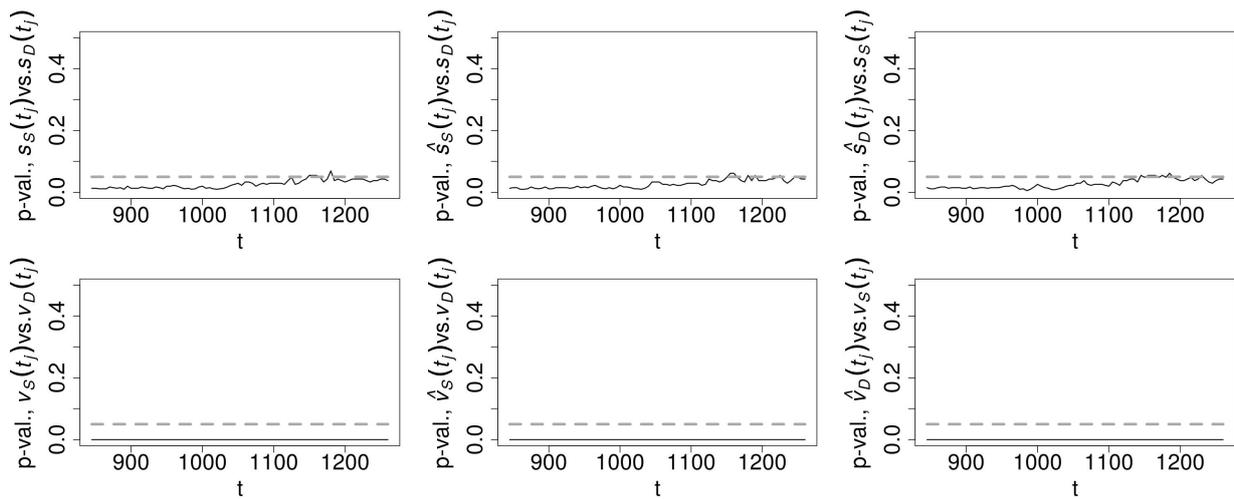
Figure A7. *K-S* test, eight-month rolling window, daily observations, cross-comparison.
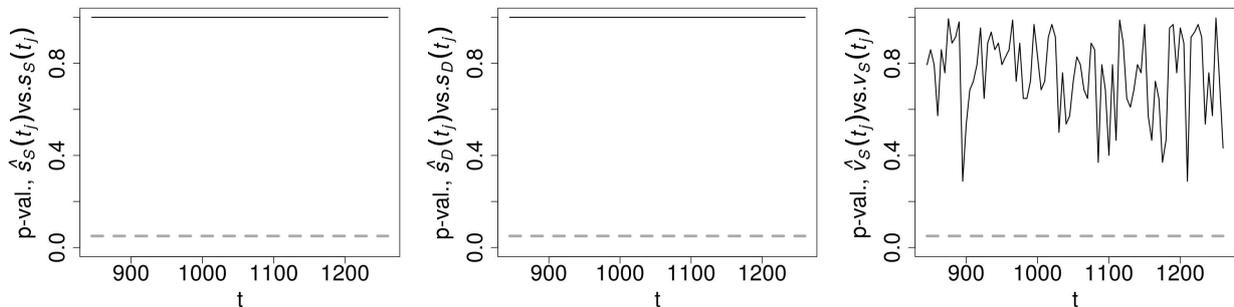
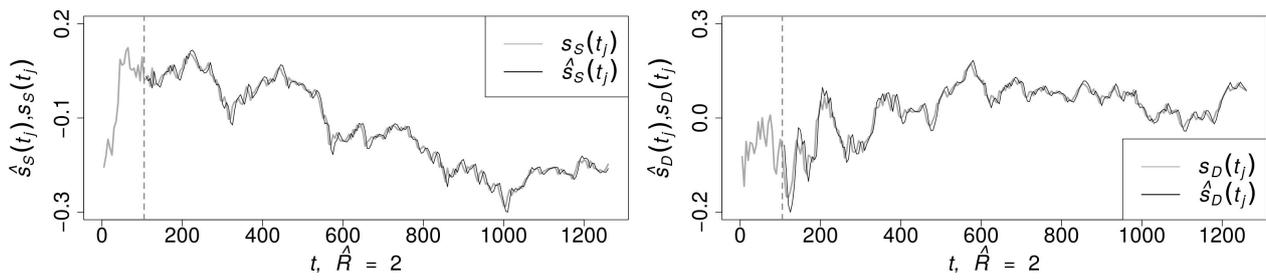