JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2026, VOL. 00, NO. 0, 1–16: Theory and Methods
https://doi.org/10.1080/01621459.2026.2687693

Online Change Rate Learning for Functional Data

Deru Kong[a], Ping Yu[b], Tiejun Tong[c], and WenWu Wang[a]

[a]School of Statistics and Data Science, Qufu Normal University, Qufu, China; [b]HKU Business School, University of Hong Kong, Pokfulam, Hong Kong; [c]Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong

ABSTRACT

With advances in modern data collection technologies, functional data are increasingly received in a streaming manner. Although numerous methods have been developed to model such data, most focus on estimating the mean and covariance functions, while giving limited attention to their derivatives. In this article, we introduce an online method for estimating the change rates of the mean and covariance functions, enabling real-time updates with high statistical efficiency. The method dynamically updates by incorporating incoming data and updating summary statistics of historical data. We establish the asymptotic normality of the estimators, classify functional data into three types based on the distributional properties (non-dense, semi-dense and ultra-dense), and moreover provide a data-driven procedure for the online bandwidth selection. By minimizing the loss of relative efficiency, we show that the proposed online change rate estimators perform comparably to the offline kernel estimators using the full data. We further compare our new estimators with representative offline and online methods through extensive simulations, demonstrating that they achieve an effective balance between statistical accuracy and computational time. Two real applications to hourly power consumption and traffic accident data, both capturing the changing trends in real time, further validate the broad applicability and substantial practical benefits of our new method. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY
Received April 2025
Accepted May 2026

KEYWORDS
Change rate; Dynamic updating process; Kernel-based method; Online learning; Streaming functional data

1. Introduction

1.1. *Offline Derivatives Estimation in Functional Data Analysis*

The increasing complexity and high dimensionality of modern data have led to a surge of interest in functional data analysis (FDA). Unlike conventional scalar or vector observations, functional data capture the smooth trajectories of underlying stochastic processes, while preserving both temporal and inter-variable dynamics. These properties allow FDA to extract richer information, thus, driving its widespread application across fields such as biomedicine, finance, and social sciences (He, Müller, and Wang 2000; Müller 2005; Zhang and Chen 2007; Karuppusami et al. 2022). As a result, FDA has become a key framework for analyzing complex dynamic systems by modeling them as infinite-dimensional stochastic processes. Within this framework, estimating the change rate, also known as the first-order derivative, is of particular theoretical and practical importance. For instance, it enables detailed individual-level characterization (Hall, Müller, and Yao 2009; Grith et al. 2018), facilitates group-level comparisons (McKeague et al. 2015; Simpkin et al. 2018), enhances predictive modeling (Mas and Pumo 2009; Rossi and Villa-Vialaneix 2011), and aids in exploring critical transitions (Sadeghioon et al. 2014).

Motivated by these diverse applications, extensive efforts have been devoted to developing methods for estimating the derivatives of functional data. Among them, nonparametric techniques such as splines, wavelets, and kernels have long served as foundational tools for the offline estimation. The spline smoothing can be interpreted as a form of the kernel smoothing under the reproducing kernel Hilbert space framework. For practical use, it requires careful placement of knots across the entire dataset and tends to oversmooth the regions where the true derivative exhibits sharp changes (Sangalli et al. 2009; Cao et al. 2012). The performance of the wavelet estimators hinges on a set of tuning parameters, including the choice of mother wavelet, the range of dilation, and the thresholding rule (Antoniadis and Fan 2001). The absence of closed-form analytical expressions for estimators in the smooth wavelet basis also restricts their broad applicability in FDA (Pigoli and Sangalli 2012). Additionally, both the spline and wavelet methods rely on global basis expansions whose coefficients are coupled across the domain through roughness penalties or thresholding, which further complicates the incremental updates as new data arrive. In contrast, the kernel-based approach is intrinsically local and admits explicit and directly computable derivative estimators which are expressed in the sufficient statistics (Liu and Müller 2009; Dai, Müller, and Tao 2018; Ghale-Joogh and Hosseini-Nasab 2021). These statistics, in turn, enable efficient and data-driven bandwidth selection that can be updated incrementally as new data arrive. This inherent adaptability makes the kernel smoothing particularly well-suited for real-time FDA.

CONTACT WenWu Wang wangwenwu@qfnu.edu.cn School of Statistics and Data Science, Qufu Normal University, 57 Jingxuan West Road, Qufu, Shandong 273165, China.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

1.2. Online Kernel Learning for Streaming Data

With the growing prevalence of streaming data in areas like critical infrastructure and environmental sensing, timely and reliable estimates of change that can be made quickly are essential for early warnings and interpretable interventions. To illustrate with examples, in power system, the ramp rates of load and generation control the frequency-security margins and the timing of reserve deployment (Entso-E 2016). In traffic safety, real-time detection of sudden acceleration and sharp movements helps prevent crashes and activates automatic emergency braking (Treiber, Hennecke, and Helbing 2000). Related change rate estimation also arises in other domains, such as healthcare (Pigoli and Sangalli 2012), earthquake warning (Allen et al. 2009), and industrial monitoring (Sadeghioon et al. 2014).

Nevertheless, traditional offline derivative estimation methods are ill-equipped for streaming data because their high computational cost and memory demands introduce latency, undermining the real-time performance and stability. In contrast, to the best of our knowledge, the online FDA literature has given little attention to the derivatives. Existing work instead focuses on other objectives, such as the online estimation of the mean and covariance functions (Yang and Yao 2023) and of the functional principal components (Patilea and Racine 2024).

Online kernel methods appear promising for streaming data by capturing the nonlinear dynamics flexibly while maintaining the computational efficiency (Lu et al. 2016; Kong and Xia 2019; Yang and Yao 2023). However, extending the online kernel methods to the estimation of change rates introduces methodological challenges. A key difficulty lies in the bandwidth selection, which is crucial for controlling the smoothness of the estimator. In the online setting, this bandwidth must be continuously adjusted to accommodate the new incoming data, creating a fundamental trade-off between the statistical and computational efficiency. Furthermore, although empirical comparisons between the online and offline methods exist, their theoretical properties remain insufficiently explored. Motivated by these considerations, this article addresses the challenges and contributes in the following three key aspects:

- We propose, for the first time, an online kernel method for estimating the change rates of the mean and covariance functions of functional data. Our approach efficiently updates the sufficient summary statistics with a candidate bandwidth sequence and continuously adapts to streaming data, achieving high computational efficiency with reduced memory usage.
- We establish theoretical guarantees by categorizing functional data based on the distributional properties, deriving convergence rates of the data-driven bandwidths, and demonstrating that our new estimators are nearly equivalent to the offline counterparts.
- We also develop efficient algorithms that enable fast updates and modifications of estimators, with simulations confirming that our new approach achieves an excellent balance between the statistical and computational efficiency.

The rest of the article is organized as follows. In Section 2, we introduce a novel online estimation framework for the first-order derivatives of the mean and covariance functions and develop the corresponding algorithms. Section 3 provides theoretical guarantees, including the asymptotic normality, a data-driven bandwidth selection criterion and the relative efficiency. Simulation studies in Section 4 demonstrate that our new online method achieves high statistical and computational efficiency. In Section 5, we apply the proposed method to a real data example. Finally, we conclude the article with some discussions in Section 6. The supplementary material provides the regularity assumptions, the corollary for the covariance function, the data-driven online bandwidth selection procedure, some extra simulations, an extended real data example, and the technical results.

2. New Method for Online Change Rate Learning

In this section, we propose a novel online estimation method for the first-order derivatives of the mean and covariance functions, and develop two algorithms to implement the proposed method efficiently for streaming functional data.

2.1. Model of Streaming Functional Data

To formalize this approach, we model streaming functional data as follows. The sample of functions is regarded as realizations of an L^2 stochastic process $\{X(t) : t \in I\}$ on an interval I. We specify the model by decomposing $X(t)$ into two parts,

$$X(t) = \mu(t) + \Phi(t), \tag{1}$$

where $\mu(t)$ is the mean function of $X(t)$ and $\Phi(t)$, as the stochastic part of $X(t)$, generates the within-subject covariance function of $X(t)$, satisfying $E(\Phi(t)) = 0$ for any $t \in I$ and $\mathrm{Cov}(\Phi(s), \Phi(t)) = \gamma(s,t)$ for all $s,t \in I$. Intuitively, $\Phi(t)$ can be interpreted as the functional random effect, representing the subject-specific deviations from $\mu(t)$. Its covariance function $\gamma(s,t)$ characterizes the within-curve variability structure, quantifying how fluctuations at time s co-vary with those at time t.

We now describe the process for generating streaming functional data. There are k blocks (periods) of the data, where $k = 1,\dots,K$, and K is the current block which increases to infinity as the data accumulate. For the kth block, there are n_k independent subjects. For the ith subject in the kth block, the process realizations are given by $\{X_{ki}(t) : k = 1,\dots,K, i = 1,\dots,n_k\}$. We allow the sampling scheme to vary across the subjects. For the ith subject in the kth block, there are m_{ki} measurements taken at the time points $\left\{T_{kij} : k = 1,\dots,K, i = 1,\dots,n_k, j = 1,\dots,m_{ki}\right\}$. Assume that the corresponding outcome observations at these time points are contaminated with random noise ϵ_{kij}. Then our model of streaming functional data is given by

$$Y_{kij} = X_{ki}(T_{kij}) + \epsilon_{kij} = \mu(T_{kij}) + \Phi_{ki}(T_{kij}) + \epsilon_{kij}, \tag{2}$$

where μ and Φ are defined in model (1), and ϵ_{kij} are independent and identically distributed with $E(\epsilon_{kij}) = 0$ and $\mathrm{Var}(\epsilon_{kij}) = \sigma^2$. The realized observations are $\left\{\left(T_{kij}, Y_{kij}\right) : k = 1,\dots,K, i = 1,\dots,n_k, j = 1,\dots,m_{ki}\right\}$. Assume that the first-order derivatives of $\mu(t)$ and $\gamma(s,t)$ exist. We aim to provide efficient, real-time estimation for the first-order derivatives of $\mu(t)$ and $\gamma(s,t)$ in model (2).

2.2. Online Change Rate Learning for Mean Function

Our first objective is to estimate $\mu^{(1)}(t)$, the first-order derivative of the mean function in model (2). In the online setting, the estimation of $\mu^{(1)}(t)$ is based on the current K data blocks, denoted by $\tilde{\mu}^{(1)}_{(K)}(t)$, for which we apply a tilde to label the objects estimated by an online method. Given the minimal model assumptions, we employ a nonparametric technique using local quadratic smoother to obtain $\tilde{\mu}^{(1)}_{(K)}(t)$. Assuming that $\mu(t)$ is three times continuously differentiable, we solve the following locally weighted least squares problem:

$$\begin{aligned}&\left(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2\right)\\ &\quad= \underset{\beta_0,\beta_1,\beta_2\in\mathbb{R}}{\arg\min} \sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{m_{ki}} \left\{ Y_{kij} - \beta_0 - \beta_1(T_{kij}-t) \right.\\ &\qquad \left. - \beta_2(T_{kij}-t)^2 \right\}^2 W_{\tilde{h}_{\mu(K)}}\left(T_{kij}-t\right),\end{aligned}$$

where $\tilde{\mu}^{(1)}_{(K)}(t) = \tilde{\beta}_1$, $\tilde{h}_{\mu(K)}$ is the bandwidth selected by an online method using the first K data blocks, and $W_h(\cdot) = W(\cdot/h)/h$ with a kernel function $W(\cdot)$. The subscript k or K denotes that the statistics use only data from the kth or Kth block; the subscript (k) or (K) denotes that the statistics are based on online learning with the kth or Kth data block as well as the summary statistics from historical data. The first-order derivative online estimation $\tilde{\mu}^{(1)}_{(K)}(t)$ can be explicitly expressed as

$$\tilde{\mu}^{(1)}_{(K)}(t) = \varepsilon_{3,2}^{\top}\left\{\sum_{k=1}^{K} P_k(t;\tilde{h}_{\mu(K)})\right\}^{-1}\left\{\sum_{k=1}^{K} q_k(t;\tilde{h}_{\mu(K)})\right\}, \quad (3)$$

where $\varepsilon_{3,2} = (0,1,0)^{\top}$ and $P_k(t;\tilde{h}_{\mu(K)})$ and $q_k(t;\tilde{h}_{\mu(K)})$ depend only on the data of the kth block and the bandwidth $\tilde{h}_{\mu(K)}$, given by

$$P_k(t;\tilde{h}_{\mu(K)}) = \sum_{i=1}^{n_k}\sum_{j=1}^{m_{ki}} W_{\tilde{h}_{\mu(K)}}(T_{kij}-t)T_{kij}(t)T_{kij}^{\top}(t),$$
$$q_k(t;\tilde{h}_{\mu(K)}) = \sum_{i=1}^{n_k}\sum_{j=1}^{m_{ki}} W_{\tilde{h}_{\mu(K)}}(T_{kij}-t)T_{kij}(t)Y_{kij},$$

with $T_{kij}(t) = (1, T_{kij}-t, (T_{kij}-t)^2)^{\top}$. The minimization problem yields a closed-form solution, which can be decomposed into two sufficient statistics as $\{\sum_{k=1}^{K} P_k(t;\tilde{h}_{\mu(K)}),\ \sum_{k=1}^{K} q_k(t;\tilde{h}_{\mu(K)})\}$. These two statistics are additive, meaning that when a new data block is added to the dataset, only these two statistics need to be updated, thus, efficiently updating the estimator (3) without the need to recompute the entire dataset.

