Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: www.tandfonline.com/journals/ubes20

Nonparametric Causal Inference with Functional Covariates

Daisuke Kurisu, Taisuke Otsu & Mengshan Xu

To cite this article: Daisuke Kurisu, Taisuke Otsu & Mengshan Xu (20 Jun 2025): Nonparametric Causal Inference with Functional Covariates, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2025.2501563

To link to this article: https://doi.org/10.1080/07350015.2025.2501563

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2025, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/07350015.2025.2501563

Nonparametric Causal Inference with Functional Covariates

Daisuke Kurisu[a], Taisuke Otsu[b], and Mengshan Xu[c]

[a]Center for Spatial Information Science, The University of Tokyo, Chiba, Japan; [b]Department of Economics, London School of Economics, London, UK; [c]Department of Economics, University of Mannheim, Mannheim, Germany

ABSTRACT

Functional data and their analysis have become increasingly popular in various fields of data science. This article considers estimation and inference of the average treatment effect under unconfoundedness when the covariates involve a functional variable, and proposes the inverse probability weighting estimator, where the propensity score is estimated by using a kernel estimator for functional variables. We establish the $\sqrt{n}$-consistency and asymptotic normality of the proposed estimator. Numerical experiments and an empirical application demonstrate the usefulness of the proposed method.

ARTICLE HISTORY
Received February 2024
Accepted April 2025

KEYWORDS
Causal inference; Functional data; Semiparametric estimation

1. Introduction

Functional data and their statistical analysis have become increasingly popular in various fields of data sciences (Ramsey and Silverman 2005; Ferraty and Vieu 2006; Hsing and Eubank 2015; Kokoszka and Reimherr 2017). One of the major tools of functional data analysis is regression analysis using functional data, such as scalar-on-function regressions, functional response models, and their generalizations to generalized linear models, sparse models, or dependent observations. As in the conventional statistical analysis for Euclidean data, these functional regression methods are useful to describe the conditional means for a response given covariates.

In this article, we are concerned with estimation and inference of the average treatment effect (ATE) of a binary treatment when covariates contain some functional data. In the conventional setup where the vector of covariates is Euclidean, various estimation methods for the ATE are available, such as the inverse probability weighting (IPW), regression-based, and doubly-robust estimators (see Imbens and Rubin 2015, for a survey). A common feature of these estimation methods is that they are developed as functionals of preliminary estimators for the conditional mean and/or propensity score functions. Since researchers typically do not have enough information to specify the parametric forms of those functions, nonparametric approaches are commonly employed. When covariates contain some functional data, such nonparametric approaches are even more important due to complex structures of functional data.

In order to address this issue for estimation of the ATE with a functional covariate, this article advocates the kernel smoothing approach with functional covariates to estimate the propensity score or conditional mean outcome functions (see Ferraty and Vieu 2006, for an overview). In particular, we show that the IPW estimator with the plug-in kernel estimator for the propensity score function is consistent, asymptotically normal at the parametric rate, and asymptotically efficient. Although this result is analogous to the one for the case of Euclidean covariates (Hirano, Imbens, and Ridder 2003), its extension to the case of functional covariates is far from trivial. A major challenge is to develop a semiparametric theory for functionals of nonparametric kernel estimators with functional covariates. Ferraty et al. (2010) derived the uniform convergence rate of the kernel estimator with functional covariates, which is a building block for semiparametric theory. Boente and Vahnovan (2017) established consistency and asymptotic normality for the kernel-based estimator of the partially linear model. This article also makes a theoretical contribution to the literature on semiparametric theory with functional data.

This article extends the scope of causal inference methods to allow functional control variables. In the literature of causal inference, Miao, Xue, and Zhang (2020) studied estimation of the ATE with functional covariates. However, they employed parametric forms to specify the propensity score function, and did not present the theoretical properties of their parametric approach. In contrast, our method is nonparametric and is robust to misspecification of the propensity score function. A recent paper by Lin, Kong, and Wang (2023) investigated causal inference for functional outcomes with Euclidean covariates; on the other hand, this article considers the case of Euclidean outcomes with functional covariates. Finally, there are also some related yet distinctly different papers in the literature, such as Wong and Chan (2018); Zhao (2019), and Singh, Xu, and Gretton (2024), developed functional covariate balancing approaches using the reproducing kernel Hilbert space method for causal inference in observational studies. Wong and Chan (2018) and Zhao (2019) typically considered the case where functional objects are functions of Euclidean covariates; in comparison, our paper focuses on the case where the covariate itself is a functional variable. Singh, Xu, and Gretton (2024) covered a more general setup, where both the treatment and covariates

CONTACT Taisuke Otsu t.otsu@lse.ac.uk Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

may be infinite-dimensional, and they derived a (nonparametric) uniform convergence rate for their estimator of the causal response curve; in contrast, our paper is concerned with the binary treatment and $\sqrt{n}$-consistent estimation of the ATE.

This article is organized as follows. Section 2 provides a brief introduction to functional data and their applications in economics. Section 3 outlines our setup and the IPW estimator with a functional covariate. Section 4 presents our main theoretical results: Section 4.1 derives the uniform convergence rate for a kernel-weighted summation with a functional covariate, and Section 4.2 establishes the $\sqrt{n}$-consistency and asymptotic normality of the IPW estimator. In Sections 5 and 6, we conduct simulation studies and an empirical application to illustrate the usefulness of the proposed method. The proofs of our theoretical results, related lemmas, and additional simulation comparisons are contained in the supplementary material.

2. Functional Data and their Application in Economics

A functional variable is defined as a random variable taking values in an infinite dimensional space, and observations of functional variables are called functional data (Definition 1.1 of Ferraty and Vieu 2006). The functional data commonly emerge in different aspects of research areas: archaeologists use the shapes of bones—functional data captured from ancient skeletons—to study diseases that affected people in the distant past; criminologists use the trajectory through life of offending—functional data that record how criminal behavior develops and changes over an individual's lifetime—to study the formation and development of criminal behaviors; meteorologists use historical weather data—functional data often recorded at intervals of minutes and spanning many years—to analyze and forecast future weather trends. For the causal inference literature, functional regressors are applied in clinical trials (e.g., Ciarleglio et al. 2018; Zhao et al. 2018 for functional magnetic resonance imaging data). We refer to the textbook by Ramsey and Silverman (2005) for various applications of functional data across disciplines such as criminology, archaeology, psychology, neurophysiology, auxology, meteorology, biomechanics, education, etc.

Particularly, economic data collected over various time spans represents one of the most extensively analyzed types of functional data, including historical price indices from various goods and financial markets as well as key macroeconomic indicators such as GDP, trade volume, exchange rates, etc. (Ramsey and Silverman 2005, chap. 3) analyze the functional data of the monthly nondurable goods manufacturing for the United States, using a technique called the phase-plane plot. Using functional covariates, Florens and Van Bellegem (2015) study fertility rates over ages, and Benatia, Carrasco, and Florens (2017) study hourly temperature and electricity consumption. Also, in Section 5.2 of our paper, we conduct a simulation study on a market regulator model, using functional covariates sampled from real-world high-frequency data of the Standard & Poor's 500 Index over the past 17 years; in Section 6, we incorporate the monthly realizations of the federal funds effective rate as functional covariates to study the effects of monetary policy on economic indicators such as inflation, industrial production, and unemployment rate.

In the examples mentioned above, a typical functional dataset usually consists of a collection of curves. However, the functional data encompasses a much broader scope than curves, including vectors of curves (such as traces of handwriting), surfaces, arrays, images, or any other more complicated infinite-dimensional mathematical objects. A recent application in economics and finance by Jiang, Kelly, and Xiu (2023) adapts a statistical pattern recognition algorithm, using patterns of historical prices—rather than conventional time series of prices—as inputs for a convolutional neural network to predict future returns. They show that their methods can yield more accurate predictions of returns and provide more profitable investment strategies.

To motivate our approach using functional data (say, $\{X_i\}_{i=1}^n$), it is insightful to compare with the conventional approach[1] using its discretized version (say, $\{X_i^{\text{disc}}\}_{i=1}^n$ with $X_i^{\text{disc}} \in \mathbb{R}^d$).[2] First, we cite a textbook explanation by (Ferraty and Vieu 2006, Chapter 1.3, p. 7):

Indeed, if for instance we consider a sample of finely discretized curves, two crucial statistical problems appear. The first comes from the ratio between the size of the sample and the number of variables (each real variable corresponding to one discretized point). The second, is due to the existence of strong correlations between the variables and becomes an ill-conditioned problem in the context of multivariate linear model. So, there is a real necessity to develop statistical methods/models in order to take into account the functional structure of this kind of data.