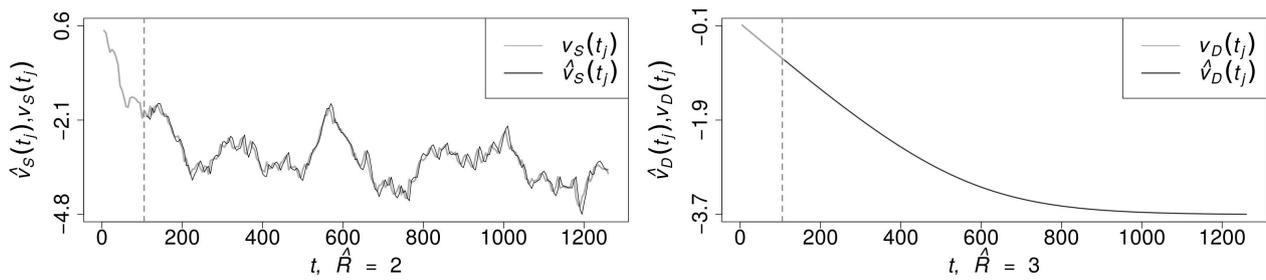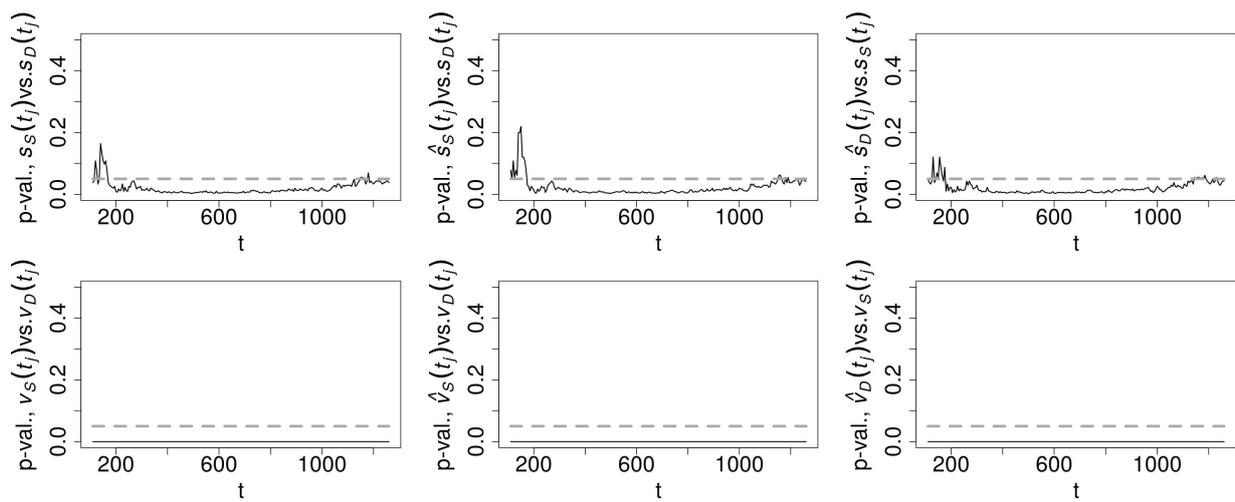**Figure A8.** *K-S* test, eight-month rolling window, daily observations, comparison with the true underlying process.