A key challenge for the online estimation is to adaptively select the bandwidth $\tilde{h}_{\mu(K)}$, which requires recalculating the bandwidth using the entire dataset as new data arrive. This recalibration introduces a computational bottleneck, as it involves incorporating all accumulated data. Since the data stream is continuous, the volume of observations and $\{P_k(t;\tilde{h}_{\mu(K)}),\ q_k(t;\tilde{h}_{\mu(K)})\}_{k=1}^{K}$ also increase over time. Storing all statistics would significantly increase the computational complexity, while not storing them would compromise the statistical efficiency of the estimator. It is hence crucial to achieve an appropriate balance between the computational and the statistical power.

To address the power challenge, we propose a method for dynamically updating the bandwidth in the online change rate learning using a candidate bandwidth sequence. This method consists of three steps. Specifically, the first step is to generate a candidate bandwidth sequence upon the arrival of each data block, the second step is to update the candidate bandwidth sequence online while considering storage costs, and the third step is to select the bandwidth closest to the current one from the updated candidate sequence. As the amount of the data increases, the online bandwidth decreases. Thus, we establish a monotonically decreasing candidate bandwidth sequence at the Kth data block, $\{\eta_{\mu(K),\tilde{l}}\}_{\tilde{l}=1}^{L}$, which satisfies $\tilde{h}_{\mu(K)} = \eta_{\mu(K),1} > \cdots > \eta_{\mu(K),L}$, where L is the length of the candidate bandwidth sequence. The expressions of $\{\eta_{\mu(K),\tilde{l}}\}_{\tilde{l}=1}^{L}$ and the selection of $\tilde{h}_{\mu(K)}$ are discussed in Section 3.2 and Theorem 4, respectively.

Assuming that there are currently K data blocks, we specify the candidate bandwidth sequence $\{\eta_{\mu(K),\tilde{l}}\}_{\tilde{l}=1}^{L}$ at block K. When the $(K+1)$th data block arrives, we select the current bandwidth $\tilde{h}_{\mu(K+1)}$ and compute the new candidate sequence $\{\eta_{\mu(K+1),\tilde{l}}\}_{\tilde{l}=1}^{L}$. We then update the candidate bandwidth sequence, referring to the newly updated sequence as the centroid. The centroid $\phi_{\mu(K+1),\tilde{l}}$ is given by

$$\phi_{\mu(K+1),\tilde{l}} = (1-\omega_{\mu(K),K+1})\phi_{\mu(K),j_{\tilde{l}(K+1)}} + \omega_{\mu(K),K+1}\eta_{\mu(K+1),\tilde{l}},$$

where the initial centroid $\phi_{\mu(0)\tilde{l}} = 0$, which results in $\phi_{\mu(1),\tilde{l}} = \eta_{\mu(1),\tilde{l}}$. The weight $\omega_{\mu(K),K+1} = s_{K+1,1}/S_{K+1,1}$ reflects the importance of the $(K+1)$th data block relative to the entire available dataset, where $S_{K+1,1} = \sum_{k=1}^{K+1} s_{k,1}$ with $s_{k,1} = \sum_{i=1}^{n_k} m_{ki}$ is the size of all available data blocks, and $s_{K+1,1} = \sum_{i=1}^{n_{K+1}} m_{K+1,i}$ is the size of the $(K+1)$th block. The index $j_{\tilde{l}(K+1)}$ corresponds to the third step in the online estimation process. We select the index $j_{\tilde{l}(K+1)}$ from the centroid $\{\phi_{\mu(K),\tilde{l}}\}_{\tilde{l}=1}^{L}$ to approximate $\eta_{\mu(K+1),\tilde{l}}$ by solving the minimization problem:

$$j_{\tilde{l}(K+1)} = \underset{i\in\{1,\ldots,L\}}{\arg\min}\left|\eta_{\mu(K+1),\tilde{l}} - \phi_{\mu(K),i}\right|.$$

The centroid dynamically updates the candidate bandwidth online by adapting new insights from the $(K+1)$th block and updating historical knowledge (from the first K blocks), which eliminates the need to store the sequence of candidate bandwidth for each block. Notably, $\phi_{\mu(K+1),\tilde{l}}$ is a weighted average of $\{\eta_{\mu(K+1),\tilde{l}}, \eta_{\mu(K),\cdot}, \ldots, \eta_{\mu(1),\cdot}\}$, where the dot depends on $j_{\tilde{l}(K+1)}$. The index $j_{\tilde{l}(K+1)}$ tries to minimize the difference between the current candidate bandwidth $\eta_{\mu(K+1),\tilde{l}}$ and the stored centroids $\{\phi_{\mu(K),\tilde{l}}\}_{\tilde{l}=1}^{L}$ from the previous block.

We then update the statistics $\sum_{k=1}^{K+1} P_k(t;\tilde{h}_{\mu(K)})$ and $\sum_{k=1}^{K+1} q_k(t;\tilde{h}_{\mu(K)})$ online by incorporating new data into the historical summary statistics. After determining the index $j_{\tilde{l}(K+1)}$, the online-updated statistics $\tilde{P}_{\mu(K+1),\tilde{l}}(t)$ and $\tilde{q}_{\mu(K+1),\tilde{l}}(t)$ are calculated as

$$\tilde{P}_{\mu(K+1),\tilde{l}}(t) = P_{K+1}(t;\eta_{\mu(K+1),\tilde{l}}) + \tilde{P}_{\mu(K),j_{\tilde{l}(K+1)}}(t),$$

$$\tilde{q}_{\mu_{(K+1)},\tilde{l}}(t) = q_{K+1}(t;\eta_{\mu_{(K+1)},\tilde{l}}) + \tilde{q}_{\mu_{(K)},j_{\tilde{l}(K+1)}}(t),$$

with the initial values $\tilde{P}_{\mu_{(0)},\tilde{l}}(t) = 0$ and $\tilde{q}_{\mu_{(0)},\tilde{l}}(t) = 0$ for $\tilde{l} = 1,\ldots,L$. Taking the updated statistic $\tilde{P}_{\mu_{(K+1)},\tilde{l}}(t)$ as an example, it combines the current statistic from the $(K+1)$th block, $P_{K+1}(t;\eta_{\mu_{(K+1)},\tilde{l}})$, with the stored statistic from the first K blocks, $\tilde{P}_{\mu_{(K)},j_{\tilde{l}(K+1)}}(t)$. This implies that $\tilde{P}_{\mu_{(K+1)},\tilde{l}}(t)$ has the summation form, $\tilde{P}_{\mu_{(K+1)},\tilde{l}}(t) = \sum_{k=1}^{K+1} P_k(t;\eta_{\mu_{(k)},\cdot})$ and similarly for $\tilde{q}_{\mu_{(K+1)},\tilde{l}}(t)$, where the dot index is equal to $\tilde{l}$ when $k = K+1$ and for other values of $k = 1,\ldots,K$, it depends on $j_{\tilde{l}(k)}$. This update ensures that both the most recent data and historical information are considered in the online estimation process.

Since $\tilde{h}_{\mu_{(K+1)}} = \eta_{\mu_{(K+1)},1}$, we select the online-updated statistics with $\tilde{l} = 1$. Finally, we obtain the online estimator with the dynamically adjustable candidate bandwidths of $\mu^{(1)}_{(K+1)}(t)$ as

$$\begin{aligned}\tilde{\mu}^{(1)}_{(K+1)}(t) &= \varepsilon_{3,2}^{\top}\left\{\tilde{P}_{\mu_{(K+1)},1}(t)\right\}^{-1}\tilde{q}_{\mu_{(K+1)},1}(t)\\ &= \varepsilon_{3,2}^{\top}\left\{\sum_{k=1}^{K+1} P_k(t;\eta_{\mu_{(k)},\cdot})\right\}^{-1}\left\{\sum_{k=1}^{K+1} q_k(t;\eta_{\mu_{(k)},\cdot})\right\},\end{aligned} \tag{4}$$

where the dot index in $\eta_{\mu_{(k)},\cdot}$ is discussed above. The estimator is dynamically updated as new data arrive, while maintaining a connection to historical data. This allows the method to efficiently handle a continuous data stream without losing prior information. We define the $K+1$ candidate bandwidths used in (4) as a pseudo-bandwidth sequence for the mean function, denoted by $\{\tilde{\eta}_{\mu_{(K+1)},k}\}_1^{K+1}$, where $\tilde{\eta}_{\mu_{(K+1)},K+1} = \eta_{\mu_{(K+1)},1}$, $\tilde{\eta}_{\mu_{(K+1)},K} = \eta_{\mu_{(K)},j_{1(K+1)}}$, etc. The major steps of the proposed online change rate learning (OCRL) method are illustrated in Figure 1. The statistics inside the solid-line boxes are updated, with the most updated versions being stored, while the older versions can be discarded. In contrast, the statistics inside the dashed-line boxes are computed for each block and can be discarded once the current update task is completed. Our OCRL method for the mean function is summarized in Algorithm 1.

In Algorithm 1, we keep only the summary statistics $\{\tilde{P}_{\mu_{(K)},\tilde{l}}(t),\ \tilde{q}_{\mu_{(K)},\tilde{l}}(t)\}_{\tilde{l}=1}^{L}$ and the current centroids $\{\phi_{\mu_{(K)},\tilde{l}}\}_{\tilde{l}=1}^{L}$, rather than storing the whole historical dataset $\{(T_{kij}, Y_{kij}) : k = 1,\ldots,K,\ i = 1,\ldots,n_k,\ j = 1,\ldots,m_{ki}\}$. By only computing and storing L statistics, the computational time can be significantly reduced, especially when L is small. However, a small L may result in low statistical efficiency. Section 3.3 discusses how to choose the appropriate value for L. A key feature of the OCRL method is its ability to prepare for future data by constructing the statistics using the candidate bandwidth sequence. As new data arrive, we select the statistics that best align with the current data, thereby achieving high estimation efficiency. This performance is further validated by the theoretical results presented in Theorem 4.

2.3. Online Change Rate Learning for Covariance Function

Our second objective is to estimate the first-order partial derivative of the covariance function in model (2). Since the estimation processes for $\gamma^{(1,0)}(s,t)$ and $\gamma^{(0,1)}(s,t)$ are similar, we focus on $\gamma^{(1,0)}(s,t)$ as an example. The partial derivative of the covariance function is denoted as $\gamma^{(u,v)}(s,t) = \frac{\partial^{u+v}}{\partial s^u \partial t^v}\gamma(s,t)$, $u,v \geq 0$. We obtain the online estimation of $\gamma^{(1,0)}(s,t)$ based on the first K data blocks, denoted by $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. Since the covariance function relies on the unknown mean function, we first obtain the online estimation of the mean function, $\tilde{\mu}_{(K)}(t)$. The raw covariance function is then derived as $\tilde{C}^{(K)}_{ki,j,l} = \{Y_{kij} - \tilde{\mu}_{(K)}(T_{kij})\}\{Y_{kil} - \tilde{\mu}_{(K)}(T_{kil})\}$, where $1 \leq j \neq l \leq m_{ki}$ with $m_{ki} \geq 2$. Next, we apply a two-dimensional local quadratic smoother to obtain $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. Assuming that $\gamma(s,t)$ is three times continuously differentiable, we solve the locally weighted least squares problem:

$$\begin{aligned}&\left(\tilde{\beta}_{00},\tilde{\beta}_{10},\tilde{\beta}_{01},\tilde{\beta}_{20},\tilde{\beta}_{02},\tilde{\beta}_{11}\right)\\ &= \mathop{\arg\min}_{\beta_{00},\beta_{10},\beta_{01},\beta_{20},\beta_{02},\beta_{11}\in\mathbb{R}} \sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{1\leq j\neq l\leq m_{ki}}\Big\{\tilde{C}^{(K)}_{ki,j,l} - \beta_{00}\\ &\quad - \beta_{10}(T_{kij}-s) - \beta_{01}(T_{kil}-t) - \beta_{20}\\ &\quad (T_{kij}-s)^2 - \beta_{02}(T_{kil}-t)^2 - \beta_{11}(T_{kij}-s)(T_{kil}-t)\Big\}^2\\ &\quad \times W_{\tilde{h}_{\gamma_{(K)}}}\left(T_{kij}-s\right) W_{\tilde{h}_{\gamma_{(K)}}}\left(T_{kil}-t\right),\end{aligned}$$

where $\tilde{\gamma}^{(1,0)}_{(K)}(s,t) = \tilde{\beta}_{10}$ and $\tilde{h}_{\gamma_{(K)}}$ is the bandwidth selected by an online method using the blocks. The online estimation of the first-order partial derivative, $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$, can be explicitly written as

$$\tilde{\gamma}^{(1,0)}_{(K)}(s,t) = \varepsilon_{6,2}^{\top}\left\{\sum_{k=1}^{K} P_k(s,t;\tilde{h}_{\gamma_{(K)}})\right\}^{-1}\left\{\sum_{k=1}^{K} q_k(s,t;\tilde{h}_{\gamma_{(K)}})\right\}, \tag{5}$$

where $\varepsilon_{6,2} = (0,1,0,0,0,0)^{\top}$, and $P_k(s,t;\tilde{h}_{\gamma_{(K)}})$ and $q_k(s,t;\tilde{h}_{\gamma_{(K)}})$ depend solely on the kth block of data and