In fact, the cited paragraph indicates three problems associated with discretizing functional data into finite-dimensional vectors and then applying conventional statistical approaches: (i) high-dimensional covariates, (ii) strong correlations among the covariates resulting from the discretization, and (iii) ignorance of the functional structure of the data. In the following, we will discuss these three problems in sequence.

Related to point (i), we further note that the researcher needs to choose the dimension of d for discretization, which is a highly nontrivial issue particularly for nonparametric regression using X_i^{disc} as (typically high-dimensional) covariates. If the chosen d is too small, it will inevitably lead to the loss of data information. Conversely, if d is too large, the nonparametric method will suffer from the curse of dimensionality. Moreover, if d is chosen to be higher than the sample size, many regular estimation approaches, such as linear and nonlinear least squares methods, cannot be implemented.

To illustrate point (ii), we reproduce (Ferraty and Vieu 2006, Figure 2.8) and display 28 curves representing yearly differentiated log electricity consumption from 1973 to 2001 in Figure 1:

[1] Here, we refer to the statistical approaches that obtain their estimators by regressing a vector of dependent variables on a vector of finite-dimensional covariates as the "conventional approach". This includes various types of common linear, nonlinear, and nonparametric estimators.

[2] Note that there is a distinction between the discretization used to conduct the conventional approach and the discretization as a practical implementation for the functional approach. In practice, functional data are usually collected as samples of discretized functional objects since inputting an infinite amount of data into a computer is technically infeasible in most cases. In this section, we focus on the first type of discretization, and leave the investigation of the second type to Appendix B of the supplementary material.

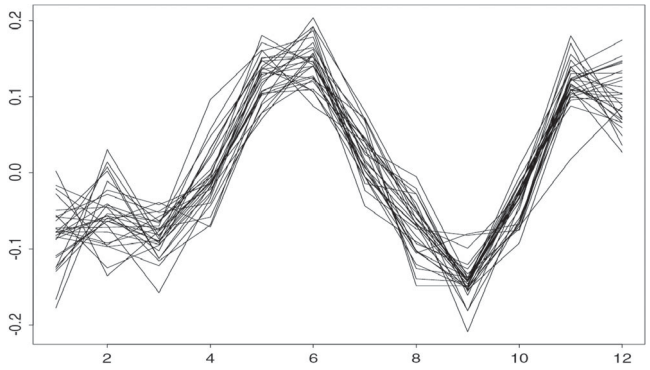

Figure 1. Yearly differentiated log electricity consumption from 1973 to 2001 (reproduced from Fig. 2.8 of Ferraty and Vieu 2006).

We observe that all curves follow a similar seasonal pattern in electricity consumption: increasing during summer, declining through autumn, and rising once more toward year's end. While each curve can be represented as a high-dimensional vector in the sample, we observe significant co-movement among curves from different years, indicating high correlations between random values from the same curve. Consequently, a conventional regression method, which uses these highly correlated covariates as direct inputs, is likely to experience a severe loss of estimation efficiency.

Now we move to point (iii). As illustrated by Geenens (2011), an issue especially important for nonparametric regression is that conventional regression methods based on discretized vectors transformed from functional covariates typically fail to consider the functional structure of this type of data, resulting in inefficient exploration of the information embedded in infinite-dimensional functional data. To see concretely this point, consider a nonparametric classification problem for discriminating forgeries from genuine signatures as in Geenens (2011). In this example, the dependent variable D_i is the indicator for genuine signatures, and $X_i(t) = (X_{1i}(t), X_{2i}(t)) \in \mathbb{R}^2$ represents the position of the pen at time t. The discretized covariates $X_i^{\text{disc}} = \{(X_{1i}(t_m), X_{2i}(t_m))\}_{m=1}^d$ is constructed by the grid $(t_1, \ldots, t_d)$. To classify genuine signatures, we can estimate $\mathbb{P}(D_i = 1 | \|X_i - x\| = 0)$ for the functional covariate or $\mathbb{P}(D_i = 1 | X_i^{\text{disc}} = x^{\text{disc}})$ for the Euclidean covariates resulting from discretizing functional covariates, where $\|\cdot\|$ is a semi-norm formally discussed in the next section. One appealing feature of the functional data approach is that the researcher can choose the seminorm $\|\cdot\|$ to measure proximity of two signatures based on the functional structure. For instance, one may set

$$\|X_1 - X_2\| = \sqrt{\int (X_1''(t) - X_2''(t))^2 dt},$$

where $X_i''(t)$ is the tangential projection of the vector of second derivatives with respect to time of $X_i(t)$. Thus, this semi-norm would account for the similarity in tangential pen acceleration between two signing processes, which is typically determined by the movement of the wrist (a feature commonly accepted as very hard to reproduce by skilled forgers). On the other hand, the conventional approach trying to estimate $\mathbb{P}(D_i = 1 | X_i^{\text{disc}} = x^{\text{disc}})$ may exhibit rather different patterns and overlook key features embedded in the functional data, such as wrist movement.

It should be noted the above distinction between two approaches is not only practically useful but also theoretically important. Note that the convergence rates of the nonparametric estimators are typically determined by decay rates of the probability $\mathbb{P}(\|X_i - x\| \le h)$ for the functional data approach or $\mathbb{P}(\|X_i^{\text{disc}} - x^{\text{disc}}\|_E \le h^{\text{disc}})$ for the discretized Euclidean data approach, where h and h^{disc} are bandwidths shrinking to zero. Then even though $\mathbb{P}(\|X_i^{\text{disc}} - x^{\text{disc}}\|_E \le h^{\text{disc}})$ decays at a very fast rate with a moderate or large value of d, the probability $\mathbb{P}(\|X_i - x\| \le h)$ with appropriately chosen $\|\cdot\|$ may decay much slower so that one may achieve a sufficiently fast convergence rate for the nonparametric estimator of $\mathbb{P}(D_i = 1 | \|X_i - x\| = 0)$ by the functional data approach.

Overall, all these discussions and examples demonstrate that the introduction of functional data analysis in economics is not only theoretically important but also holds substantial potential for practical applications.

3. Setup and Estimator

Let us introduce the basic setup. For each unit $i = 1, \ldots, n$, we observe an indicator variable D_i for a treatment ($D_i = 1$ if treated and $D_i = 0$ otherwise), and an outcome variable

$$Y_i = \begin{cases} Y_i(0) & \text{if } D_i = 0 \\ Y_i(1) & \text{if } D_i = 1 \end{cases},$$

where $Y_i(0)$ and $Y_i(1)$ are potential outcomes for $D_i = 0$ and 1, respectively. This article is concerned with the situation where in addition to Euclidean covariates $Z_i \in \mathbb{R}^k$, the researcher observes a functional covariate X_i whose support $\mathcal{X}$ is a subset of a semi-metric space $\mathcal{M}$ with a semi-norm. Let $\|\cdot\|$ be a semi-norm for $\mathcal{X}$. Based on the sample $\{D_i, Y_i, X_i, Z_i\}_{i=1}^n$ of size n, we wish to conduct inference on the average treatment effect (ATE) $\theta = \mathbb{E}[Y_i(1) - Y_i(0)]$. To identify θ, we impose the following conventional assumption. Let $p(x, z) = \mathbb{P}\{D_i = 1 | \|X_i - x\| = 0, Z_i = z\}$ be the propensity score. Note that similar to the conditioning of the event $Z_i = z$ for (continuous) Euclidean covariates Z_i, the event $\|X_i - x\| = 0$ for the functional covariate X_i is typically a null event.

Assumption 1. There exists a positive constant $c \in (0, 1/2)$ such that $p(x, z) \in [c, 1 - c]$ for each $x \in \mathcal{X}$ and $z \in \mathbb{R}^k$. D_i and $\{Y_i(0), Y_i(1)\}$ are conditionally independent given (X_i, Z_i).

Under this assumption, the ATE θ can be identified as

$$\theta = \mathbb{E}\left[\frac{D_i Y_i}{p(X_i, Z_i)} - \frac{(1 - D_i) Y_i}{1 - p(X_i, Z_i)}\right]. \tag{3.1}$$

In order to estimate θ, we first estimate the propensity score $p(X_i, Z_i)$ by kernel smoothing, that is,

$$\hat{p}(X_i, Z_i) = \frac{\sum_{j \neq i}^n D_j K(\|X_j - X_i\|/h) K_z((Z_j - Z_i)/h_z)}{\sum_{j \neq i}^n K(\|X_j - X_i\|/h) K_z((Z_j - Z_i)/h_z)}, \tag{3.2}$$

where $K : [0, \infty) \to \mathbb{R}$ and $K_z : \mathbb{R}^k \to \mathbb{R}$ are kernel functions, and h and h_z are bandwidths. It should be noted that the propensity score $p(x, z)$ is defined as the conditional probability given that $\|X_i - x\| = 0$ by using the semi-norm $\|\cdot\|$, and its estimator $\hat{p}(x, z)$ is also defined based on the kernel weights

with respect to this semi-norm. Kernel localization by using a semi-norm is commonly applied in nonparametric methods with functional covariates (see Ferraty and Vieu 2006, for an overview). As presented in Remark 1 and numerical illustrations in Section 5, a large part of explanatory tools for functional data consists in displaying them in low-dimensional spaces, so allowing for a semi-norm for kernel smoothing is important in our setup. An intuitive example of semi-norms for the Euclidean space $\mathbb{R}^d$ is a norm for a lower dimensional Euclidean subspace of $\mathbb{R}^d$.