Figure A9. Functional data prediction, one-month rolling window, daily returns.

Figure A10. Functional data prediction, one-month rolling window, daily volatility.

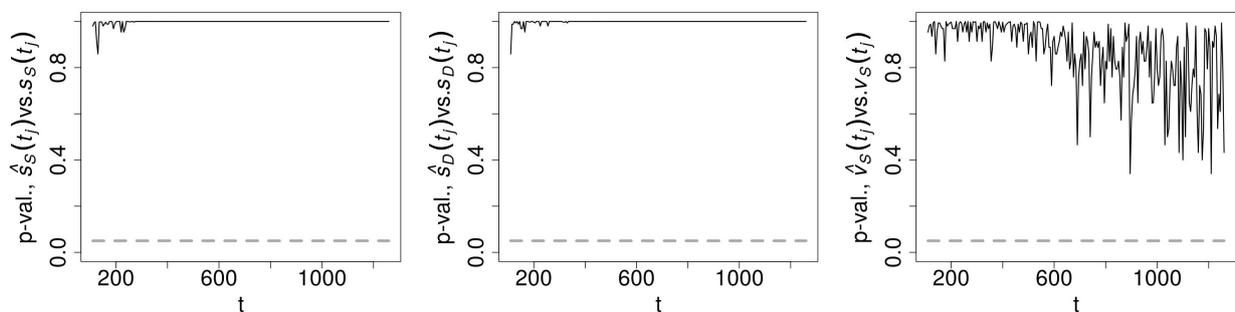**Figure A11.** *K-S* test, one-month rolling window, daily observations, cross-comparison.

Figure A12. *K-S* test, one-month rolling window, daily observations, comparison with the true underlying process.

Appendix B.2. MLE Results

This section presents detailed information on the 420 rolling-window estimates. Figures A14–A18 present the estimation results for each parameter. The three consecutive plots, from left to right, illustrate the distribution of the estimated values eliminating