Algorithm 1 OCRL for the mean function

Initialization: $K \leftarrow 0$, $S_{K,1} \leftarrow 0$, $\phi_{\mu_{(K)},\tilde{l}} \leftarrow 0$, $\tilde{P}_{\mu_{(K)},\tilde{l}} \leftarrow 0$, $\tilde{q}_{\mu_{(K)},\tilde{l}} \leftarrow 0$, $\tilde{l} = 1,\ldots,L$
While $K < \infty$ **do**
  Input: $\left\{(T_{Kij}, Y_{Kij}) : i = 1,\ldots,n_K,\ j = 1,\ldots,m_{Ki}\right\}$
  for $K = K+1$ **do**
    $S_{K,1} \leftarrow S_{K,1} + \sum_{i=1}^{n_K} m_{Ki}$
    compute the online bandwidth $\tilde{h}_{\mu_{(K)}}$ (see Section S3 in the supplementary material)
      for $\tilde{l} = 1$ to L **do**
        specify the candidate bandwidth $\tilde{\eta}_{\mu_{(K)},\tilde{l}}$
        compute the index $j_{\tilde{l}(K)}$
        update the centroid $\phi_{\mu_{(K)},\tilde{l}}$ and the statistics $\tilde{P}_{\mu_{(K)},\tilde{l}}(t)$ and $\tilde{q}_{\mu_{(K)},\tilde{l}}(t)$
      Store $\phi_{\mu_{(K)},\tilde{l}}$, $\tilde{P}_{\mu_{(K)},\tilde{l}}(t)$ and $\tilde{q}_{\mu_{(K)},\tilde{l}}(t)$
      end for
  end for
  Output: $\tilde{\mu}^{(1)}_{(K)}(t) = \varepsilon_{3,2}^{\top}\left\{\tilde{P}_{\mu_{(K)},1}(t)\right\}^{-1}\tilde{q}_{\mu_{(K)},1}(t)$
end While

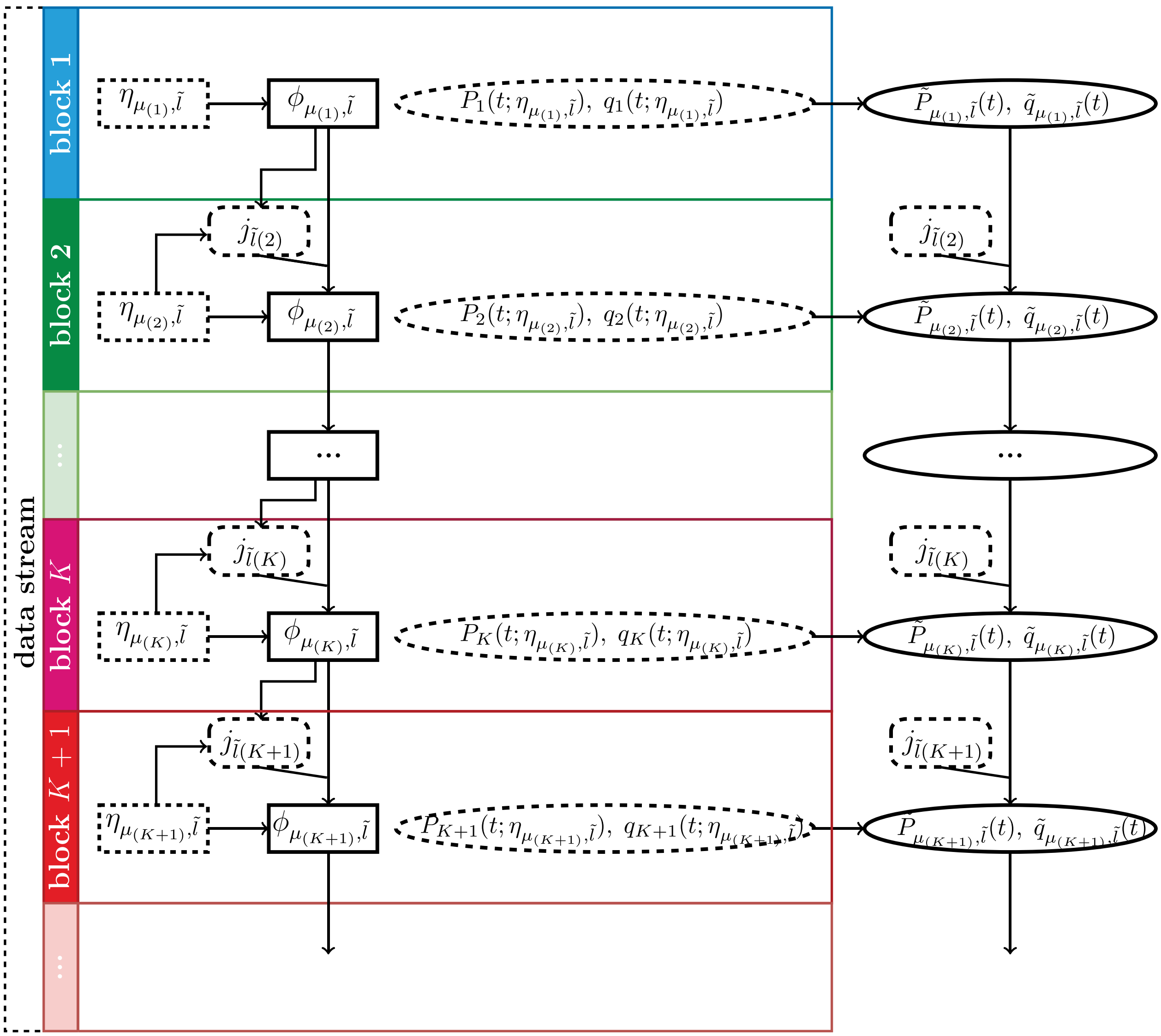

Figure 1. Process schematic of the proposed OCRL method for the mean function.

the bandwidth $\tilde{h}_{\gamma(K)}$, given by

$$P_k(s,t;\tilde{h}_{\gamma(K)}) = \sum_{i=1}^{n_k}\sum_{1\le j\neq l\le m_{ki}} W_{\tilde{h}_{\gamma(K)}}(T_{kij}-s) \times W_{\tilde{h}_{\gamma(K)}}(T_{kil}-t)\boldsymbol{T}_{ki,j,l}(s,t)\boldsymbol{T}_{ki,j,l}^{\top}(s,t),$$

$$q_k(s,t;\tilde{h}_{\gamma(K)}) = \sum_{i=1}^{n_k}\sum_{1\le j\neq l\le m_{ki}} W_{\tilde{h}_{\gamma(K)}}(T_{kij}-s) \times W_{\tilde{h}_{\gamma(K)}}(T_{kil}-t)\boldsymbol{T}_{ki,j,l}(s,t)\tilde{C}_{ki,j,l},$$

with $\boldsymbol{T}_{ki,j,l}(s,t) = (1, T_{kij}-s, T_{kil}-t, (T_{kij}-s)^2, (T_{kil}-t)^2, (T_{kij}-s)(T_{kil}-t))^{\top}$. The two-dimensional data introduce more serious challenges in estimating $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. As the complexity of the model increases and the data volume to be managed grows, the demands on the memory and the computational resources also rise, which can negatively impact the statistical efficiency.

To tackle these challenges, we use an idea similar to the online change rate estimation for the mean function. We generate a candidate bandwidth sequence for each block. Assuming that there are K data blocks, we specify a candidate bandwidth sequence $\{\eta_{\gamma(K),\tilde{l}}\}_{\tilde{l}=1}^{L}$, and set $\tilde{h}_{\gamma(K)} = \eta_{\gamma(K),1} > \cdots > \eta_{\gamma(K),L}$. When the $(K+1)$th data block arrives, we select the current online bandwidth $\tilde{h}_{\gamma(K+1)}$ and compute the candidates $\{\eta_{\gamma(K+1),\tilde{l}}\}_{\tilde{l}=1}^{L}$. The candidate bandwidth sequence is then updated online, and we obtain the centroids $\{\phi_{\gamma(K+1),\tilde{l}}\}_{\tilde{l}=1}^{L}$ as

$$\phi_{\gamma(K+1),\tilde{l}} = (1-\omega_{\gamma(K),K+1})\phi_{\gamma(K),j_{\tilde{l}(K+1)}} + \omega_{\gamma(K),K+1}\eta_{\gamma(K+1),\tilde{l}},$$

where the initial centroid $\phi_{\gamma(0)\tilde{l}} = 0$, which results in $\phi_{\gamma(1),\tilde{l}} = \eta_{\gamma(1),\tilde{l}}$, and the weight $\omega_{\gamma(K),K+1} = s_{K+1,2}/S_{K+1,2}$ with $S_{K+1,2} = \sum_{k=1}^{K+1} s_{k,2}$, $s_{k,2} = \sum_{i=1}^{n_k} m_{ki}(m_{ki}-1)$ and $s_{K+1,2} = \sum_{i=1}^{n_{K+1}} m_{K+1,i}(m_{K+1,i}-1)$. We solve the index $j_{\tilde{l}(K+1)}$ through the minimization problem:

$$j_{\tilde{l}(K+1)} = \arg\min_{i\in\{1,\ldots,L\}} \left|\eta_{\gamma(K+1),\tilde{l}} - \phi_{\gamma(K),i}\right|.$$

We then deduce the online-updated statistics $\{\tilde{P}_{\gamma(K+1),\tilde{l}}(s,t),\ \tilde{q}_{\gamma(K+1),\tilde{l}}(s,t)\}_{\tilde{l}=1}^{L}$ as

$$\tilde{P}_{\gamma(K+1),\tilde{l}}(s,t) = P_{K+1}(s,t;\eta_{\gamma(K+1),\tilde{l}}) + \tilde{P}_{\gamma(K),j_{\tilde{l}(K+1)}}(s,t),$$

Algorithm 2 OCRL for the covariance function

Initialization: $K \leftarrow 0$, $S_{K,2} \leftarrow 0$, $\phi_{\gamma(K),\tilde{l}} \leftarrow 0$, $\tilde{P}_{\gamma(K),\tilde{l}} \leftarrow 0$, $\tilde{q}_{\gamma(K),\tilde{l}} \leftarrow 0$, $\tilde{l} = 1, \ldots, L$
While $K < \infty$ **do**
 Input: $\{(T_{Kij}, Y_{Kij}) : i = 1, \ldots, n_K,\ j = 1, \ldots, m_{Ki}\}$
 for $K = K + 1$ **do**
  $S_{K,2} \leftarrow S_{K,2} + \sum_{i=1}^{n_K} m_{Ki}(m_{Ki} - 1)$
  compute the online bandwidth $\tilde{h}_{\gamma(K)}$ (see Section S3 in the supplementary material)
  for $\tilde{l} = 1$ to L **do**
   specify the candidate bandwidth $\tilde{\eta}_{\gamma(K),\tilde{l}}$
   compute the index $j_{\tilde{l}(K)}$
   update the centroid $\phi_{\gamma(K),\tilde{l}}$ and the statistics $\tilde{P}_{\gamma(K),\tilde{l}}(t)$ and $\tilde{q}_{\gamma(K),\tilde{l}}(t)$
   Store $\phi_{\gamma(K),\tilde{l}}$, $\tilde{P}_{\gamma(K),\tilde{l}}(t)$ and $\tilde{q}_{\gamma(K),\tilde{l}}(t)$
  end for
 end for
 Output: $\tilde{\gamma}_{(K)}^{(1,0)}(s,t) = \varepsilon_{6,2}^{\top}\left\{\tilde{P}_{\gamma(K),1}(t)\right\}^{-1}\tilde{q}_{\gamma(K),1}(t)$
end While

$$\tilde{q}_{\gamma(K+1),\tilde{l}}(s,t) = q_{K+1}(s,t;\eta_{\gamma(K+1),\tilde{l}}) + \tilde{q}_{\gamma(K),j_{\tilde{l}(K+1)}}(s,t),$$

where the initial values $\tilde{P}_{\gamma(0),\tilde{l}}(s,t) = 0$ and $\tilde{q}_{\gamma(0),\tilde{l}}(s,t) = 0$. Finally, the online estimator with the dynamically adjustable candidate bandwidths for $\gamma_{(K+1)}^{(1,0)}(s,t)$ is given by

$$\begin{aligned}\tilde{\gamma}_{(K+1)}^{(1,0)}(s,t) &= \varepsilon_{6,2}^{\top}\left\{\tilde{P}_{\gamma(K+1),1}(s,t)\right\}^{-1}\tilde{q}_{\gamma(K+1),1}(s,t)\\ &= \varepsilon_{6,2}^{\top}\left\{\sum_{k=1}^{K+1}P_k(s,t;\eta_{\gamma_{(k)},\cdot})\right\}^{-1}\left\{\sum_{k=1}^{K+1}q_k(s,t;\eta_{\gamma_{(k)},\cdot})\right\},\end{aligned} \tag{6}$$

where the dot index in $\eta_{\gamma(k),\cdot}$ is equal to $\tilde{l}$ when $k = K+1$ and for other values of $k = 1, \ldots, K$, it depends on $j_{\tilde{l}(k)}$. We also define the $K+1$ candidate bandwidths used in (6) as a pseudo-bandwidth sequence for the covariance function, denoted by $\{\tilde{\eta}_{\gamma(K+1),k}\}_1^{K+1}$, where $\tilde{\eta}_{\gamma(K+1),K+1} = \eta_{\gamma(K+1),1}$, $\tilde{\eta}_{\gamma(K+1),K} = \eta_{\gamma(K),j_{1(K+1)}}$, etc. We refer to the above estimation process as OCRL for the covariance function, and summarize the corresponding algorithm in Algorithm 2.