Based on this nonparametric estimator, the inverse probability weighting estimator for θ can be estimated by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{D_i Y_i}{\hat{p}(X_i, Z_i)} - \frac{(1-D_i)Y_i}{1-\hat{p}(X_i, Z_i)}\right]. \tag{3.3}$$

Although we focus on the Nadaraya-Watson or local constant estimator $\hat{p}(X_i, Z_i)$ to estimate the propensity score, it is possible to extend our asymptotic analysis to the local linear-type estimator as in Ferraty and Nagy (2022). However, it is beyond the scope of this article to study whether our asymptotic theory can be adapted to other estimation methods, such as sieve-based methods.

Since this is the first paper that introduces the kernel smoothing approach to estimate the ATE with a functional covariate, we present our theoretical results for the inverse probability weighting (IPW) estimator $\hat{\theta}$ in (3.3) in the next section. Although it is beyond the scope of this article, it is of interest to study alternative estimation methods for the ATE, such as the nonparametric regression-based and doubly robust estimators as in Farrell (2015) for the case of Euclidean covariates (see Remark 4 for a further discussion).

4. Asymptotic Theory

4.1. Uniform Convergence Rate of Kernel Weighted Sums

Hereafter to simplify the presentation, we focus on the IPW estimator $\hat{\theta}$ with no Euclidean covariates. In particular, we study the asymptotic properties of the IPW estimator with no Euclidean covariates

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{D_i Y_i}{\hat{p}(X_i)} - \frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}\right],$$

where

$$\hat{p}(X_i) = \frac{\sum_{j\neq i}^{n} D_j K(\|X_j - X_i\|/h)}{\sum_{j\neq i}^{n} K(\|X_j - X_i\|/h)}. \tag{4.1}$$

To this end, as a building block, we first establish the uniform convergence rate of the following general weighted summation:

$$\hat{\Psi}(x) = \frac{1}{n\varphi(h)}\sum_{i=1}^{n} K(\|X_i - x\|/h) W_i, \tag{4.2}$$

where $\{W_i\}_{i=1}^n$ is a sequence of scalar random variables of $W \in \mathbb{R}$, and $\varphi(h)$ is defined in Assumption 2(ii). Intuitively, $n\varphi(h)$ can be viewed as the effective sample size for nonparametric estimation. For example, the Nadaraya-Watson estimator for the conditional mean or propensity score can be represented by the ratio of the weighted summations in the form of (4.2). The uniform convergence rate of $\hat{\Psi}(x)$ was studied by Ferraty et al. (2010). In order to achieve the $\sqrt{n}$-consistency for the semiparametric object $\hat{\theta}$, we need to establish a sufficiently fast uniform convergence rate for $\hat{\Psi}(x)$ over $\mathcal{X}$ (typically $o_p(n^{-1/4})$). The treatment on the stochastic part is an adapted version of Ferraty et al. (2010) to our focus on causal inference, and the analysis on the bias part using higher-order kernels is new in the literature.

Let $\mathcal{B}(x,h) = \{y \in \mathcal{X} : \|x-y\| \leq h\}$ be a ball centered around $x \in \mathcal{X}$ with radius h and $F(x,h) = \mathbb{P}\{X \in \mathcal{B}(x,h)\}$ be the small ball probability of X. $F(x,\cdot)$ may be interpreted as the cumulative distribution function of $\|x - X\|$ for given x. For $\varepsilon > 0$, a finite set $\mathcal{G} \subset \mathcal{F}$ is called an ε-covering of $\mathcal{F}$ with respect to $\|\cdot\|$ if for any $f \in \mathcal{F}$ there exists $g \in \mathcal{G}$ such that $\|f - g\| \leq \varepsilon$. The minimum cardinality of a ε-covering of $\mathcal{F}$ with respect to $\|\cdot\|$ is called the covering number of $\mathcal{F}$ with respect to $\|\cdot\|$ and denoted by $\mathcal{N}(\mathcal{F}, \|\cdot\|, \varepsilon)$. Also for any positive sequences $\{c_{1n}\}$ and $\{c_{2n}\}$, $c_{1n} \sim c_{2n}$ means $c_{1n}/c_{2n} \to C$ for some $C \in (-\infty, \infty)$ as $n \to \infty$. To derive the uniform convergence rate of the stochastic part $\hat{\Psi}(x) - \mathbb{E}[\hat{\Psi}(x)]$, we impose the following assumptions.

Assumption 2. (i) $\{D_i, Y_i, X_i\}_{i=1}^n$ is an iid sample of $(D, Y, X) \in \{0,1\} \times \mathbb{R} \times \mathcal{X}$. The ε-covering number of $\mathcal{X}$ satisfies $\log \mathcal{N}(\mathcal{X}, \|\cdot\|, \varepsilon) = O((\log(1/\varepsilon))^\eta)$ for some $\eta > 0$.

(ii) There exist positive constants C_φ, $c_F < C_F$, and $c_f < C_f$, an integer $\alpha \geq 2$, and a nonnegative function $f(x)$ on $x \in \mathcal{X}$ such that $\varphi(h) = C_\varphi h^\alpha$,

$$\begin{aligned} &c_F \varphi(h) f(x) \leq F(x,h) \leq C_F \varphi(h) f(x), \\ &c_f \leq \inf_{x\in\mathcal{X}} f(x) \leq \sup_{x\in\mathcal{X}} f(x) \leq C_f, \end{aligned} \tag{4.3}$$

hold for all $x \in \mathcal{X}$.

(iii) $\sup_{x\in\mathcal{X}} \mathbb{E}[|Y|^\zeta \,|\, \|X - x\| = 0] \leq C_Y$ for some $\zeta > 2$ and $C_Y < \infty$.

(iv) $K : [0,\infty) \to \mathbb{R}$ has support on $[0,1]$ and is continuously differentiable on $(0,1)$. Further, K is Lipschitz on $[0,\infty)$, that is, $|K(v) - K(v_1)| \leq C_K |v - v_1|$ for some $C_K < \infty$ and all $v, v_1 \in [0,\infty)$, and it holds that $\int K(z)dz = 1$ and

$$\begin{aligned} &\int z^\ell K(z)dz = 0 \quad \text{for } \ell = 1, \ldots, \alpha-2, \alpha, \ldots, q-1, \\ &\int z^{\alpha-1} K(z)dz \neq 0, \quad \int z^q K(z)dz \neq 0, \end{aligned}$$

for some positive integer $q \geq \alpha + 1$. Furthermore, $F(x,s)$ is differentiable in s, and

$$\begin{aligned} &\int K(s/h)\left\{\frac{dF(x,s)}{ds}\right\} ds \sim \bar{f}(x)\int K(s/h)\left\{\frac{d\varphi(s)}{ds}\right\} ds, \\ &\int K^2(z)\left\{\frac{dF(x,zh)}{dz}\right\} dz \sim \bar{f}(x)\int K^2(z)\left\{\frac{d}{dz}\varphi(zh)\right\} dz, \end{aligned} \tag{4.4}$$

where $\bar{f}(x)$ is a nonnegative functional on $x \in \mathcal{X}$ with $0 < c_{\bar{f}} \leq \inf_{x\in\mathcal{X}} \bar{f}(x) \leq \sup_{x\in\mathcal{X}} \bar{f}(x) \leq C_{\bar{f}} < \infty$ for some constants $c_{\bar{f}}$ and $C_{\bar{f}}$.

(v) As $n \to \infty$, $h \to 0$ and $n\varphi(h)(\log n)^{-\eta} \to \infty$ where η is the constant in Condition (i).

Assumption 2 (i) is on the data.[3] Although we focus on the independent data, it can be extended to allow weakly dependent data under certain mixing conditions as in Ferraty and Vieu (2006, chap. 11). The condition on the covering number is a key to establishing the uniform convergence rate as in Ferraty et al. (2010). This assumption covers: a compact subset of a separable Hilbert space with a projection semi-norm ($\eta = 1$) (see Remark 1), and the unit ball of the Cameron-Martin space with the supremum norm ($\eta = 2$) (see, van der Vaart and van Zanten 2007).[4]

Note that in contrast to the Euclidean case (as in Li, Racine, and Wooldridge 2009), we assume the total boundedness of $\mathcal{X}$. Indeed for the uniform convergence result in Proposition 1, the total boundedness of $\mathcal{X}$ imposed in Assumption 2(i) is sufficient. Note that if the notion of total boundedness is defined for a given metric space, then every compact set is totally bounded and every totally bounded set is bounded.