the 2.5% tail on each side to avoid plot distortion due to extreme values, the estimation bias, and the rejection rate, respectively. Within each plot, the label "*SS*" corresponds to the utilization of the SV model for estimating SV underlying processes, "*SD*" corresponds to the employment of the SV model for DV underlying process estimation, "*DD*" corresponds the use of the DV model for estimating DV underlying processes, and "*DS*" corresponds to the use of the DV model for SV underlying process estimation. These figures suggest that MLE offered dependable estimations under accurate specifications. However, its performance lacked consistency when the underlying process was misspecified.
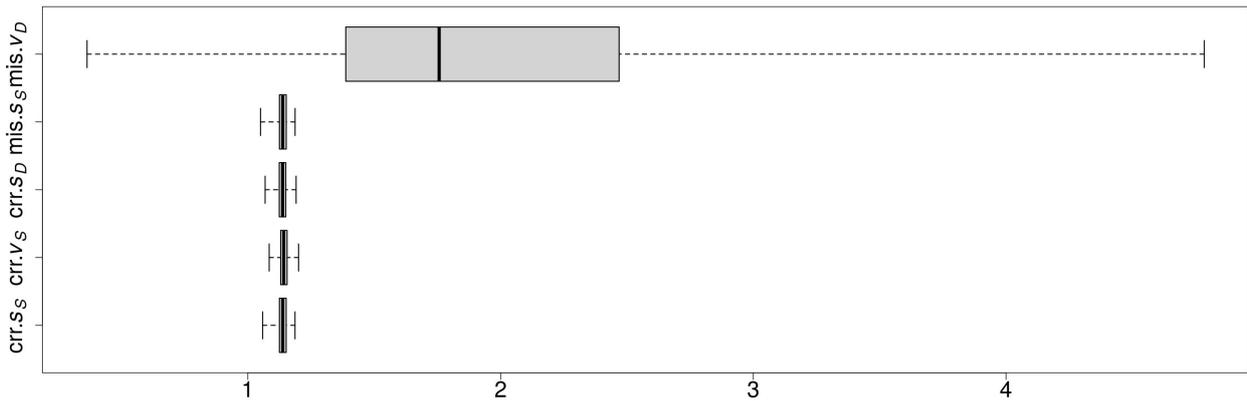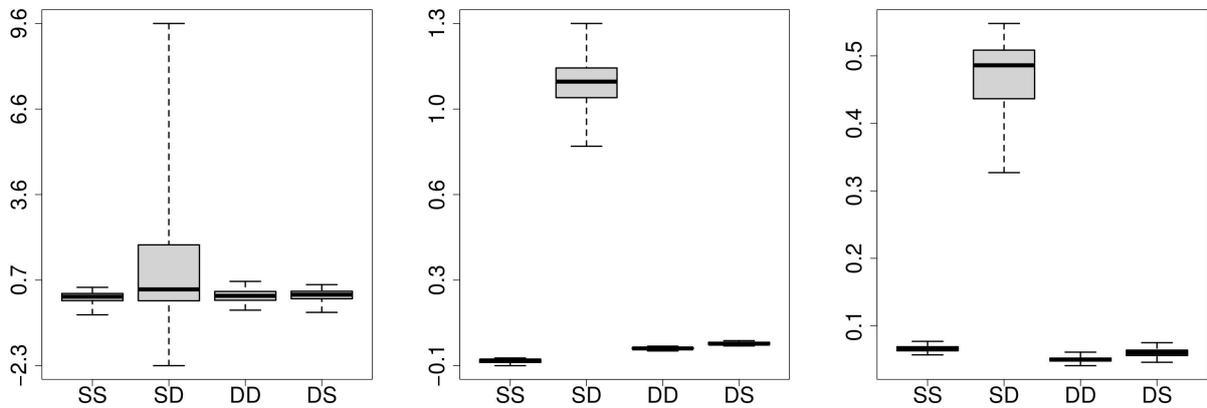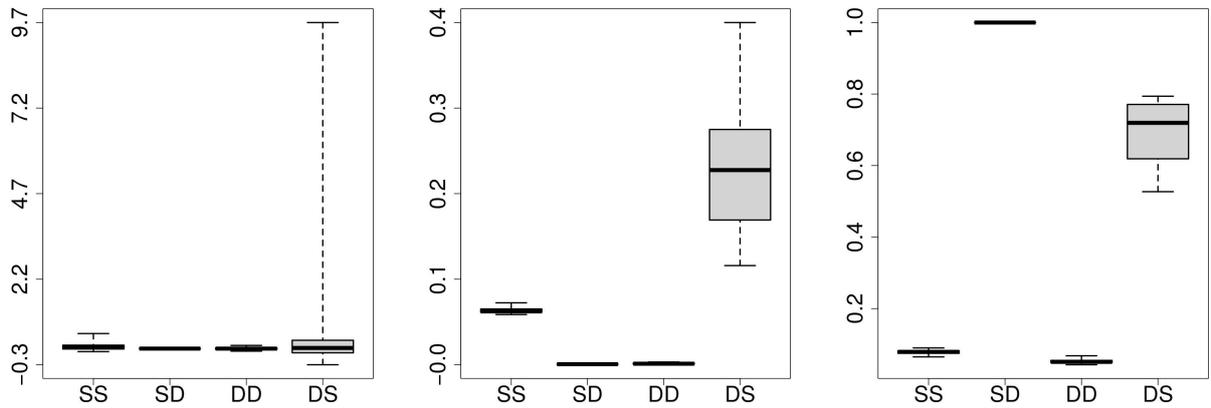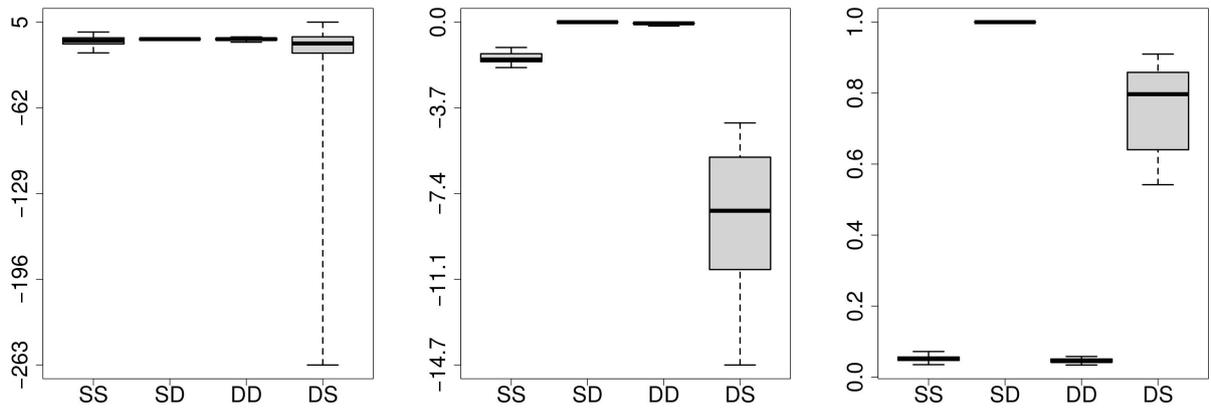