We briefly highlight the computational and memory efficiency of OCRL. Unlike the offline methods that repeatedly compute over the full data history, OCRL incrementally updates the sufficient statistics. When the $(K+1)$th block arrives, the update requires only $O(Ls_{K+1,d})$ operations, where L is fixed and typically small. As a result, the per-update cost remains constant, even as the dataset grows. In contrast, the offline kernel method costs $O(S_{K+1,d})$ operations per update and grows at least linearly with the cumulative data size. This shift from growing to constant per-update cost underscores the clear computational advantage of OCRL for streaming data. In addition, the memory footprint of OCRL is just $O(L)$, a substantial saving compared to $O(S_{K+1,d})$ as required in the offline methods to store all past data. These properties make OCRL highly scalable for real-time streaming data while maintaining statistical accuracy comparable to the offline methods.

3. Theoretical Results

In the theoretical development, we establish the asymptotic normality and evaluate the convergence properties of the online learning bandwidths, as well as compare the statistical efficiency of the proposed method with that of its offline counterpart. For ease of notation, we also define $\bar{m}_{j,K} = S_{K,j}/N_K$, where $S_{K,j} = \sum_{k=1}^{K} s_{k,j}$ with $s_{k,j} = \sum_{i=1}^{n_k} m_{ki}(m_{ki}-1)\cdots(m_{ki}-j+1)$ for $j \geq 1$, and several key functional terms that will be used in the asymptotic results, including $\alpha(W) = \int x^2 W(x)\,dx$, $\beta(W) = \int x^4 W(x)\,dx$, $R(W) = \int W^2(x)\,dx$ and $U(W) = \int x^2 W^2(x)\,dx$. Lastly, $\xrightarrow{d}$ represents the convergence in distribution.

3.1. Asymptotic Normality

We first establish the asymptotic normality of $\tilde{\mu}_{(K)}^{(1)}(t)$ in Theorem 1. The distributional properties of $\tilde{\mu}_{(K)}^{(1)}(t)$ are influenced by the average number of the measurements per subject $\bar{m}_{1,K}$, together with the total number of the subjects $N_K = \sum_{k=1}^{K} n_k$. Depending on the relative order of $\bar{m}_{1,K}$ and $N_K^{3/4}$, three distinct forms of the asymptotic normality emerge, corresponding to different sampling schemes.

Theorem 1. Suppose that Assumptions 1–3 in Section S1 of the supplementary material hold and $\limsup_K \bar{m}_{2,K}/\bar{m}_{1,K}^2 < \infty$. For a fixed interior point $t \in [\tilde{h}_{\mu(K)}, 1-\tilde{h}_{\mu(K)}]$, as $K \to \infty$, we have

$$\sqrt{\frac{S_{K,1}}{\Gamma_\mu(t)\rho_{\mu(K),-3}}}\left\{\tilde{\mu}_{(K)}^{(1)}(t) - \mu^{(1)}(t) - \frac{1}{6}\frac{\beta(W)}{\alpha(W)}\mu^{(3)}(t)\rho_{\mu(K),2}\right\} \xrightarrow{d} N(0,1),$$

where

$$\Gamma_\mu(t) = \frac{U(W)}{\alpha^2(W)}\frac{\gamma(t,t)+\sigma^2}{f(t)} + \frac{S_{K,2}}{S_{K,1}\rho_{\mu(K),-3}}\frac{1}{f^2(t)} \times\left[\gamma(t,t)\left\{f^{(1)}(t)\right\}^2 + \gamma^{(1,1)}(t,t)f^2(t) + 2\gamma^{(1,0)}(t,t)f(t)f^{(1)}(t)\right],$$

and $\rho_{\mu(K),\alpha} = \sum_{k=1}^{K}\omega_{\mu(K),k}\tilde{\eta}_{\mu(K),k}^{\alpha}$ with $\omega_{\mu(K),k} = s_{k,1}/S_{K,1}$ and $\tilde{\eta}_{\mu(K),k}$ being the pseudo-bandwidth chosen in the kth block for $\tilde{\mu}_{(K)}(t)$.

As shown in Theorem 1, the normalized asymptotic variance factor $\Gamma_\mu(t)$ involves the derivatives of the covariance function $\gamma^{(1,1)}(t,t)$ and $\gamma^{(1,0)}(t,t)$, which do not appear in the asymptotic theory for the mean function estimation. Their presence indicates that $\mu^{(1)}(t)$ is influenced not only by the marginal variance but also by the local dynamics of the covariance function. Furthermore, the second term in $\Gamma_\mu(t)$ satisfies $S_{K,2}/(S_{K,1}\rho_{\mu(K),-3}) = O(\bar{m}_{2,K}/\bar{m}_{1,K}^2(N_K\bar{m}_{1,K}^{-4/3})^{-3/7})$. This implies that the phase transition threshold is determined by the growth rate of $N_K\bar{m}_{1,K}^{-4/3}$. Three distinct types of asymptotic normality depend on the relative order of $\bar{m}_{1,K}$, with respect to $N_K^{3/4}$, and the order of $\tilde{h}_{\mu(K)}$. These results highlight the relationship between the sampling schemes, the bandwidth, and the estimator's performance.

Corollary 1. Suppose that the conditions in Theorem 1 hold.

(i) When $\bar{m}_{1,K}/N_K^{3/4} \to 0$ and $\tilde{h}_{\mu(K)} = O(S_{K,1}^{-1/7})$, we have

$$\Gamma_\mu(t) = \frac{U(W)}{\alpha^2(W)} \frac{\gamma(t,t)+\sigma^2}{f(t)}$$

and

$$\sqrt{\frac{S_{K,1}}{\Gamma_\mu(t)\rho_{\mu(K),-3}}} \left\{\tilde{\mu}_{(K)}^{(1)}(t) - \mu^{(1)}(t) - \frac{1}{6}\frac{\beta(W)}{\alpha(W)}\mu^{(3)}(t)\rho_{\mu(K),2}\right\} \xrightarrow{d} N(0,1).$$

(ii) When $\bar{m}_{1,K}/N_K^{3/4} \to C$ and $S_{K,2}\tilde{h}_{\mu(K)}^3/S_{K,1} \to C_1$ with $0 < C, C_1 < \infty$, we have

$$\Gamma_\mu(t) = \frac{U(W)}{\alpha^2(W)} \frac{\gamma(t,t)+\sigma^2}{f(t)} + C_1 \frac{1}{f^2(t)}\Big[\gamma(t,t)\left\{f^{(1)}(t)\right\}^2 + \gamma^{(1,1)}(t,t)f^2(t) + 2\gamma^{(1,0)}(t,t)f(t)f^{(1)}(t)\Big]$$

and

$$\sqrt{\frac{S_{K,1}^2}{\Gamma_\mu(t)S_{K,2}}} \left\{\tilde{\mu}_{(K)}^{(1)}(t) - \mu^{(1)}(t) - \frac{1}{6}\frac{\beta(W)}{\alpha(W)}\mu^{(3)}(t)\rho_{\mu(K),2}\right\} \xrightarrow{d} N(0,1).$$

(iii) When $\bar{m}_{1,K}/N_K^{3/4} \to \infty$, $\tilde{h}_{\mu(K)} = o(N_K^{-3/4})$, and $\bar{m}_{1,K}\tilde{h}_{\mu(K)}^3 \to \infty$, we have

$$\Gamma_\mu(t) = \frac{1}{f^2(t)}\Big[\gamma(t,t)\left\{f^{(1)}(t)\right\}^2 + \gamma^{(1,1)}(t,t)f^2(t) + 2\gamma^{(1,0)}(t,t)f(t)f^{(1)}(t)\Big]$$

and

$$\sqrt{\frac{S_{K,1}^2}{\Gamma_\mu(t)S_{K,2}}} \left\{\tilde{\mu}_{(K)}^{(1)}(t) - \mu^{(1)}(t)\right\} \xrightarrow{d} N(0,1).$$

Corollary 1 sheds light on three partitions of the asymptotic normal distribution, grounded in the degree of the data density. Drawing upon the partition criteria established by Zhang and Wang (2016) and Yang and Yao (2023), we classify the data into different types. Specifically, the data are first categorized as dense and non-dense, based on whether the convergence rate of the estimator can achieve $S_{K,1}^{-1/2}$. Within the dense category, the data are further subdivided into semi-dense and ultra-dense, depending on the presence or absence of bias terms of the estimator. Under this framework, Case (i) which achieves the optimal convergence rate of $S_{K,1}^{-2/7}$, corresponds to the non-dense functional data. This rate is comparable to that of the derivative estimation using a one-dimensional local quadratic smoother. Cases (ii) and (iii) correspond to the semi-dense and ultra-dense functional data, respectively. In both of the dense cases, the variance of $\tilde{\mu}_{(K)}^{(1)}(t)$ incorporates the derivative terms, complicating the analytical process. Therefore, in the context of the online change rate estimation, we categorize functional data into three types: non-dense, semi-dense and ultra-dense. This categorization parallels the framework used for the online mean function estimation in Yang and Yao (2023).

Next, we establish the asymptotic normality of $\tilde{\gamma}_{(K)}^{(1,0)}(s,t)$. We also define the following expectations, $E_1(s,t) = E[\{Y_1 - \mu(T_1)\}^2\{Y_2 - \mu(T_2)\}^2 | T_1 = s, T_2 = t]$, $E_2(s,t) = E[\{Y_1 - \mu(T_1)\}^2\{Y_2 - \mu(T_2)\}\{Y_3 - \mu(T_3)\}|T_1 = s, T_2 = t, T_3 = t]$, where $\{(T_i, Y_i) : i = 1,2,3\}$ are the observations for the same subject, and E_Φ denotes the expectation with respect to Φ.

Theorem 2. Suppose that Assumptions 1, 2, and 4 in Section S1 of the supplementary material hold, $\limsup_K \bar{m}_{4,K}/\bar{m}_{2,K}^2 < \infty$, $\limsup_K \bar{m}_{2,K}/\bar{m}_{1,K}^2 < \infty$, and $\limsup_K S_{K,1}S_{K,3}/S_{K,2}^2 < \infty$. For a fixed interior point $(s,t) \in [\tilde{h}_{\gamma(K)}, 1-\tilde{h}_{\gamma(K)}]^2$, as $K \to \infty$, we have

$$\sqrt{\frac{S_{K,2}}{\Gamma_\gamma(s,t)\rho_{\gamma(K),-4}}} \left\{\tilde{\gamma}_{(K)}^{(1,0)}(s,t) - \gamma^{(1,0)}(s,t) - \frac{1}{6}\frac{\beta(W)}{\alpha(W)} \times \gamma^{(3,0)}(s,t)\rho_{\gamma(K),2} - \frac{1}{2}\alpha(W)\gamma^{(1,2)}(s,t)\rho_{\gamma(K),2}\right\} \xrightarrow{d} N(0,1),$$

where

$$\Gamma_\gamma(s,t) = \{1 + I(s=t)\}\left\{\frac{R(W)U(W)}{\alpha^2(W)}\frac{E_1(s,t)}{f(s)f(t)} + \frac{S_{K,3}\rho_{\gamma(K),-3}}{S_{K,2}\rho_{\gamma(K),-4}}\frac{U(W)}{\alpha^2(W)}\frac{E_2(s,t)}{f(s)}\right\} + \frac{S_{K,4}}{S_{K,2}\rho_{\gamma(K),-4}} \times \frac{E_\Phi\left[\left\{\Phi(s)\Phi(t)f^{(1)}(s) + \Phi^{(1)}(s)\Phi(t)f(s)\right\}^2\right] - \left\{\gamma^{(1,0)}(s,t)f(s) + \gamma(s,t)f^{(1)}(s)\right\}^2}{f^2(s)},$$

and $\rho_{\gamma(K),\alpha} = \sum_{k=1}^K \omega_{\gamma(K),k}\tilde{\eta}_{\gamma(K),k}^\alpha$ with $\omega_{\gamma(K),k} = s_{k,2}/S_{K,2}$ and $\tilde{\eta}_{\gamma(K),k}$ being the pseudo-bandwidth chosen in the kth block for $\tilde{\gamma}_{(K)}(s,t)$.