Assumption 2(ii) is on the small ball probability $F(x,h)$ whose decay is characterized by a polynomial $\varphi(h) = C_\varphi h^\alpha$ of the bandwidth h. We emphasize that the kernel estimator $\hat{p}(X_i, Z_i)$ for the propensity score and the inverse probability weighting estimator $\hat{\theta}$ in (3.3) for the ATE do not involve $\varphi(h)$, and thus the researcher typically does not have to know or choose the constants C_φ and α. These constants appear in the intermediate object $\hat{\Psi}(x)$ in (4.2) and are used for the asymptotic analysis.

Intuitively, when $X \in \mathbb{R}$, $F(x,h)$ can be approximated by $2f_X(x)h$ with the Lebesgue density f_X of X, and this assumption reduces to the one with $\varphi(h) \sim h$. Thus, the requirement $c_f \le \inf_{x\in\mathcal{X}} f(x)$ in (4.3) can be understood as the one that the Lebesgue density f_X should be bounded away from zero in the Euclidean case (in our case, the density or mass of $\|X-x\|$ at zero should be bounded away from zero). Furthermore, as illustrated in Remark 1, intuitively the constant α may be interpreted as the dimension of $\mathcal{X}$. In physical sciences, α is also known as the fractal order for fractal-type processes (Ferraty and Vieu 2006, chap. 13).

Assumption 2(iii) is a standard moment condition. Assumption 2(iv) requires a sufficiently higher-order kernel (of order $q \ge \alpha + 1$) compared to the decay rate of the small ball probability. In order to control the bias term in Proposition 2, another requirement on the kernel order q will be introduced. The condition on $\int z^{\alpha-1}K(z)dz$ (called $C_{\varphi,K}$ in Proposition 2) is imposed to avoid degeneracy of a variance component. The requirement in (4.4) intuitively rules out highly non-separable functional forms on the derivative $\frac{dF(x,s)}{ds}$. The function $\bar{f}(x)$ is an approximately multiplicative component in $\frac{dF(x,s)}{ds}$ depending only on x. Under the condition (4.3) in Assumption 2(ii), this requirement is typically satisfied with $f(x) = \bar{f}(x)$. We refer to (Ferraty and Vieu 2006, chap. 13) for specific examples. Assumption 2(v) is on the bandwidth h. Its second condition says that h cannot decay too fast to guarantee a sufficient amount of small ball probability to control the variance term.

Under these assumptions, the uniform convergence rate of $\hat{\Psi}(x) - \mathbb{E}[\hat{\Psi}(x)]$ is obtained as follows.

Proposition 1. Let W be a function of (D, Y, X) satisfying $\sup_{x\in\mathcal{X}} \mathbb{E}[|W|^\zeta \,|\, \|X - x\| = 0] \le C_W$ for some $\zeta > 2$ and $C_W < \infty$. Under Assumption 2, it holds

$$\sup_{x\in\mathcal{X}} |\hat{\Psi}(x) - \mathbb{E}[\hat{\Psi}(x)]| = O_p\left(\sqrt{\frac{(\log n)^\eta}{n\varphi(h)}}\right).$$

Note that the convergence rate depends on the decay rate $\varphi(h)$ of the small ball probability, and the component η controls the covering number of $\mathcal{X}$.

We next present two results to control the bias term of the kernel estimators with a functional covariate. In our context, the bias term can be written as (see Lemma 1 in Supplement)

$$\mathbb{E}\left[\frac{1}{n\varphi(h)}\sum_{i=1}^n K(\|X_i - x\|/h)\{g(X_i) - g(x)\}\right],$$

where $g(x) = \mathbb{E}[W \,|\, \|X-x\| = 0]$. Let $\hat{\varphi}(x) = \frac{1}{n}\sum_{i=1}^n K(\|X_i - x\|/h)$ be an estimator of the normalizing component of the small ball probability, $C_{\varphi,K} = \int z^{\alpha-1}K(z)dz$, and $\psi_g(s) = \mathbb{E}[g(X) - g(x) \,|\, \|X - x\| = s]$ for a functional $g : \mathcal{X} \to \mathbb{R}$. As clarified in Proposition 2, $\hat{\varphi}(x)/\varphi(h)$ can be considered as an estimator of $C_{\varphi,K}\bar{f}(x)$, where $\bar{f}(x)$ may be interpreted as the Lebesgue density for the Euclidean case. We add the following conditions.

Assumption 3. Let $\gamma > 0$ and let γ_0 be the integer such that $\gamma_0 < \gamma \le \gamma_0 + 1$. ψ_g is γ_0-times differentiable on $(0,1]$ and the γ_0th derivative $\psi_g^{(\gamma_0)}$ is $(\gamma - \gamma_0)$-Hölder continuous on $[0,1]$. Furthermore,

$$\int \psi_g(s)K(s/h)\left\{\frac{dF(x,s)}{ds}\right\}ds \sim \bar{f}(x)\int \psi_g(s)K(s/h)\left\{\frac{d\varphi(s)}{ds}\right\}ds, \tag{4.5}$$

where $\bar{f}(x)$ is defined in Assumption 2 (iv).

The first condition is on the smoothness of the function g (or ψ_g) to be estimated. The second condition is used to guarantee approximations of functionals of the small ball probability in terms of $\varphi(\cdot)$. The requirement in (4.5) is considered a generalization of (4.4) in Assumption 2 (iv) (i.e., (4.4) corresponds to the case of $\psi_g(s) = 1$). Under this assumption, we obtain the following results to control the bias term.

[3] This article focuses on the case of the binary treatment so that the expected potential outcomes $\mathbb{E}[Y_i(d)]$ for $d = 0, 1$ are two-dimensional objects. In a recent paper, Singh, Xu, and Gretton (2024) proposed a reproducing Hilbert space kernel-based ridge regression estimator for the causal response curve $\theta(d) = \mathbb{E}[Y_i(d)]$ defined over a Polish space $\mathcal{D}$, which may be continuous or infinite-dimensional. Their paper also allows functional covariates and presents a uniform convergence rate for their estimator $\tilde{\theta}(d)$, that is, the rate for $\sup_{d\in\mathcal{D}} |\tilde{\theta}(d) - \theta(d)|$. Since $\theta(d)$ is an infinite-dimensional object in Singh, Xu, and Gretton (2024), their convergence rate is typically slower than the parametric $\sqrt{n}$-rate. In contrast, the ATE $\theta(1) - \theta(0)$ of our interest is finite-dimensional and the $\sqrt{n}$-consistent estimation may be possible as shown in Theorem 1.

[4] The condition on the covering number is high level so that we do not need to specify the dimension of the domain of the functions $X \in \mathcal{X}$. The covering number $\mathcal{N}(\mathcal{F}, \|\cdot\|, \varepsilon)$ typically gets larger as the dimension of the domain increases. For example, an upper bound of the covering number of the unit ball of the reproducing kernel Hilbert space (RKHS) $H_\sigma([0,1]^d)$ generated by the Gaussian radial basis function kernel $K(x,y) = \exp(-\sigma^{-2}\|x - y\|_E^2)$ for $a, b \in \mathbb{R}^d$ and $\sigma > 0$ with the Euclidean norm $\|\cdot\|_E$ is given as $\log \mathcal{N}(H_\sigma([0,1]^d), \|\cdot\|_\infty, \varepsilon) \lesssim (\log(1/\varepsilon))^{d+1}$, where $\|\cdot\|_\infty$ is the supremum norm (see, Proposition 1 of Zhou (2002), for more details). Note that the RKHS $H_\sigma([0,1]^d)$ includes the Cameron-Martin space associated with a Gaussian measure as a special case (see Example 3 in Ferraty et al. 2010). In this example, the constant η in Assumption 2 (i) can be interpreted as the dimension of the domain.

Proposition 2. Under Assumption 2, it holds

$$\sup_{x\in\mathcal{X}} |\varphi(h)^{-1}\mathbb{E}[\hat{\varphi}(x)] - C_{\varphi,K}\bar{f}(x)| = o(1).$$

Additionally, suppose that Assumption 2 (iv) is satisfied with $q \geq \max\{\alpha+1, \gamma_0+\alpha-1\}$, and that Assumption 3 holds true. Then it holds

$$\sup_{x\in\mathcal{X}} |\mathbb{E}[(g(X)-g(x))K(\|X-x\|/h)]\varphi(h)^{-1}| = O(h^{\gamma}).$$

The first result of this proposition is a useful intermediate to characterize the limit of the denominator of the kernel estimator in (4.1) with a functional covariate. The second result establishes the order of the bias term of the kernel estimator. Similar to the conventional kernel estimator, the bias gets smaller for a given sequence of h as the function ψ_g becomes smoother (i.e., γ gets larger).