Figure A13. Unaffected MLE vs. FDA.

(**a**) Distribution of the estimates.   (**b**) Distribution of the empirical bias.   (**c**) Distribution of the rejection rates.

Figure A14. Estimation of $\hat{a}$.

(**a**) Distribution of the estimates.   (**b**) Distribution of the empirical bias.   (**c**) Distribution of the rejection rates.

Figure A15. Estimation of $\hat{\alpha}$.
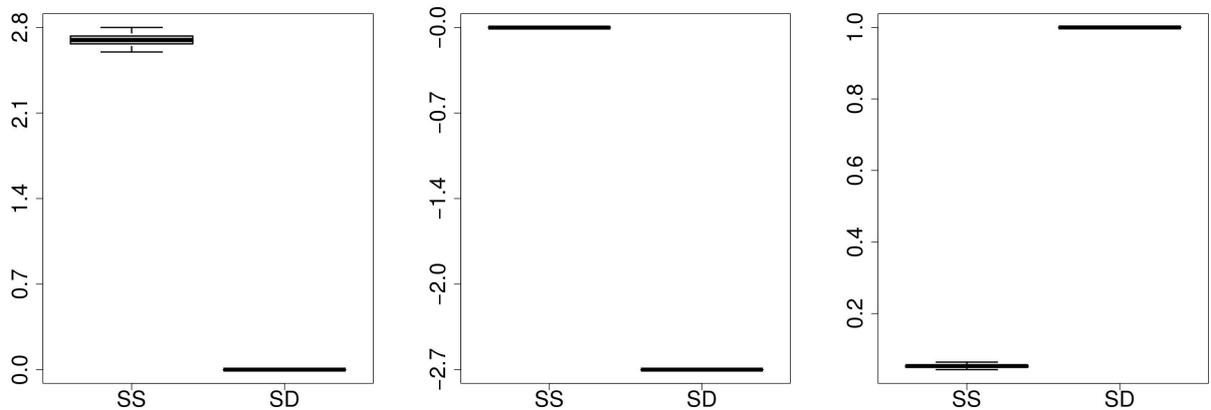

(**a**) Distribution of the estimates.   (**b**) Distribution of the empirical bias.   (**c**) Distribution of the rejection rates.

Figure A16. Estimation of $\hat{\beta}$.