Theorem 2 shows that with an appropriately chosen $\tilde{h}_{\gamma(K)}$, the proposed online change rate estimator of $\gamma^{(1,0)}(s,t)$ is asymptotically normally distributed. By employing the relative order, we establish a comprehensive framework to understand the behavior of the estimators under the varying data densities. Theorems 1 and 2 precisely characterize the asymptotic distribution of the change rate estimators in streaming functional data. Moreover, the partitioning of data into the distinct density regimes plays a pivotal role in practical applications, as the data density and the bandwidth selection critically influence the statistical efficiency of the estimators. In the subsequent simulation studies and real-world examples, we investigate both the dense and non-dense data to verify the excellent performance of the proposed online estimators under different sampling schemes. Similar to Corollary 1, we also deduce the corresponding result for $\tilde{\gamma}_{(K)}^{(1,0)}(s,t)$ in Section S2 of the supplementary material.

3.2. Online Bandwidth Selection

To select the bandwidth, we apply the mean integrated squared error (MISE) as a global loss measure to balance the bias and variance. By minimizing the MISEs, we can derive the optimal bandwidths as follows.

Corollary 2. Suppose that the assumptions for Theorems 1 and 2 hold. The MISEs for the online estimators (4) and (6) are

$$\mathrm{MISE}\left(\tilde{\mu}^{(1)}_{(K)}\right) = \frac{1}{36}\frac{\beta^2(W)}{\alpha^2(W)}\theta_\mu \rho^2_{\mu(K),2} + \frac{\nu_\mu}{S_{K,1}}\rho_{\mu(K),-3},$$
$$\mathrm{MISE}\left(\tilde{\gamma}^{(1,0)}_{(K)}\right) = \frac{1}{36}\frac{\beta^2(W)}{\alpha^2(W)}\theta_\gamma \rho^2_{\gamma(K),2} + \frac{\nu_\gamma}{S_{K,2}}\rho_{\gamma(K),-4},$$

where $\theta_\mu = \int \left\{\mu^{(3)}(t)\right\}^2 f(t)dt, \nu_\mu = \int \Gamma_\mu(t)f(t)dt, \theta_\gamma = \iint \{\gamma^{(3,0)}(s,t) + 3\gamma^{(1,2)}(s,t)\alpha^2(W)/\beta(W)\}^2 f(s)f(t)dsdt$ and $\nu_\gamma = \iint \Gamma_\gamma(s,t)f(s)f(t)dsdt$. Further by minimizing the MISEs, the optimal forms of $\tilde{h}_{\mu(K)}$ and $\tilde{h}_{\gamma(K)}$ are

$$\tilde{h}^{*opt}_{\mu(K)} = \left\{27\frac{\alpha^2(W)}{\beta^2(W)}\frac{\nu_\mu}{\theta_\mu}\right\}^{1/7} S_{K,1}^{-1/7} \quad \text{and}$$
$$\tilde{h}^{*opt}_{\gamma(K)} = \left\{27\frac{\alpha^2(W)}{\beta^2(W)}\frac{\nu_\gamma}{\theta_\gamma}\right\}^{1/8} S_{K,2}^{-1/8}. \tag{7}$$

In practice, however, $\theta_\mu, \theta_\gamma, \nu_\mu$, and ν_γ in (7) are unknown, making the theoretical bandwidths not directly accessible. To solve the problem, we develop a data-driven approach that substitutes these unknown quantities with their sample-based estimators,

$$\tilde{h}^{opt}_{\mu(K)} = \left\{27\frac{\alpha^2(W)}{\beta^2(W)}\frac{\tilde{\nu}_\mu}{\tilde{\theta}_\mu}\right\}^{1/7} S_{K,1}^{-1/7} \quad \text{and}$$
$$\tilde{h}^{opt}_{\gamma(K)} = \left\{27\frac{\alpha^2(W)}{\beta^2(W)}\frac{\tilde{\nu}_\gamma}{\tilde{\theta}_\gamma}\right\}^{1/8} S_{K,2}^{-1/8}, \tag{8}$$

where $\tilde{\theta}_\mu, \tilde{\theta}_\gamma, \tilde{\nu}_\mu$ and $\tilde{\nu}_\gamma$ are the pilot estimation obtained by an online method (see Section S3 in the supplementary material). The online learning bandwidths $\tilde{h}^{opt}_{\mu(K)}$ and $\tilde{h}^{opt}_{\gamma(K)}$ serve as the practical counterparts to the optimal bandwidths under the MISE criterion.

Theorem 3. Suppose that the assumptions for Theorems 1 and 2 hold. Then as $K \to \infty$,

(i) for the non-dense functional data, we have

$$\frac{\tilde{h}^{opt}_{\mu(K)} - \tilde{h}^{*opt}_{\mu(K)}}{\tilde{h}^{*opt}_{\mu(K)}} = O_p\left(S_{K,1}^{-2/5}\right) \quad \text{and}$$
$$\frac{\tilde{h}^{opt}_{\gamma(K)} - \tilde{h}^{*opt}_{\gamma(K)}}{\tilde{h}^{*opt}_{\gamma(K)}} = O_p\left(S_{K,2}^{-1/3}\right),$$

(ii) for the dense functional data, we have

$$\frac{\tilde{h}^{opt}_{\mu(K)} - \tilde{h}^{*opt}_{\mu(K)}}{\tilde{h}^{*opt}_{\mu(K)}} = O_p\left(S_{K,1}^{-2/9}\right) \quad \text{and}$$
$$\frac{\tilde{h}^{opt}_{\gamma(K)} - \tilde{h}^{*opt}_{\gamma(K)}}{\tilde{h}^{*opt}_{\gamma(K)}} = O_p\left(S_{K,2}^{-1/4}\right).$$

The convergence rates of the bandwidths differ between the dense and non-dense functional data. This discrepancy arises because, as indicated in Theorems 1 and 2, the variance of the estimators for the dense data involves higher-order derivatives, necessitating the use of local higher-order smoothers to for select the bandwidth. Theorem 3 demonstrates that the difference between the theoretical bandwidths in (7) and the online learning bandwidths in (8) gradually converges to 0 as the sample size increases, provided that the pilot online learning bandwidths and the candidate bandwidths are appropriately chosen. Furthermore, the convergence rates of the online learning bandwidths for the dense functional data are slower than those for the non-dense functional data, which reflects the additional complexity of the dense data regime.

3.3. Statistical Efficiency

To evaluate the proposed online method, we take the offline kernel method as a baseline for comparison. We obtain the offline estimators of $\mu^{(1)}(t)$ and $\gamma^{(1,0)}(s,t)$ using local quadratic smoothers based on the first K data blocks, denoted by $\hat{\mu}^{(1)}_{(K)}(t)$ and $\hat{\gamma}^{(1,0)}(s,t)$, for which we apply a hat to label the objects estimated by the offline kernel method. In the offline setting, the subscript (k) or (K) denotes that the statistics are based on the full data up to k or K, respectively. The bandwidths $\hat{h}_{\mu(K)}$ and $\hat{h}_{\gamma(K)}$ are selected using the offline kernel method. The MISEs of the offline method for estimating $\mu^{(1)}(t)$ and $\gamma^{(1,0)}(s,t)$ are expressed as

$$\mathrm{MISE}\left(\hat{\mu}^{(1)}_{(K)}\right) = \frac{1}{36}\frac{\beta^2(W)}{\alpha^2(W)}\theta_\mu \hat{h}^4_{\mu(K)} + \frac{\nu_\mu}{S_{K,1}}\hat{h}^{-3}_{\mu(K)},$$
$$\mathrm{MISE}\left(\hat{\gamma}^{(1,0)}_{(K)}\right) = \frac{1}{36}\frac{\beta^2(W)}{\alpha^2(W)}\theta_\gamma \hat{h}^4_{\gamma(K)} + \frac{\nu_\gamma}{S_{K,2}}\hat{h}^{-4}_{\gamma(K)}.$$

To quantify the efficiency of the proposed online change rate estimation relative to the offline change rate estimation, we define the relative efficiency as

$$\mathrm{reff}\left(\tilde{\mu}^{(1)}_{(K)}\right) = \frac{\mathrm{MISE}\left(\hat{\mu}^{(1)}_{(K)}\right)}{\mathrm{MISE}\left(\tilde{\mu}^{(1)}_{(K)}\right)} \quad \text{and}$$
$$\mathrm{reff}\left(\tilde{\gamma}^{(1,0)}_{(K)}\right) = \frac{\mathrm{MISE}\left(\hat{\gamma}^{(1,0)}_{(K)}\right)}{\mathrm{MISE}\left(\tilde{\gamma}^{(1,0)}_{(K)}\right)}. \tag{9}$$

These ratios measure how well the proposed OCRL method performs using the candidate bandwidths compared to the offline kernel approach. If the relative efficiency is less than but close to 1, this indicates that the performance of our online method achieves near-equivalent performance to the offline kernel method. The use of the candidate bandwidths does not cause a significant loss in accuracy, thus, demonstrating the statistical efficiency of the OCRL method.

Theorem 4. Suppose that Assumptions 1–5 in Section S1 of the supplementary material hold. If the candidate bandwidths are chosen as

$$\eta_{\mu(K),\tilde{l}} = \left(\frac{L-\tilde{l}+1}{L}\right)^{1/7} \tilde{h}^{opt}_{\mu(K)} \quad \text{and}$$
$$\eta_{\gamma(K),\tilde{l}} = \left(\frac{L-\tilde{l}+1}{L}\right)^{1/8} \tilde{h}^{opt}_{\gamma(K)}, \tag{10}$$

then for the online estimators (4) and (6), as $K \to \infty$, their relative efficiencies satisfy

$$\text{reff}\left(\tilde{\mu}^{(1)}_{(K)}\right) \geq \left(1 + 0.2596L^{-1} + 0.0030L^{-2}\right)^{-1} + O_p\left(S_{K,1}^{-2/9}\right),$$
$$\text{reff}\left(\tilde{\gamma}^{(1,0)}_{(K)}\right) \geq \left(1 + 0.2604L^{-1} + 0.0018L^{-2}\right)^{-1} + O_p\left(S_{K,2}^{-1/4}\right).$$

Theorem 4 shows that selecting the candidate bandwidths as specified in (10) ensures that each value in the pseudo-bandwidth sequences closely approximates the online learning bandwidths. This selection method effectively achieves high relative efficiency in the proposed estimators, and the constants appearing in the bounds are independent of the choice of kernel function. We illustrate the lower bounds of the relative efficiency for the proposed online change rate estimators $\tilde{\mu}^{(1)}_{(K)}(t)$ and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$ in Figure 2. As the length of the candidate bandwidth sequence L increases, the improvement rate in relative efficiency gradually diminishes. When $L = 5$, the relative efficiency of the proposed OCRL method surpasses 95% for both $\tilde{\mu}^{(1)}_{(K)}(t)$ and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. Even with a smaller $L = 3$, the relative efficiency still exceeds 92%. Therefore, achieving an optimal balance between statistical accuracy and computational time does not necessitate a large L. By effectively minimizing the computational expenses while preserving high relative efficiency, the candidate bandwidths ensure the practical feasibility of the proposed online estimation framework.

4. Simulation Studies

This section compares the proposed OCRL method with several representative alternatives, including the offline kernel method described in Section 3.3, the offline spline and wavelet methods, and a stochastic gradient descent (SGD) algorithm. To save space, we present only the comparison results with the offline kernel method in the main text; detailed results for the other three methods are provided in Section S4 of the supplementary material. Moreover, to facilitate the practical application, the code for implementing OCRL in simulations and real data applications is also publicly available on GitHub at *https://github.com/online-learn-ing/online-change-rate-learning-functional-data*. All analyses were conducted in R using a Hygon C86 7375 32-core Processor (2.0 GHz).