Remark 1. (Example) We close this subsection with an example where Assumptions 2 and 3 are satisfied. Let $L^2_{\mathbb{R}}([0,1])$ be the Hilbert space of all real-valued functions that are square integrable with respect to the Lebesgue measure on $[0,1]$ with the L^2-norm given by $\|a\|_2 = \sqrt{\langle a,a\rangle}$ where $\langle a,b\rangle = \int_0^1 a(t)b(t)dt$ and $a,b \in L^2_{\mathbb{R}}([0,1])$. Let $\{\Xi_{1,i},\ldots,\Xi_{\alpha,i}\}_{i=1}^n$ be an iid sequence with compactly supported density f_Ξ and let $\{q_k(t)\}_{k=1}^\infty$ be an orthogonal basis of $L^2_{\mathbb{R}}([0,1])$. Define $X_i(t) = \sum_{k=1}^\alpha \Xi_{k,i}q_k(t)$. In this case, the space $\mathcal{X}$ for X can be defined as $\mathcal{X} = \{X = \sum_{k=1}^\alpha \xi_k q_k : (\xi_1,\ldots,\xi_\alpha) \in \mathrm{supp}(f_\Xi)\} \subset L^2_{\mathbb{R}}([0,1])$. Note that $X \in \mathcal{X}$ is determined by the α-dimensional coefficient vector $\xi = (\xi_1,\ldots,\xi_\alpha)'$. Furthermore, $\mathcal{X}$ is a compact set with respect to the norm $\|X\| := \sqrt{\langle X,X\rangle} = \sqrt{\sum_{k=1}^\alpha \xi_k^2}$. If the object $g(X) = g(\sum_{k=1}^\alpha \xi_k q_k)$ is sufficiently smooth on $\mathrm{supp}(f_\Xi)$, then Assumptions 2 and 3 are typically satisfied. For example, if we assume that there exists a function $\bar{g} : \mathbb{R}^\alpha \to \mathbb{R}$ such that $g(\sum_{k=1}^\alpha \xi_k q_k) = \bar{g}(\xi_1,\ldots,\xi_\alpha)$ and that $\bar{g}$ and f_Ξ are (γ_0+1)-times continuously partially differentiable on the interior of $\mathrm{supp}(f_\Xi)$, then one can see that Assumptions 2 (i), (ii), (iv), and (v) and 3 are satisfied with $\eta = 1$, $\varphi(h) \sim h^\alpha$, and $f = \bar{f} = f_\Xi$.

Remark 2 (Euclidean covariates). Although the presentation will be tedious, when we have additional Euclidean covariates $Z \in \mathcal{Z}$ for a compact $\mathcal{Z} \subset \mathbb{R}^{d_z}$, the uniform convergence rate of the kernel-weighted summation $\hat{\Psi}(x,z) = \frac{1}{n\varphi(h)h_z^{d_z}}\sum_{i=1}^n K(\|X_i - x\|/h)K_z((Z_i-z)/h_z)W_i$ will be

$$\sup_{x\in\mathcal{X},z\in\mathcal{Z}} |\hat{\Psi}(x,z) - \mathbb{E}[\hat{\Psi}(x,z)]| = O_p\left(\sqrt{\frac{(\log n)^\eta}{n\varphi(h)h_z^{d_z}}}\right),$$

and the bias term will be of an analogous order that involves a power of h_z.

Remark 3 (Conditional ATE). Although the focus of this article is on estimation of the ATE θ, the theoretical results for $\hat{\Psi}(x)$ established in this subsection can be applied directly to derive the uniform convergence rate for the nonparametric kernel estimator of the conditional ATE $m_d(x) = \mathbb{E}[Y(d)|\,\|X-x\| = 0]$ for $d = 0, 1$, that is

$$\hat{m}_d(x) = \frac{\sum_{i=1}^n \mathbb{I}\{D_i = d\}Y_iK(\|X_i-x\|/h)}{\sum_{i=1}^n \mathbb{I}\{D_i=d\}K(\|X_i-x\|/h)}. \tag{4.6}$$

The asymptotic properties of the numerator and denominator of this estimator can be obtained by applying Propositions 1 and 2 with setting $W_i = \mathbb{I}\{D_i = d\}Y_i$ and $\mathbb{I}\{D_i = d\}$, respectively.

4.2. Asymptotic Property of Inverse Probability Weighting Estimator

Based on the uniform convergence results in the last subsection, we now derive the asymptotic distribution of the IPW estimator $\hat{\theta}$ in (3.3).

Theorem 1. Suppose Assumptions 1, 2, and 3 hold true and $g \in \{p, p^2, \tau p, m_0\}$. Furthermore, assume $nh^{2\gamma} \to 0$ and $n\varphi(h)^2 \to \infty$ as $n \to \infty$. Then we have

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \mathbb{E}\left[\frac{\sigma^2(X,D)\{D-p(X)\}^2}{p^2(X)\{1-p(X)\}^2}\right] + \mathrm{var}(\tau(X))\right),$$

where $\sigma^2(x,d) = \mathbb{E}[(Y-m_d(X))^2|\,\|X-x\| = 0, D = d]$, $m_d(x) = \mathbb{E}[Y|\,\|X-x\| = 0, D = d]$, and $\tau(x) = \mathbb{E}[Y(1) - Y(0)|\,\|X-x\| = 0]$.

This theorem says that even if the propensity score $p(X)$ is estimated by the kernel estimator using a functional covariate X, we can still achieve the $\sqrt{n}$-consistency for the semiparametric IPW estimator $\hat{\theta}$. The asymptotic normality is obtained by adapting the U-statistic argument for semiparametric estimators to the present setup. We also note that the above asymptotic variance equals the asymptotic efficiency bound for the ATE (Hahn 1998). Therefore, the proposed IPW estimator $\hat{\theta}$ is asymptotically efficient.

The asymptotic variance of $\hat{\theta}$ can be estimated by

$$\hat{V}_\theta = \frac{1}{n}\sum_{i=1}^n \frac{\hat{U}_i^2\{D_i - \hat{p}(X_i)\}^2}{\hat{p}^2(X_i)\{1-\hat{p}(X_i)\}^2} + \frac{1}{n}\sum_{i=1}^n (\hat{\tau}(X_i) - \check{\theta})^2,$$

where $\hat{U}_i = Y_i - (\hat{m}_0(X_i) + \hat{\tau}(X_i)D_i)$, $\check{\theta} = \frac{1}{n}\sum_{i=1}^n \hat{\tau}(X_i)$, and

$$\hat{m}_0(X_i) = \frac{\hat{p}(X_i)\{\hat{r}(X_i) - \hat{r}_1(X_i)\}}{\hat{p}(X_i)\{1-\hat{p}(X_i)\}},$$
$$\hat{\tau}(X_i) = \frac{\hat{r}_1(X_i) - \hat{r}(X_i)\hat{p}(X_i)}{\hat{p}(X_i)\{1-\hat{p}(X_i)\}},$$
$$\hat{r}(X_i) = \frac{\sum_{j\neq i} Y_jK(\|X_j-X_i\|/h)}{\sum_{j\neq i} K(\|X_j-X_i\|/h)},$$
$$\hat{r}_1(X_i) = \frac{\sum_{j\neq i} Y_jD_jK(\|X_j-X_i\|/h)}{\sum_{j\neq i} K(\|X_j-X_i\|/h)}.$$

The consistency of this variance estimator is obtained as follows.

Theorem 2. Under the same assumption of Theorem 1, it holds

$$\hat{V}_\theta \xrightarrow{p} \mathbb{E}\left[\frac{\sigma^2(X,D)\{D-p(X)\}^2}{p^2(X)\{1-p(X)\}^2}\right] + \mathrm{Var}(\tau(X)).$$

By Theorems 1 and 2, we obtain the t-ratio $\frac{\sqrt{n}(\hat{\theta}-\theta)}{\sqrt{\hat{V}_\theta}} \xrightarrow{d} N(0,1)$ to conduct asymptotic inference on θ.