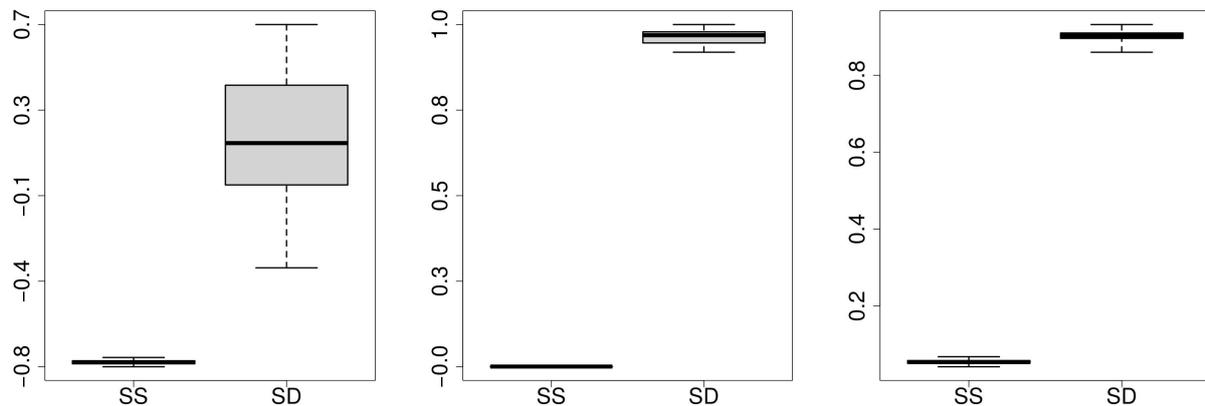(**a**) Distribution of the estimates.   (**b**) Distribution of the empirical bias.   (**c**) Distribution of the rejection rates.

Figure A17. Estimation of $\hat{\sigma}$.

(**a**) Distribution of the estimates.　　(**b**) Distribution of the empirical bias.　　(**c**) Distribution of the rejection rates.

Figure A18. Estimation of $\hat{\rho}$.

References

1. Buccheri, G.; Corsi, F.; Flandoli, F.; Livieri, G. The continuous-time limit of score-driven volatility models. *J. Econom.* **2021**, *221*, 655–675. [CrossRef]
2. Merton, R.C. On estimating the expected return on the market: An exploratory investigation. *J. Financ. Econ.* **1980**, *8*, 323–361. [CrossRef]
3. Merton, R.C.; Samuelson, P.A. *Continuous-Time Finance*; Blackwell: Boston, MA, USA, 1992.
4. Stentoft, L. American option pricing with discrete and continuous time models: An empirical comparison. *J. Empir. Financ.* **2011**, *18*, 880–902. [CrossRef]
5. Badescu, A.; Cui, Z.; Ortega, J.-P. Non-affine GARCH option pricing models, variance-dependent kernels, and diffusion limits. *J. Financ. Econom.* **2017**, *15*, 602–648. [CrossRef]
6. Corradi, V. Reconsidering the continuous time limit of the GARCH (1, 1) process. *J. Econom.* **2000**, *96*, 145–153. [CrossRef]
7. Corradi, V.; Distaso, W. *Deterministic versus Stochastic Volatility*; Manuscript; University of Warwick: Coventry, UK, 2010.
8. Das, S.R. The surprise element: Jumps in interest rates. *J. Econom.* **2002**, *106*, 27–65. [CrossRef]
9. Duan, J.-C.; Ritchken, P.; Sun, Z. Approximating GARCH-JUMP models, jump-diffusion processes, and option pricing. *Math. Financ. Int. J. Math. Stat. Financ. Econ.* **2006**, *16*, 21–52. [CrossRef]
10. Giraitis, L.; Leipus, R.; Surgailis, D. Recent advances in ARCH modelling. In *Long Memory in Economics*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 3–38.
11. Hafner, C.M.; Laurent, S.; Violante, F. Weak diffusion limits of dynamic conditional correlation models. *Econom. Theory* **2017**, *33*, 691–716. [CrossRef]
12. Trifi, A. Issues of aggregation over time of conditional heteroscedastic volatility models: What kind of diffusion do we recover? *Stud. Nonlinear Dyn. Econom.* **2006**, *10*, 4. [CrossRef]
13. Nelson, D.B. ARCH models as diffusion approximations. *J. Econom.* **1990**, *45*, 7–38. [CrossRef]
14. Alexander, C.; Lazar, E. The continuous limit of weak GARCH. *Econom. Rev.* **2021**, *40*, 197–216. [CrossRef]
15. Singleton, K.J. *Empirical Dynamic Asset Pricing*; Princeton University Press: Princeton, NJ, USA, 2009.
16. Sundaresan, S.M. Continuous-time methods in finance: A review and an assessment. *J. Financ.* **2000**, *55*, 1569–1622. [CrossRef]
17. Aydin, D.; Yilmaz, E. Censored nonparametric time-series analysis with autoregressive error models. *Comput. Econ.* **2021**, *58*, 169–202. [CrossRef]
18. Bosq, D. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*; Springer Science & Business Media: New York, NY, USA, 2012; Volume 110.
19. Fan, J.; Yao, Q. *Nonlinear Time Series: Nonparametric and Parametric Methods*; Springer: New York, NY, USA, 2003; Volume 20.
20. Härdle, W.; Lütkepohl, H.; Chen, R. A review of nonparametric time series analysis. *Int. Stat. Rev.* **1997**, *65*, 49–72. [CrossRef]
21. Heiler, S. *A Survey on Nonparametric Time Series Analysis*; Technical Report, CoFE Discussion Paper; ZBW—Leibniz-Informationszentrum Wirtschaft: Kiel, Germany, 1999.
22. Kleppe, T.S.; Yu, J.; Skaug, H.J. Maximum likelihood estimation of partially observed diffusion models. *J. Econom.* **2014**, *180*, 73–80. [CrossRef]
23. Kutoyants, Y.A. *Identification of Dynamical Systems with Small Noise*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 300.
24. Pena, D.; Tiao, G.C.; Tsay, R.S. *A Course in Time Series Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
25. Ryabko, D. *Asymptotic Nonparametric Statistical Analysis of Stationary Time Series*; Springer: Berlin/Heidelberg, Germany, 2019.