4.1. Set-Up

In model (2) with streaming functional data, we set the mean and covariance functions as

$$\mu(t) = 5\sin(2\pi t) \quad \text{and} \quad \gamma(s,t) = \sum_{m=1}^{4} \lambda_m \phi_m(s)\phi_m(t),$$

where $s,t \sim U(0,1)$, $\lambda_m = 1/(m+1)^2$ with $m = 1,\ldots,4$, $\phi_1(t) = \sqrt{2}\cos(2\pi t), \phi_2(t) = \sqrt{2}\sin(2\pi t), \phi_3(t) = \sqrt{2}\cos(4\pi t)$ and $\phi_4(t) = \sqrt{2}\sin(4\pi t)$. The stochastic part is $\Phi_{ki}(T_{kij}) = \sum_{m=1}^{4} \xi_m \phi_m(T_{kij})$, where ξ_m are independently drawn from $N(0,\lambda_m)$. The errors ϵ_{kij} are independent and identically distributed with $N(0, 0.5^2)$. We evaluate the method under both the dense and non-dense functional data. Specifically, two scenarios for the number of subjects n_k and measurements m_{ki} are considered. For the dense data, m_{ki} are generated from the normal distribution $N(25, 2^2)$ and $n_k = 3$; for the

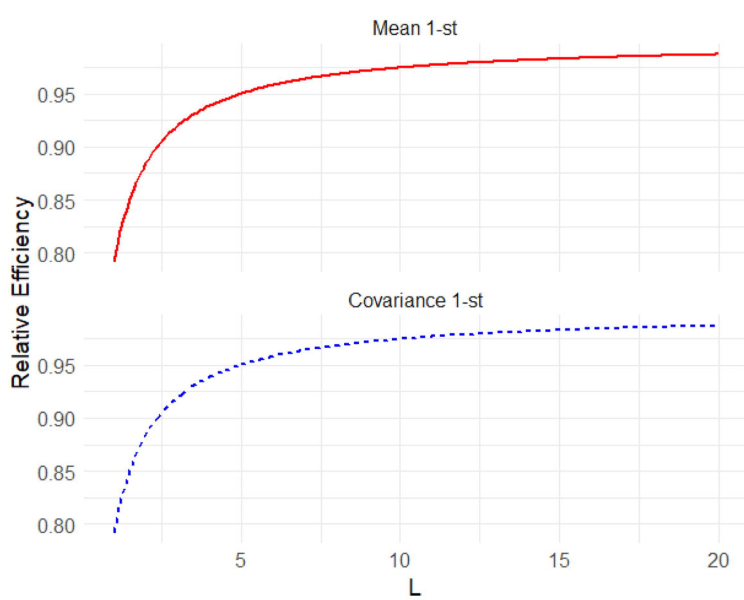

Figure 2. The lower bounds of relative efficiency for the proposed online method for estimating $\tilde{\mu}^{(1)}_{(K)}(t)$ (solid line) and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$ (dashed line).

non-dense data, m_{ki} are generated from the normal distribution $N(8, 2^2)$ and $n_k \sim N(18, 3^2)$. These numbers are rounded to the nearest integers.

In our OCRL implementation, we consider $L \in \{3, 5, 10, 20\}$, representing the length of the candidate bandwidth sequences for $\tilde{\mu}^{(1)}_{(K)}(t)$ and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. For the pilot estimation $\tilde{\theta}_\mu, \tilde{\theta}_\gamma, \tilde{\nu}_\mu, \tilde{\nu}_\gamma$ in (8), we set the parameter $G = 0.8^{1/d}$, where $d = 1$ for $\tilde{\mu}^{(1)}_{(K)}(t)$ and $d = 2$ for $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. A full sensitivity analysis of OCRL to G is shown in Figure S3 of the supplementary material. For the online estimation of $\gamma^{(1,0)}(s,t)$, we compute the online raw covariances based on the mean function estimates produced by OCRL with $L = 10$. Denote the length of candidate bandwidth sequence for the pilot estimation as J. For the computational convenience, we set $J = L$ for $\tilde{\mu}^{(1)}_{(K)}(t)$ and $J = 3$ for $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$. To further reduce the computational burden, we stop updating the pilot estimation of the online learning bandwidth for the covariance function when the data block exceeds 300. By this point, the pilot estimates have effectively stabilized, and further updates yield negligible gains; a justification of convergence is provided in Figure S4 of the supplementary material. Throughout the article, we use the Epanechnikov kernel. As in the classical nonparametric smoothing theory, the statistical performance is usually not sensitive to the specific kernel choice; therefore, similar results are expected for other standard kernels. For the bandwidth selection in the offline kernel method, please refer to Section S3 in the supplementary material. We perform the derivative estimates at every 20 blocks and simulate 100 replications with each streaming data consisting of $K = 1000$ data blocks. The capability of the proposed online change rate estimation is assessed from both the statistical and computational perspectives.

4.2. Statistical Efficiency

We first evaluate the statistical efficiency by examining the difference between the online and offline learning bandwidths. Figure 3 presents the boxplots of $(\tilde{h}^{opt}_{(K)} - \hat{h}^{opt}_{(K)})/\hat{h}^{opt}_{(K)}$ for $\tilde{\mu}^{(1)}_{(K)}(t)$ and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$ based on the first 100, 500, 1000 data blocks, where $\hat{h}^{opt}_{(K)}$ is the optimal form of $\hat{h}_{(K)}$ under the MISE criterion in practice. The differences between the online and offline learning bandwidths fluctuate slightly around 0, and thus the online learning bandwidths are extremely close to the offline counterparts. This indicates that the loss of the proposed OCRL method in efficiency is minimal. If we further use the candidate bandwidths, the performance of the proposed estimators do not degrade significantly. For the first-order derivative of the mean function under the dense data, the boxplots exhibit greater spread and are overall farther away from 0. This may lead to greater variability in the statistical efficiency in this case compared to the other three cases.

To visualize the dynamic updating process of the OCRL method, we further plot the pseudo-bandwidth sequence $\{\tilde{\eta}_{\mu_{(K)},k}\}_1^K$ with $L = 10$ under the dense sampling scheme based on the first 100, 500, 1000 data blocks in Figure 4. As more data become available, the proposed OCRL method dynamically picks up the candidate bandwidths that increasingly approximate the current bandwidth. This intuitively illustrates that the proposed OCRL method can effectively compete with the offline kernel method in terms of statistical efficiency. Similar results for $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$ and the non-dense sampling scheme are included in the supplementary material.

Next, we evaluate the estimation power of the proposed OCRL method using the relative efficiency, based on the empirical MISEs (eMISEs), as follows:

$$\mathrm{reff_e}\left(\tilde{\mu}^{(1)}_{(K)}\right) = \frac{\mathrm{eMISE}\left(\check{\mu}^{(1)}_{(K)}\right)}{\mathrm{eMISE}\left(\tilde{\mu}^{(1)}_{(K)}\right)} \quad \text{and}$$

$$\mathrm{reff_e}\left(\tilde{\gamma}^{(1,0)}_{(K)}\right) = \frac{\mathrm{eMISE}\left(\check{\gamma}^{(1,0)}_{(K)}\right)}{\mathrm{eMISE}\left(\tilde{\gamma}^{(1,0)}_{(K)}\right)},$$

where

$$\mathrm{eMISE}(\check{\mu}^{(1)}_{(K)}) = \frac{1}{100}\sum_{rep=1}^{100}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{m_{ki}}\left\{\check{\mu}^{(1)[rep]}_{(K)}(T_{kij}) - \mu^{(1)}(T_{kij})\right\}^2,$$

$$\mathrm{eMISE}(\tilde{\mu}^{(1)}_{(K)}) = \frac{1}{100}\sum_{rep=1}^{100}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{m_{ki}}\left\{\tilde{\mu}^{(1)[rep]}_{(K)}(T_{kij}) - \mu^{(1)}(T_{kij})\right\}^2,$$

$$\mathrm{eMISE}(\check{\gamma}^{(1,0)}_{(K)}) = \frac{1}{100}\sum_{rep=1}^{100}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{1\le j\ne l\le m_{ki}}\left\{\check{\gamma}^{(1,0)[rep]}_{(K)}(T_{kij}, T_{kil}) - \gamma^{(1,0)}(T_{kij}, T_{kil})\right\}^2,$$

$$\mathrm{eMISE}(\tilde{\gamma}^{(1,0)}_{(K)}) = \frac{1}{100}\sum_{rep=1}^{100}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{1\le j\ne l\le m_{ki}}\left\{\tilde{\gamma}^{(1,0)[rep]}_{(K)}(T_{kij}, T_{kil}) - \gamma^{(1,0)}(T_{kij}, T_{kil})\right\}^2.$$

Here, $\tilde{\mu}^{(1)[rep]}_{(K)}$ and $\tilde{\gamma}^{(1,0)[rep]}_{(K)}$ are the estimates obtained from OCRL, and $\check{\mu}^{(1)[rep]}_{(K)}$ and $\check{\gamma}^{(1,0)[rep]}_{(K)}$ are the estimates obtained from the competitors, including the three offline methods and the online method based on the SGD algorithm.

We compare the empirical relative efficiencies with the theoretical relative efficiencies lower bounds of OCRL versus the offline kernel method for $\tilde{\mu}^{(1)}_{(K)}(t)$ and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$ in Figure 5. Regardless of whether functional data are dense or not, or the number of the data blocks, the empirical relative efficiencies of the proposed OCRL method are consistently higher than the theoretical lower bounds under each candidate bandwidth sequence length. This highlights the strong statistical performance of our online method. As the number of the data blocks K grows beyond 300, the empirical relative efficiencies stabilize. The empirical relative efficiencies of $\tilde{\mu}^{(1)}_{(K)}(t)$ for the dense data exhibit significant peaks greater than 0 when the data block size is small. This indicates that the OCRL method can outperform the classical offline kernel method in handling $\mu^{(1)}_{(K)}(t)$ under the dense sampling, especially for small sample sizes. Compared to the spline, wavelet and SGD methods, the relative efficiencies

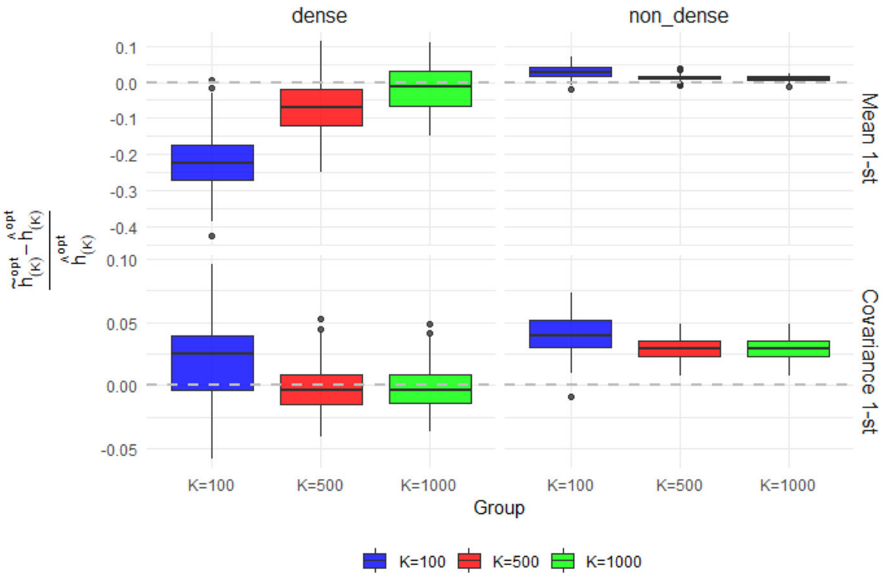

Figure 3. The boxplots of $(\tilde{h}^{opt}_{(K)} - \hat{h}^{opt}_{(K)})/\hat{h}^{opt}_{(K)}$ for $\mu^{(1)}_{(K)}(t)$ and $\gamma^{(1,0)}_{(K)}(s,t)$ under the dense and non-dense sampling schemes based on the first 100, 500, 1000 data blocks, where $\tilde{h}^{opt}_{(K)}$ are obtained using $L = 10$.