Applying almost the same argument to show Theorem 1, it is also possible to show a bootstrap version of the t-ratio. Let

$S_i = (Y_i, X_i, D_i)$ and let $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{S_i}$ denote its empirical distribution. Conditionally on the data $\{S_1, \ldots, S_n\}$, generate $S_1^b = (Y_1^b, X_1^b, D_1^b), \ldots, S_n^b \sim_{\text{iid}} \mathbb{P}_n$, and compute the bootstrap counterpart

$$\hat{\theta}^b = \frac{1}{n}\sum_{i=1}^n \left(\frac{D_i^b Y_i^b}{\hat{p}^b(X_i^b)} - \frac{(1-D_i^b)Y_i^b}{1-\hat{p}^b(X_i^b)} \right),$$

which is defined by replacing S_i with S_i^b. An analogous argument to (Li, Racine, and Wooldridge 2009, Theorem 2.2) yields the validity of this bootstrap procedure as follows.

Theorem 3. Under the same assumption of Theorem 1, it holds

$$\sup_{r\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\sqrt{n}(\hat{\theta}^b - \hat{\theta})}{\sqrt{\hat{V}_\theta}} \le r \,\middle|\, \{S_i\}_{i=1}^n \right) - \Phi(r) \right| \xrightarrow{p} 0,$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$.

Remark 4 (Alternative estimation methods). Although it is beyond the scope of this article, it is of interest to study alternative estimation methods for the ATE θ. For example, based on the estimator for the conditional mean in (4.6), the nonparametric regression-based estimator for θ is obtained as $\hat{\theta}_R = \frac{1}{n}\sum_{i=1}^n \{\hat{m}_1(X_i) - \hat{m}_0(X_i)\}$. Another more intriguing example is a doubly robust estimator, which can be constructed as

$$\hat{\theta}_{DR} = \frac{1}{n}\sum_{i=1}^n \left[\frac{D_i(Y_i - \hat{m}_1(X_i))}{\hat{p}(X_i)} + \hat{m}_1(X_i) - \frac{(1-D_i)(Y_i - \hat{m}_0(X_i))}{1-\hat{p}(X_i)} - \hat{m}_0(X_i) \right],$$

by using the propensity score estimator $\hat{p}(X_i)$ and regression estimators $\hat{m}_1(X_i)$ and $\hat{m}_0(X_i)$. In the case of high-dimensional Euclidean covariates, Farrell (2015) studied asymptotic properties of an analogous doubly robust estimator, where the propensity score and regression functions are estimated by a group lasso method, and established its doubly robust property.

In the current context, a key sufficient condition to achieve the $\sqrt{n}$-consistency and asymptotic normality in Theorem 1 for our estimator $\hat{\theta}$ is a sufficiently fast uniform convergence rate of the propensity score estimator (i.e., $\sup_{x\in\mathcal{X}} |\hat{p}(x) - p(x)| = o_p(n^{-1/4})$) based on Lemma 1 in Appendix. As clarified in Section 4.1, this sufficiently fast convergence rate requires the entropy condition on $\mathcal{X}$ (in Assumption 2 (i)) and smoothness conditions combined with the conditions on the small ball probability $\varphi(h)$ (in Assumptions 2 (i) and (iv) and 3). On the other hand, based on the theoretical developments in Farrell (2015), we can expect that the doubly robust estimator $\hat{\theta}_{DR}$ will achieve the same limiting distribution in Theorem 1 under different key sufficient conditions: (i) $\frac{1}{n}\sum_{i=1}^n (\hat{p}(X_i) - p(X_i))^2 = o_p(1)$, (ii) $\frac{1}{n}\sum_{i=1}^n (\hat{m}_d(X_i) - m_d(X_i))^2 = o_p(1)$ for $d = 0, 1$, and (iii) the product rate conditions

$$\sqrt{\frac{1}{n}\sum_{i=1}^n D_i(\hat{p}(X_i) - p(X_i))^2}\sqrt{\frac{1}{n}\sum_{i=1}^n D_i(\hat{m}_1(X_i) - m_1(X_i))^2} = o_p(n^{-1/2}),$$

$$\sqrt{\frac{1}{n}\sum_{i=1}^n (1-D_i)(\hat{p}(X_i) - p(X_i))^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (1-D_i)(\hat{m}_0(X_i) - m_0(X_i))^2} = o_p(n^{-1/2}).$$

These conditions are typically weaker than the key condition $\sup_{x\in\mathcal{X}} |\hat{p}(x) - p(x)| = o_p(n^{-1/4})$ for our IPW estimator $\hat{\theta}$: (i) and (ii) are mild L^2-consistency conditions, and (iii) is easier to satisfy if either the propensity score or regression function is easier to estimate. We note that as far as their regularity conditions are satisfied, both the IPW and doubly robust estimators will exhibit the same limiting distribution and achieve the semiparametric efficiency bound. For example, Cattaneo (2010) obtained the same limiting distribution for the IPW estimator as the one by Farrell (2015) for the doubly robust estimator to estimate multivalued treatment effects.

As studied in Farrell (2015) for the case of Euclidean covariates, the asymptotic analysis for the doubly robust estimator $\hat{\theta}_{DR}$ (and also the regression-based estimator $\hat{\theta}_R$) in the present setup requires a separate investigation and we leave it for future research.

5. Numerical Illustrations

In this section, we present three numerical illustrations to show the effectiveness of our proposed method.

5.1. Comparison with Parametric Approach

5.1.1. Correctly Specified Case

We consider the following data generating process (DGP):

$$Y = \theta \cdot D + \sum_{j=1}^4 \beta_j C_j + \varepsilon, \qquad X(t) = \sum_{j=1}^4 C_j \phi_j(t),$$

$$p(X) = \mathbb{P}(D = 1|X) = \frac{\exp\left(\int_0^1 \eta(t)X(t)dt\right)}{1 + \exp\left(\int_0^1 \eta(t)X(t)dt\right)}, \tag{5.1}$$

where $\varepsilon \sim N(0,1)$. The functions $\{\phi_j(t)\}_{j=1}^4$ are the first four elements of the Fourier basis on $[0,1]$, that is, $\phi_1 = 1$, $\phi_2(t) = \sqrt{2}\cos(2\pi t)$, $\phi_3(t) = \sqrt{2}\sin(2\pi t)$, and $\phi_4(t) = \sqrt{2}\cos(4\pi t)$, and the Fourier coefficients $\{C_j\}_{j=1}^4$ are generated by $C_j \sim_{iid}$ Uniform$[0,1]$. Note that the random coefficients C_j's can be regarded as observed since the functional covariate $X(t)$ is fully observed, and its Fourier expansion (if exists) is unique. The function $\eta(t)$ can be expressed as $\eta(t) = \sum_{j=1}^4 \gamma_j \phi_j(t)$. For the parameters in (5.1), we set $\theta = 0.5$ and $\gamma = \beta = (0.3, 0.3, 0.1, 0.1)'$. The researcher aims to estimate the ATE θ without knowledge of γ and β.

To estimate the propensity score function $p(\cdot)$, we employ the proposed nonparametric estimator $\hat{p}(X_i) = \frac{\sum_{j\neq i}^n D_j K(\|X_j - X_i\|/h)}{\sum_{j\neq i}^n K(\|X_j - X_i\|/h)}$ and the parametric logit estimator. Then to estimate the ATE, we apply the IPW estimator in (3.3) for both methods. The logit model for the propensity score is specified as

$$p(X) = \frac{\exp\left(\sum_{j=1}^4 \gamma_j C_j\right)}{1 + \exp\left(\sum_{j=1}^4 \gamma_j C_j\right)}. \tag{5.2}$$

Note that under the setting (5.1), this logit model is correctly specified because $\int_0^1 \eta(t)\, X(t)dt = \sum_{j=1}^4 \gamma_j C_j$, given that the Fourier basis is orthonormal.

In this setting, since the functional covariates reside in a sieve space of dimension $d = 4$ (a detail known to the user), we can heuristically choose $h = n^{-\frac{1}{d+4}} = n^{-\frac{1}{8}}$. To measure the distance of two functional covariates, we employ the L^2-distance: $\|X_j - X_i\|_2 = \sqrt{\int_0^1 \{X_j(t) - X_i(t)\}^2 dt}$. In the simulation, this distance is approximated by $\sqrt{\frac{1}{101}\sum_{t=0}^{100}\left\{X_j\left(\frac{t}{100}\right) - X_i\left(\frac{t}{100}\right)\right\}^2}$, that is, the average of squared distances evaluated at 101 equally spaced points on $[0, 1]$.

The simulation results based on 1000 Monte Carlo replications are presented in Table 1, where the Monte Carlo means and MSEs rescaled by n are reported. The left panel shows the simulation results of the IPW estimator based on the logit model, and the right panel shows the results based on the nonparametric kernel estimator. From Table 1, we can see that both methods seem to consistently estimate the ATE $\theta = 0.5$. Since the logit model of $p(\cdot)$ is correctly specified in this DGP, the parametric method slightly outperforms our proposed method in terms of the MSE. Both methods appear to provide approximately correct empirical coverages, and the average sizes of their confidence intervals (CIs) remain comparable between the two methods across all sample sizes.