26. Zhu, B.; Song, P.X.-K.; Taylor, J.M. Stochastic functional data analysis: A diffusion model-based approach. *Biometrics* **2011**, *67*, 1295–1304. [CrossRef]

27. Aït-Sahalia, Y.; Kimmel, R. Maximum likelihood estimation of stochastic volatility models. *J. Financ. Econ.* **2007**, *83*, 413–452. [CrossRef]

28. Durham, G.B.; Gallant, A.R. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econ. Stat.* **2002**, *20*, 297–338. [CrossRef]

29. Berry, T.; Giannakis, D.; Harlim, J. Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E* **2015**, *91*, 032915. [CrossRef]

30. Matzner-Løfber, E.; Gannoun, A.; Gooijer, J.G.D. Nonparametric forecasting: A comparison of three kernel-based methods. *Commun. Stat.-Theory Methods* **1998**, *27*, 1593–1617. [CrossRef]

31. Lahiri, S.N. On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhyā: Indian J. Stat. Ser. A* **1996**, *58*, 403–417.

32. Cressie, N.A. *Statistics for Spatial Data*; Wiley: New York, NY, USA, 1991.

33. Kurisu, D. On nonparametric inference for spatial regression models under domain expanding and infill asymptotics. *Stat. Probab. Lett.* **2019**, *154*, 108543. [CrossRef]

34. Chen, X. Large sample sieve estimation of semi-nonparametric models. *Handb. Econom.* **2007**, *6*, 5549–5632.

35. Chen, X.; Liao, Z. Sieve semiparametric two-step GMM under weak dependence. *J. Econom.* **2015**, *189*, 163–186. [CrossRef]

36. Grenander, U. *Abstract Inference*; Wiley: New York, NY, USA; Toronto, ON, Canada, 1981.

37. Chen, T.; DeJuan, J.; Tian, R. Distributions of GDP across versions of the Penn World Tables: A functional data analysis approach. *Econ. Lett.* **2018**, *170*, 179–184. [CrossRef]

38. Ramsay, J. *Functional Data Analysis*; Wiley Online Library: Hoboken, NJ, USA, 2005.

39. Adebayo, F.A.; Sivasamy, R.; Shangodoyin, D.K. Forecasting stock market series with ARIMA model. *J. Stat. Econom. Methods* **2014**, *3*, 65–77.