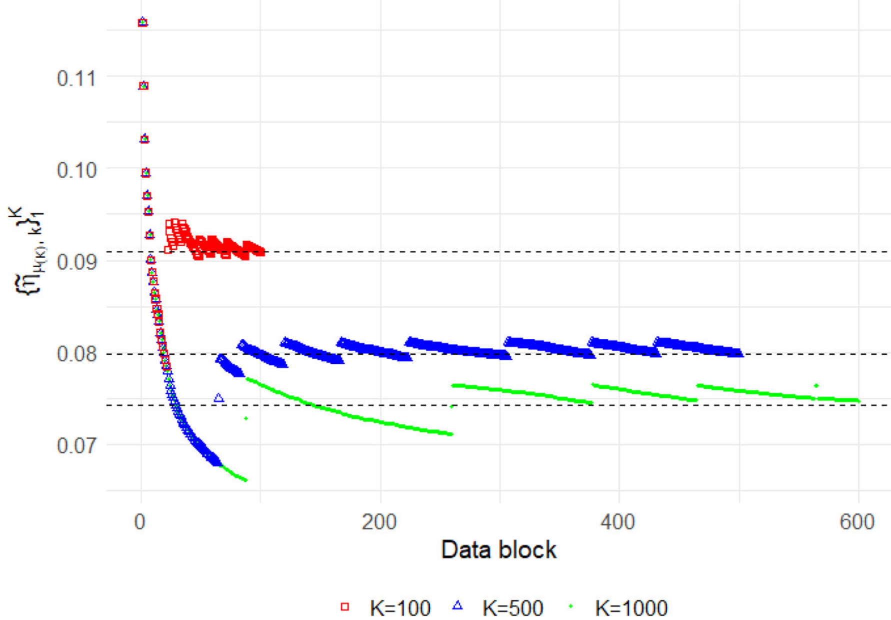

Figure 4. The pseudo-bandwidth sequence $\{\tilde{\eta}_{\mu_{(K)},k}\}_1^K$ with $L = 10$ under the dense sampling scheme based on the first 100, 500, 1000 data blocks.

on the logarithmic scale substantially exceed zero in almost all cases, which underscores the robustness and clear advantages of OCRL in the derivative estimation of streaming functional data. For the first derivative of the mean function $\mu^{(1)}(t)$, OCRL performs comparably to the spline method, achieving the relative efficiencies as high as 95% with $L = 5$. When estimating $\gamma^{(1,0)}(s,t)$, OCRL delivers superior performance relative to the spline method. In every scenario, OCRL outperforms both the wavelet and SGD methods. This superiority is most pronounced for estimating $\gamma^{(1,0)}(s,t)$, where the eMISEs of the wavelet method are roughly 15 times larger than those of OCRL and the eMISEs of SGD are larger by factors of nearly 40 or even 70 times. For more details, please refer to Figures S5–S7 in the supplementary material.

4.3. Computational Efficiency

To assess the computational time of the proposed OCRL method, we also compare it with the competitors. Figure 6 displays the average computational time in seconds over 100 replications for the proposed OCRL method with $L \in \{3, 5, 10, 20\}$ and the offline kernel method. The reported runtime of OCRL

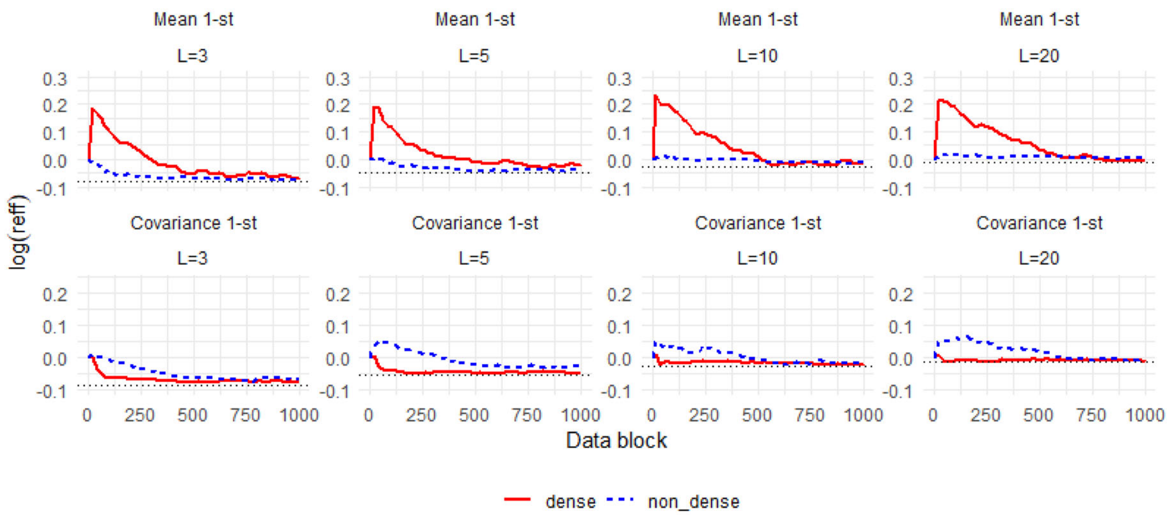

Figure 5. The empirical relative efficiencies and the theoretical relative efficiencies lower bounds (dotted lines) of OCRL versus the offline kernel method for $\tilde{\mu}^{(1)}_{(K)}(t)$ and $\tilde{\gamma}^{(1,0)}_{(K)}(s,t)$ under the dense (solid lines) and the non-dense data (dashed lines) with $L \in \{3, 5, 10, 20\}$. The y-axis is plotted on a logarithmic scale.

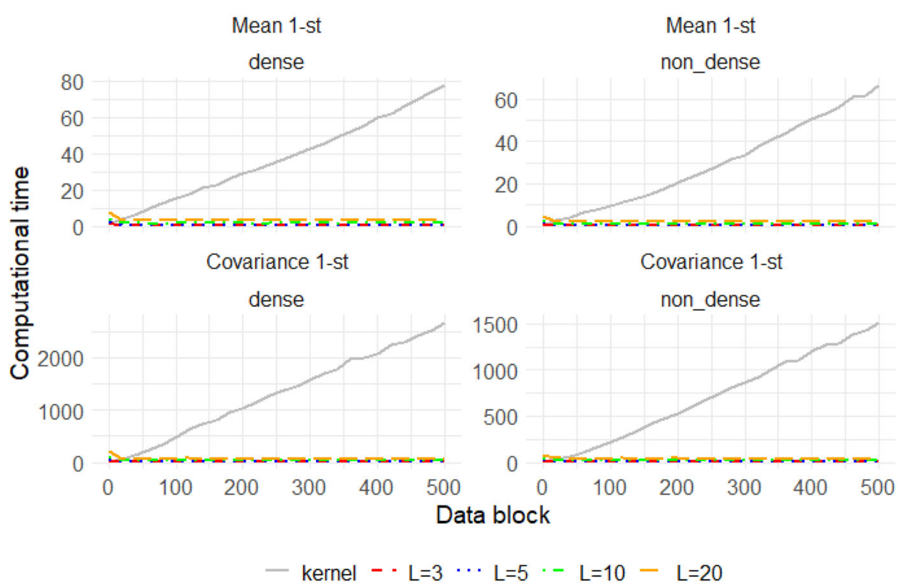

Figure 6. The comparison of the computational time in seconds between the proposed OCRL method and the offline kernel method for $\mu^{(1)}(t)$ and $\gamma^{(1,0)}(s,t)$ under the dense and non-dense sampling schemes.

corresponds to the cumulative computational time over all data blocks. The results demonstrate that our online method maintains a relatively constant computational time for the data blocks of similar sizes, showing the insensitivity to the accumulation of additional data blocks. There is no linear increase in time or memory required for a single update, underscoring its scalability and computational efficiency. In contrast, the offline kernel, spline and wavelet methods suffer dramatic slowdowns. Their runtime grows by factors ranging from tens to thousands when K is large. Our OCRL method demonstrates the significant computational advantage as the data volume grows. Although the computational speed of SGD is very fast, its estimates are unreliable. In contrast, OCRL achieves an effective balance between statistical accuracy and computational time. For more details, please refer to Figures S8–S10 in the supplementary material. The runtime of OCRL depends on the length of the candidate bandwidth sequence L. Theorem 4 indicates that a longer candidate bandwidth sequence improves the relative efficiency but increases the computational time. Based on the simulation results, we recommend the setting of $L = J = 10$ for $\mu^{(1)}_{(K)}(t)$ and $(L, J) = (5, 3)$ for $\gamma^{(1,0)}_{(K)}(s,t)$ in practical applications.

5. Application to Hourly Power Consumption Data

In this section, we apply the proposed OCRL method to analyze the derivative of the mean function for real-world functional data, demonstrating from a practical perspective that OCRL achieves high statistical efficiency with a satisfactory updating speed. By analyzing the hourly power consumption data, which is a critical issue in the energy sector, we reveal that the dynamic properties of this data are closely linked to human behavior patterns.

As global populations grow and economies develop, energy consumption is rapidly increasing. The power system, as a key element in energy conversion and distribution, plays a pivotal role in promoting sustainable development and enhancing quality of life. Studying the hourly power consumption not only helps optimize the power system operation, but also provides vital support for improving energy efficiency and facilitating the integration of renewable energy sources. We collect the hourly power consumption data in megawatts from the U.S. Energy Information Administration website *https://www.eia.gov/electricity*, covering 8 control zones from June 1, 2013 to August 2, 2018. We focus on the within-day trend of energy consumption between 0:00 and 23:00 and assume that the power consumption is independent for each day. Treating control zones as subjects and hourly consumption as measurements, we consider each day as a block, yielding $K = 1889$ data blocks and $n_k = 8$ subjects. For each subject, we design two measurement schemes: a dense scheme with $m_{ki} = 24$ and a non-dense scheme with $m_{ki} = 6$ subsampled from the dense scheme.

We use OCRL to estimate the change rate of the mean function, compare the results with the offline kernel method, and treat the offline kernel estimates based on all available data as the baseline. Table 1 reports the computational time for the proposed OCRL method versus the offline kernel method. For dense data, the offline kernel method takes at least several hundred times longer than OCRL; for non-dense data, it takes at least several dozen times longer. As the data block size increases, the runtime of the offline kernel method surges, whereas the online learning substantially saves the computational resources and reduces the computational complexity. We plot the first-order derivative estimation of the mean function at the end of year 2013, 2014, 2016, and 2018 in Figure 7. Regardless of whether the data block is small or large, OCRL achieves the statistical efficiency comparable to that of the offline kernel method. As the data stream accumulates, OCRL increasingly approximates the curves estimated from the full data. Our OCRL method not only achieves high computational efficiency but also demonstrates excellent estimation performance.

Next, we analyze the meaning implied by the change rate. The power consumption throughout the day exhibits a clear day-night pattern. Between 5:00 and 8:00, as residents wake up, go to work, and commercial activities begin, electricity consumption gradually increases at an accelerating rate. Subsequently,

Table 1. Comparison of computational time (in seconds) between the OCRL and offline kernel methods under different data blocks (K) and densities.

Time(s)	$K = 214$		$K = 579$		$K = 1310$		$K = 1889$	
	Dense	Non-dense	Dense	Non-dense	Dense	Non-dense	Dense	Bon-dense
OCRL	0.241	0.070	0.243	0.062	0.240	0.071	0.239	0.070
Offline	13.151	0.745	35.977	1.316	94.038	2.890	139.746	4.444
Offline/OCRL	54.568	10.643	148.053	21.226	391.825	40.704	584.711	63.486

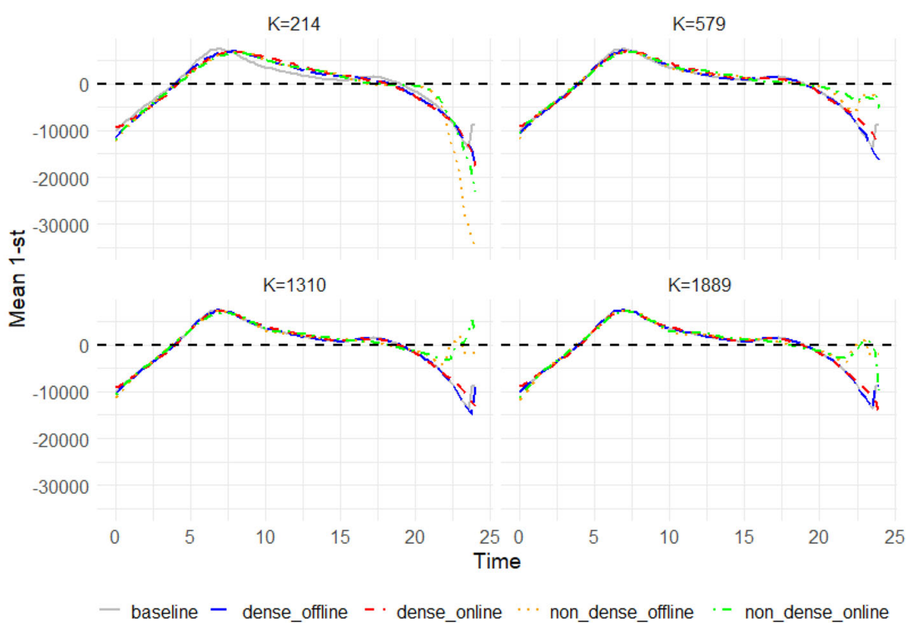

Figure 7. The first-order derivative estimation of the mean function using the proposed OCRL method and the offline kernel method based on the dense and non-dense cases of energy consumption data. The baseline, represented by the solid line, shows $\hat{\mu}_{(1)}(t)$ using the offline kernel method based on the full data from 2013 to 2018.