5.1.2. Misspecified Case

In this subsection, we consider another DGP that mirrors DGP (5.1), with the only difference being the specification of the propensity score:

$$p(X) = \frac{l + \sum_{j=1}^4 \exp(-C_j^2)}{u + \sum_{j=1}^4 \exp(-C_j^2)}, \tag{5.3}$$

where we set $l = 0.1$ and $u = 0.6$ to prevent the propensity score from approaching 0 or 1 (see Assumption 1). For the parametric method, we continue to employ the logit model in (5.2), which is misspecified under the DGP in (5.3). The implementation of our IPW estimator remains identical to that for the DGP in (5.1).

From Table 2, it is evident that the misspecified propensity score model fails to consistently estimate $\theta = 0.5$, exhibiting a persistent bias of roughly 0.1. Accordingly, its empirical coverage gradually falls below 95% and continues to decrease as the sample size increases. In contrast, our proposed method appears to be consistent: the biases steadily converge to zero, and the stable MSEs (rescaled by a factor of n) suggest a $\sqrt{n}$-consistent estimation of the ATE under the DGP in (5.3). The MSEs of the logit model are approximately twice as large as those of the proposed nonparametric method, and they exhibit a trend of further increasing (after being rescaled by a factor of n). In addition, our approach provides approximately correct empirical coverages and generally smaller CIs.

5.2. Functional ATE with User-Specified Semi-Norm

In this subsection, we generate a dataset from a market regulator model to evaluate the effectiveness of the proposed method.

5.2.1. Market Regulator Model

Consider an imaginary financial market overseen by a market regulator whose objective is to ensure smooth functioning of the market. Her guiding principle is that the greater the market's volatility, the more likely she will implement the circuit breaker. While most investors (say, the general public) may be completely unaware of this principle, it is known to a small group of investors (say, economists). However, this group of investors remains unaware of the exact criteria employed by the regulator. All investors seek to assess the necessity of appointing such a market regulator by estimating the ATE of activating a circuit breaker. Thus, in this subsection, the treatment variable D_i indicates whether the regulator activates a circuit breaker after day i.

The regulator and investors observe the daily stock price index X, which is a random function evaluated at $m + 1$ equally spaced points throughout the opening time of the stock market. In the following, we set $m = 100$.

The outcome variable Y represents welfare measures of financial market health, such as systematic risk or Moody's Sovereign Credit Ratings. Suppose these criteria are standardized such that a larger Y indicates a higher level of welfare, and we can model the outcome as

Table 1. Correctly specified logit versus nonparametric kernel estimation.

	Logit					Nonparametric			
n	mean	n·MSE	EC	AL	n	mean	n·MSE	EC	AL
100	0.5058	4.3345	0.9490	0.8243	100	0.5197	4.0786	0.9450	0.7934
200	0.5057	4.3618	0.9420	0.5739	200	0.5215	4.3875	0.9400	0.5624
500	0.4993	4.2752	0.9470	0.3602	500	0.5165	4.4701	0.9450	0.3565
1000	0.4976	4.3546	0.9490	0.2541	1000	0.5145	4.7667	0.9310	0.2523

NOTE: The true ATE is $\theta = 0.5$. "MSE" stands for the mean square error; "EC" and "AL" stand for the empirical coverage and the average length for the 95% confidence interval.

Table 2. Misspecified logit versus nonparametric kernel estimation.

	Logit					Nonparametric			
n	mean	n·MSE	EC	AL	n	mean	n·MSE	EC	AL
100	0.3908	23.418	0.9520	1.6471	100	0.5170	9.0029	0.9350	1.1244
200	0.4100	18.312	0.9400	1.0989	200	0.5150	8.7165	0.9430	0.8145
500	0.4050	20.559	0.9090	0.6663	500	0.5052	10.012	0.9410	0.5235
1000	0.4056	22.980	0.8740	0.4616	1000	0.4969	9.0506	0.9440	0.3714

NOTE: The true ATE is $\theta = 0.5$. "MSE" stands for the mean square error; "EC" and "AL" stand for the empirical coverage and the average length for the 95% confidence interval.

$$Y = D \cdot [\theta_1 \cdot \mathbb{I}\{TV_m(X) > z\} + \theta_2 \cdot \mathbb{I}\{TV_m(X) \le z\}] + \varepsilon, \quad (5.4)$$

where $TV_m(X) = \sum_{\ell=0}^{m-1} |X(t_{\ell+1}) - X(t_\ell)|$ may be interpreted as the "total variation" of the random function $X(t)$ *measured at* these $m+1$ points. (We use quotation marks because sometimes the actual total variation of a function might not exist. For example, the total variation for a Wiener process is infinite; see the first setup of Section 5.2.2).

The constant z in (5.4) is a threshold for the market volatility with $\theta_1 > 0$ and $\theta_2 < 0$, which indicates that if the market volatility exceeds z, the regulator's intervention has a positive impact; however, if the volatility is below this threshold, her intervention becomes counterproductive. The market regulator does not know the threshold z, and her treatment rule D is determined by

$$\begin{aligned} D &= \mathbb{I}\{\eta \le p(X)\}, \qquad \eta \sim \text{Uniform}[0,1], \\ p(X) &= l + (u-l) \cdot \mathbb{I}\{TV_m(X) > z_r\}, \end{aligned} \quad (5.5)$$

where η is independent of (X, ε), l and u are the lower and upper bounds of the propensity score satisfying $0 < l < u < 1$, and z_r is her subjective threshold. In a financial market, η may represent long-term variations in regulatory policies, market sentiment and risk appetite, or technological infrastructure. These factors significantly influence systemic risk but may have minimal impacts on daily stock prices. Given that η is assumed to be independent of both X and ε, Assumption 1 is fulfilled.

Both z and z_r are unknown to all investors. Investors in Groups A and B employ our nonparametric method to estimate the propensity score, and then apply the IPW presented in (3.3) to estimate the ATE. The investors in Group A, who are not aware that the regulator's treatment rule fundamentally relies on the measure of daily price fluctuations, inappropriately use the L^2-norm for their first-stage estimation. The investors in Group B, who are informed about the regulator's decision-making principles, apply a user-defined semi-metric that reflects these principles:

$$d_{TV_m}(x_1, x_2) = |TV_m(x_1) - TV_m(x_2)|, \quad (5.6)$$

for $x_1, x_2 \in \mathcal{X}$, where the space $\mathcal{X}$ will be defined in Section 5.2.2. Based on this, the investors in Group B estimate the propensity score by

$$\hat{p}(X_i) = \frac{\sum_{j \ne i}^{n} D_j K(d_{TV_m}(X_i, X_j)/h)}{\sum_{j \ne i}^{n} K(d_{TV_m}(X_i, X_j)/h)}. \quad (5.7)$$

In addition, the investors in Group C are aware of the general principle. However, they decide to use a parametric logit model to estimate the propensity score

$$p(X) = \frac{\exp(1 + \beta \cdot TV_m(X))}{1 + \exp(1 + \beta \cdot TV_m(X))}, \quad (5.8)$$

where β is an unknown parameter.

5.2.2. Simulation Setups

Here, we examine two data-generating processes for X. In the first scenario, we generate the data by

$$X(t) = X(t - \Delta t) + \Delta X(t), \qquad \Delta X(t) \sim_{\text{iid}} N(0, \Delta t \cdot 100).$$

We let $X(t)$ be evaluated at $m+1$ equally spaced points on $[0, 1]$, with m set to 100. Thus, a random function can be sampled as

$$X\left(\frac{\ell}{100}\right) = \begin{cases} 0 & \text{for } \ell = 0 \\ \sum_{j=1}^{\ell} e_j & \text{for } \ell = 1, \ldots, m \end{cases},$$

where $e_i \sim_{\text{iid}} N(0, 1)$.

In the second scenario, $X(t)$ is sampled from the real-world financial market. We use the high-frequency data of the Standard & Poor's 500 Index (S&P 500 hereafter) that are collected at 1-min intervals.[5] The data spans from April 2, 2007 to October 2, 2024, encompassing 4524 daily realizations of random functions. We have adjusted the data collection frequency to $m = 100$ for each day to maintain a consistent discussion throughout this section. We use this pool of 4524 functions as our population, from which we draw $X(t)$ with replacement for our simulation study.

Figure 2 displays some sample realizations of X in both scenarios. The left panel presents 30 random draws for the case where $X(t)$ follows a Wiener process on the interval [0,1]; the right panel presents 30 daily series spanning from August 22, 2024 to October 2, 2024.

Figures 3 and 4 display realizations of the random function X with different levels of $TV_m(X)$.