40. Khan, S.; Alghulaiakh, H. ARIMA model for accurate time series stocks forecasting. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *11*, 7. [CrossRef]

41. Mondal, P.; Shit, L.; Goswami, S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *Int. J. Comput. Sci. Eng. Appl.* **2014**, *4*, 13. [CrossRef]

42. Huang, M.; Rojas, R.; Convery, P. Forecasting stock market movements using Google Trend searches. *Empir. Econ.* **2022**, *59*, 2821–2839. [CrossRef]

43. Kumar, M.; Thenmozhi, M. Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *Int. J. Banking Account. Financ.* **2014**, *5*, 284–308. [CrossRef]

44. Dai, Z.; Dong, X.; Kang, J.; Hong, L. Forecasting stock market returns: New technical indicators and two-step economic constraint method. *N. Am. J. Econ. Financ.* **2020**, *53*, 101216. [CrossRef]

45. Li, A.; Bastos, G.S. Stock market forecasting using deep learning and technical analysis: A systematic review. *IEEE Access* **2020**, *8*, 185232–185242. [CrossRef]

46. Lo, A.W.; Mamaysky, H.; Wang, J. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *J. Financ.* **2000**, *55*, 1705–1765. [CrossRef]

47. Agrawal, J.G.; Chourasia, V.; Mittra, A. State-of-the-art in stock prediction techniques. *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.* **2013**, *2*, 1360–1366.

48. Chen, Y.J.; Chen, Y.M.; Lu, C.L. Enhancement of stock market forecasting using an improved fundamental analysis-based approach. *Soft Comput.* **2017**, *21*, 3735–3757. [CrossRef]

49. Seng, D.; Hancock, D.R. Fundamental analysis and the prediction of earnings. *Int. J. Bus. Manag.* **2012**, *7*, 32. [CrossRef]

50. Jouzdani, N.M.; Panaretos, V.M. Functional data analysis with rough sampled paths? *arXiv* **2021**, arXiv:2105.12035.

51. Kadison, R.V.; Liu, Z. The Heisenberg relation-mathematical formulations. *SIGMA Symmetry Integr. Geom. Methods Appl.* **2014**, *10*, 009. [CrossRef]

52. Davidson, K.R.; Donsig, A.P. *Real Analysis and Applications*; Springer: New York, NY, USA, 2010.

53. Claeskens, G.; Krivobokova, T.; Opsomer, J.D. Asymptotic properties of penalized spline estimators. *Biometrika* **2009**, *96*, 529–544. [CrossRef]

54. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327. [CrossRef]

55. Engle, R.F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econom. J. Econom. Soc.* **1982**, *50*, 987–1007. [CrossRef]

56. Meddahi, N. *An Eigenfunction Approach for Volatility Modeling*; Cahier de Recherche; University of Montreal: Montreal, QC, Canada, 2001.

57. Christoffersen, P.; Elkamhi, R.; Feunou, B.; Jacobs, K. Option valuation with conditional heteroskedasticity and nonnormality. *Rev. Financ. Stud.* **2010**, *23*, 2139–2183. [CrossRef]

58. Wu, X.; Zhou, H.; Wang, S. Estimation of market prices of risks in the GARCH diffusion model. *Econ. Res.—Ekon. Istraživanja* **2018**, *31*, 15–36. [CrossRef]

59. Wang, Y. Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Stat.* **2002**, *30*, 754–783. [CrossRef]

60. Ledoit, O.; Santa-Clara, P. *Relative Pricing of Options with Stochastic Volatility*; Los Angeles Finance Working Paper; University of California: Los Angeles, CA, USA, 1998; pp. 9–98.

61. Hansen, P.R.; Lunde, A. A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *J. Appl. Econom.* **2005**, *20*, 873–889. [CrossRef]

62. Leitch, G.; Tanner, J.E. Economic forecast evaluation: Profits versus the conventional error measures. *Am. Econ. Rev. 1991*, *81*, 580–590.

63. Aït-Sahalia, Y. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **2002**, *70*, 223–262. [CrossRef]

64. Chui, C.K. Concerning rates of convergence of Riemann sums. *J. Approx. Theory* **1971**, *4*, 279–287. [CrossRef]

65. Davidson, J. *Stochastic Limit Theory: An Introduction for Econometricians*; OUP Oxford: Oxford, UK, 1994.