Table 2. The estimated temperature effects on the change rate by time period.

Time period	AIC	linear p-value	nonlinear p-value
0:00–6:00	462.800	3.381e-09**	0.004588*
6:00–12:00	427.230	0.002351*	1.000
12:00–18:00	391.680	<2.2e-16**	0.001336*
18:00–24:00	495.070	0.9905	0.0972

NOTE: The statistical significance of temperature effects is denoted by asterisks: ** $p \leq 0.001$ (highly significant), * $p \leq 0.05$ (significant). Values without asterisks indicate no statistical significance ($p > 0.05$).

from 12:00 to 18:00, the rate of increase in power consumption slows down and flattens out due to the continuation of working hours and the reduction of some commercial activities. However, as work hours end, household activities increase in the evening, including lighting, entertaining, cooking, etc., which cause power consumption to climb slightly again, creating a second small peak. Understanding these fluctuations in power consumption is crucial for the planning and scheduling of energy markets. To address this, we can implement time-of-day tariffs, promote power-saving technologies and equipment, and encourage users to adjust their consumption habits. These measures can help balance the electricity demand and reduce the load pressure during peak hours (Entso-E 2016).

Having established the day power consumption pattern, we now divide the 24 hr into four time periods to examine how the summer temperature influences the change rate in the power consumption over time. For each time period j, we employ the generalized additive model:

$$\mu_j^{(1)}(t) = \beta_0 + \beta_1 \cdot \text{Temperature}_t + f_j(\text{Temperature}_t) + \varepsilon_t,$$

where β_1 captures the linear temperature effect, and $f_j(\cdot)$ is the smoothing term capturing nonlinear temperature effect. Table 2 reports the Akaike Information Criterion (AIC) for assessing the goodness of fit for the model (lower values indicate better fit), along with p-values for the linear and nonlinear temperature effects. During 0:00 and 6:00, temperature exhibits a highly significant linear and a significant nonlinear association with the change rate, reflecting the impact of environmental control systems during the periods of low human activity. Between 6:00 and 12:00, the linear effect of temperature remains significant, while the nonlinear component is not evident, likely due to the increased human-driven electricity use. The effect of temperature is most pronounced from 12:00 to 18:00, as well as the lowest AIC among the time slices shows comparatively better fit. This variation is consistent with the peak outdoor temperatures and the heightened use of cooling equipment. In the evening, from 18:00 to 24:00, temperature shows no significant association with the change rate, indicating that electricity consumption during this period is primarily shaped by household routines rather than weather.

6. Conclusion and Discussion

The primary objective of this article was to recover the derivatives of the mean and covariance functions, which are crucial for capturing the underlying dynamic features and enabling the real-time analysis of streaming functional data. We proposed an online change rate learning (OCRL) method based on the kernel smoothing, designed to process streaming functional data with both the statistical and computational efficiency. By constructing and storing sufficient statistics with the candidate bandwidth sequence, OCRL efficiently approximates the full-data estimators without accessing the raw historical data. The model seamlessly adapts incoming data and updates the summary sufficient statistics of historical data, while maintaining the stable computational and memory costs. We further analyzed the theoretical properties of OCRL under certain regularity conditions, and showed by simulations that our new online method offers a superior tradeoff between statistical accuracy and computational time.

To demonstrate its practical value, we then applied our new OCRL method to the hourly power consumption data. The analysis revealed distinct human activity patterns, characterized by the pronounced morning and evening peaks, and quantified how temperature affects the change rates over time. As another compelling example, we also studied the U.S. traffic accident data, with details provided in Section S5 of the supplementary material. Treating each day as a block and each state as a subject, we log-transformed the hourly accident counts and analyzed both the dense and non-dense sampling regimes. Compared to the offline kernel method, OCRL delivered the real-time and interpretable change rate estimates that closely match the estimates based on the full data while reducing the computational time by at least several hundred times. The estimated change rates captured the key temporal dynamics, including a sharp increase in the accident rates during the early morning, as well as the notable peaks during the morning and evening commute hours and a smaller rise around lunchtime. These findings demonstrated the broad applicability and powerful practical value of OCRL.

While our study provides valuable insights, several avenues remain open for future work. First, the development of online spline smoothing or online wavelet decomposition is an important direction, and it requires breakthroughs in balancing the computational cost, memory usage and statistical accuracy. Incremental principal component analysis (IPCA) enables online updates of functional components, and exploring IPCA-based frameworks can expand the scope and enhance the utility of online FDA (Cardot and Degras 2018). Alternatively, the stochastic gradient descent method is statistically inefficient, calling for new algorithms that improve accuracy while maintaining low computational cost. Second, refining the online bandwidth selection procedures in OCRL will enhance the robustness (Wand and Jones 1994; Zhang and Wang 2016). Incorporating temporal and spatial correlation structures in recent second-generation FDA will broaden the applicability, but it requires addressing nontrivial theoretical and algorithmic challenges in the online setting (Koner and Staicu 2023). Furthermore, the proposed OCRL framework can be extended to the concept drift in streaming data by incorporating the forgetting mechanisms, such as the exponential forgetting factors or the sliding windows (Brzezinski and Stefanowski 2014). Last but not least, OCRL can also be applied to the derivative-based causal inference in econometrics, including contexts such as the regression kink design (Card et al. 2015) and the marginal treatment effect (Heckman and Vytlacil 2005). In addition, the dynamic candidate concept that underlies OCRL can be employed in streaming high-dimensional models, such as the online least absolute shrinkage and selection operator, enabling real-time updates of sparsity as new data arrive (Nanshan, Zhang, and Cao 2025).

Supplementary Materials

The supplementary materials provide the regularity assumptions, the corollary for the covariance function, the data-driven online bandwidth selection procedure, some extra simulations, an extended real data example and the technical results.

Acknowledgments

The authors sincerely thank the editor, the associate editor, and the reviewers for their constructive and insightful comments, which have greatly contributed to the significant improvement of this article. The authors are also grateful to Shandong University for providing the server access that supported the code simulations.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

WenWu Wang's work was supported by the Shandong Provincial Natural Science Foundation (No. ZR2024MA058), the Humanities and Social Sciences Fund of the Ministry of Education of China (No. 25YJA910006), and the National Natural Science Foundation of China (No. 12071248). Tiejun Tong's research was supported by the General Research Fund of Hong Kong (HKBU12300123).

References

Allen, R. M., Gasparini, P., Kamigaichi, O., and Böse, M. (2009), "The Status of Earthquake Early Warning Around the World: An Introductory Overview," *Seismological Research Letters*, 80, 682–693. [2]

Antoniadis, A., and Fan, J. (2001), "Regularization of Wavelet Approximations," *Journal of the American Statistical Association*, 96, 939–967. [1]

Brzezinski, D., and Stefanowski, J. (2014), "Combining Block-based and Online Methods in Learning Ensembles from Concept Drifting Data Streams," *Information Sciences*, 265, 50–67. [15]

Cao, G., Wang, J., Wang, L., and Todem, D. (2012), "Spline Confidence Bands for Functional Derivatives," *Journal of Statistical Planning and Inference*, 142, 1557–1570. [1]

Card, D., Lee, D. S., Pei, Z., and Weber, A. (2015), "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 83, 2453–2483. [15]

Cardot, H., and Degras, D. (2018), "Online Principal Component Analysis in High Dimension: Which Algorithm to Choose?" *International Statistical Review*, 86, 29–50. [14]

Dai, X., Müller, H. G., and Tao, W. (2018), "Derivative Principal Component Analysis for Representing the Time Dynamics of Longitudinal and Functional Data," *Statistica Sinica*, 28, 1583–1609. [1]

Entso-E, R. (2016), "Frequency Stability Evaluation Criteria for the Synchronous Zone of Continental Europe," Technical Report, Belgium. [2,14]

Ghale-Joogh, H. S., and Hosseini-Nasab, S. M. E. (2021), "On Mean Derivative Estimation of Longitudinal and Functional Data: From Sparse to Dense," *Statistical Papers*, 62, 2047–2066. [1]

Grith, M., Wagner, H., Härdle, W. K., and Kneip, A. (2018), "Functional Principal Component Analysis for Derivatives of Multivariate Curves," *Statistica Sinica*, 28, 2469–2496. [1]

Hall, P., Müller, H. G., and Yao, F. (2009), "Estimation of Functional Derivatives," *The Annals of Statistics*, 37, 3307–3329. [1]

He, G., Müller, H. G., and Wang, J.-L. (2000), "Extending Correlation and Regression from Multivariate to Functional Data," in *Asymptotics in Statistics and Probability*, pp. 197–210, Berlin: De Gruyter. [1]

Heckman, J. J., and Vytlacil, E. (2005), "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. [15]

Karuppusami, R., Antonisamy, B., and Premkumar, P. S. (2022), "Functional Principal Component Analysis for Identifying the Child Growth Pattern Using Longitudinal Birth Cohort Data," *BMC Medical Research Methodology*, 22, 76–85. [1]

Koner, S., and Staicu, A. M. (2023), "Second-Generation Functional Data," *Annual Review of Statistics and Its Application*, 10, 547–572. [14]

Kong, E., and Xia, Y. (2019), "On the Efficiency of Online Approach to Nonparametric Smoothing of Big Data," *Statistica Sinica*, 29, 185–201. [2]

Liu, B., and Müller, H. G. (2009), "Estimating Derivatives for Samples of Sparsely Observed Functions, with Application to Online Auction Dynamics," *Journal of the American Statistical Association*, 104, 704–717. [1]

Lu, J., Hoi, S. C., Wang, J., Zhao, P., and Liu, Z.-Y. (2016)"Large Scale Online Kernel Learning," *Journal of Machine Learning Research*, 17, 1–43. [2]

Mas, A., and Pumo, B. (2009), "Functional Linear Regression with Derivatives," *Journal of Nonparametric Statistics*, 21, 19–40. [1]

McKeague, I. W., Brown, A. S., Bao, Y., Hinkka-Yli-Salomäki, S., Huttunen, J., and Sourander, A. (2015), "Autism with Intellectual Disability Related to Dynamics of Head Circumference Growth During Early Infancy," *Biological Psychiatry*, 77, 833–840. [1]

Müller, H. G. (2005), "Functional Modelling and Classification of Longitudinal Data," *Scandinavian Journal of Statistics*, 32, 223–240. [1]

Nanshan, M., Zhang, N., and Cao, J. (2025), "Online Functional Principal Component Analysis on a Multidimensional Domain," arXiv preprint arXiv:2505.02131. [15]

Patilea, V., and Racine, J. S. (2024)"Locally Adaptive Online Functional Data Analysis," Department of Economics Working Papers 2024-04, McMaster University. [2]

Pigoli, D., and Sangalli, L. M. (2012), "Wavelets in Functional Data Analysis: Estimation of Multidimensional Curves and their Derivatives," *Computational Statistics & Data Analysis*, 56, 1482–1498. [1,2]

Rossi, F., and Villa-Vialaneix, N. (2011), "Consistency of Functional Learning Methods based on Derivatives," *Pattern Recognition Letters*, 32, 1197–1209. [1]

Sadeghioon, A. M., Metje, N., Chapman, D. N., and Anthony, C. J. (2014), "SmartPipes: Smart Wireless Sensor Networks for Leak Detection in Water Pipelines," *Journal of Sensor and Actuator Networks*, 3, 64–78. [1,2]

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009), "Efficient Estimation of Three-Dimensional Curves and their Derivatives by Free-Knot Regression Splines, Applied to the Analysis of Inner Carotid Artery Centrelines," *Journal of the Royal Statistical Society*, Series C, 58, 285–306. [1]

Simpkin, A. J., Durban, M., Lawlor, D. A., MacDonald-Wallis, C., May, M. T., Metcalfe, C., and Tilling, K. (2018), "Derivative Estimation for Longitudinal Data Analysis: Examining Features of Blood Pressure

Measured Repeatedly During Pregnancy," *Statistics in Medicine*, 37, 2836–2854. [1]

Treiber, M., Hennecke, A., and Helbing, D. (2000), "Congested Traffic States in Empirical Observations and Microscopic Simulations," *Physical Review E*, 62, 1805–1824. [2]

Wand, M. P., and Jones, M. C. (1994), *Kernel Smoothing*, New York: Chapman & Hall/CRC. [14]

Yang, Y., and Yao, F. (2023), "Online Estimation for Functional Data," *Journal of the American Statistical Association*, 118, 1630–1644. [2,7]

Zhang, J.-T., and Chen, J. (2007)"Statistical Inferences for Functional Data," *The Annals of Statistics*, 35, 1052–1079. [1]

Zhang, X., and Wang, J.-L. (2016), "From Sparse to Dense Functional Data and Beyond," *The Annals of Statistics*, 44, 2281–2321. [7,14]