For the outcome model (5.4), we set $\theta_1 = 10$, $\theta_2 = -1$, and $z = Q_x^{TV_m}(0.8)$, where $Q_x^{TV_m}(\alpha)$ denotes the αth quantile of the random variable $TV_m(X)$. Under this setting, the true ATE is $\theta = -1 \times 0.8 + 10 \times 0.2 = 1.2$. For the propensity score model in (5.5), we set $z_r = Q_x^{TV_m}(0.5)$, $l = 0.3$, and $u = 0.7$. In addition, we define $r = Q_x^{TV_m}(0.9) - Q_x^{TV_m}(0.1)$ for both scenarios. In the estimation of the first scenario, the investors in Group A choose a bandwidth of $h = r \cdot n^{\frac{1}{8}}$, while those in Group B choose $h = r \cdot n^{\frac{1}{5}}$. In the second scenario, since the real data have a greater range of variation, we let the investors in Group A choose a bandwidth of $h = \text{mean}(TV_m(X)) \cdot r \cdot n^{\frac{1}{8}}$, to avoid the numerical issue during the simulation. The bandwidth choice for investors in Group B remains $h = r \cdot n^{\frac{1}{5}}$.[6]

5.2.3. Simulation Results

The simulation results based on 1000 Monte Carlo replications for both scenarios are presented in Tables 3 and 4, respectively. For both tables, the Monte Carlo means and MSEs rescaled by n are reported.

[5] The data is dowloaded from the website of BacktestMarket, *https://www.backtestmarket.com/en/sp-500-1m.*

[6] In practice, one may choose the bandwidth by adapting a data-driven method to estimate or conduct inference on the propensity score function $p(\cdot)$, such as cross-validation or L_∞-based selector by Bissantz et al. (2007). However, our ATE estimator $\hat{\theta}$ is semiparametric and typically requires undersmoothing to ensure that the bias term of the first-stage estimator $\hat{p}(\cdot)$ is asymptotically negligible. This issue remains an open question even when dealing with conventional Euclidean covariates and we leave such investigation for future research.

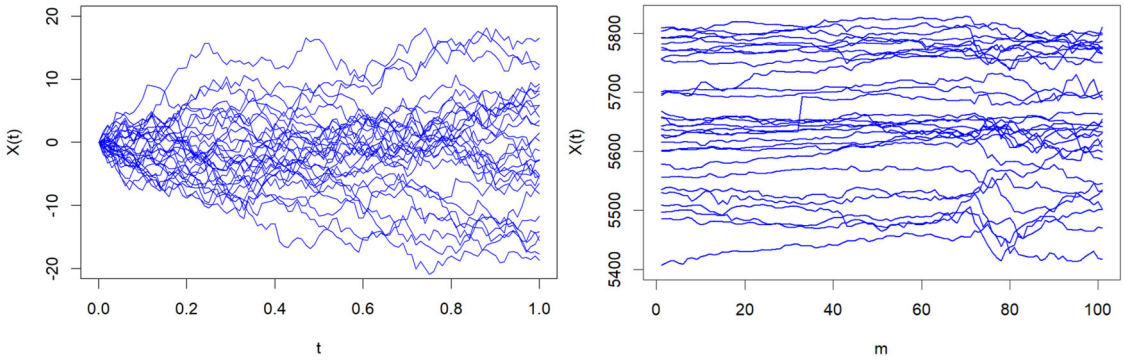

Figure 2. 30 random realizations of X in both scenarios.
NOTE: The left figure presents 30 random draws for the case where $X(t)$ follows a Wiener process on the interval [0,1]; the right panel presents 30 daily series of S&P 500 spanning from August 22, 2024 to October 2, 2024.

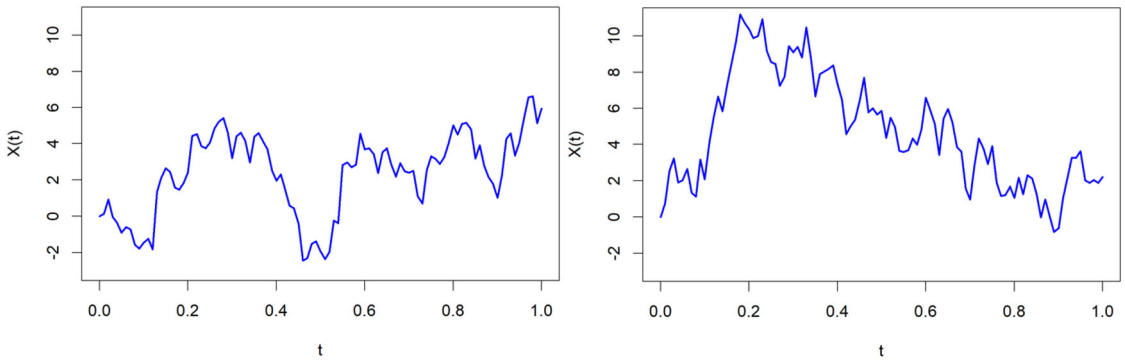

Figure 3. Two random realizations of X with different $TV_m(X)$, where $X \sim$ Wiener[0, 1].
NOTE: The left figure displays a random function with $TV_m(X) = 71.47$, which is approximately the 10th percentile of $TV_m(X)$, and the right figure displays the one with $TV_m(X) = 90.29$, which is approximately the 90th percentile of $TV_m(X)$.

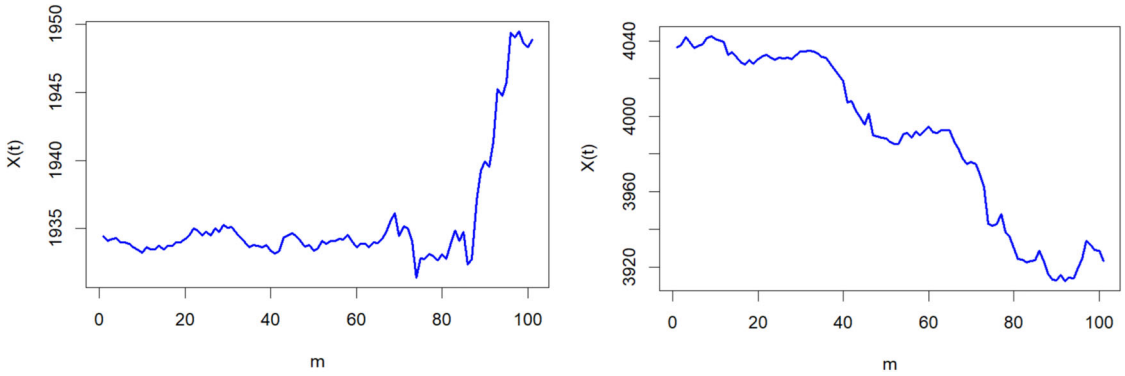

Figure 4. Two random realizations of X with different $TV_m(X)$, where X represents daily realizations of the S&P 500.
NOTE: The left figure displays an observation of S&P 500 with $TV_m(X) = 54.93$, which is approximately the 10th percentile of $TV_m(X)$ of the population we choose for our study, a pool of 4524 daily realizations of the S&P 500 indices between April 2, 2007 to October 2, 2024. The right figure displays one with $TV_m(X) = 285.31$, which is approximately the 90th percentile of $TV_m(X)$.

Table 3. Misspecified logit versus nonparametric methods under two different metrics for the first scenario.

	Group A				Group B				Group C			
n	mean	n·MSE	EC	AL	mean	n·MSE	EC	AL	mean	n·MSE	EC	AL
100	1.927	74.43	0.69	1.895	1.391	13.51	0.99	1.891	1.081	12.14	0.99	1.948
200	1.938	129.6	0.40	1.339	1.352	13.01	0.99	1.338	1.095	13.26	0.98	1.372
500	1.928	291.2	0.07	0.847	1.292	12.36	0.99	0.848	1.094	24.62	0.97	0.863
1000	1.927	560.8	0.00	0.599	1.266	11.58	0.99	0.600	1.096	39.22	0.95	0.610
2000	1.927	1101	0.00	0.423	1.245	11.02	0.99	0.425	1.093	68.43	0.89	0.431
5000	1.924	2705	0.00	0.267	1.229	10.46	0.99	0.269	1.092	151.5	0.58	0.272

NOTE: $X \sim$ Wiener[0, 1]. The true ATE is $\theta = 1.2$. For the propensity score estimation, Group A investors employ the kernel method with the L^2-norm, Group B investors employ the kernel method with the user-defined metric in (5.6), and Group C investors employ the parametric logit model in (5.8). "MSE" stands for the mean square error; "EC" and "AL" stand for the empirical coverage and the average length for the 95% confidence interval.

Table 4. Misspecified logit versu nonparametric methods under two different metrics for the second scenario.

	Group A				Group B				Group C			
n	mean	n·MSE	EC	AL	mean	n·MSE	EC	AL	mean	n·MSE	EC	AL
100	1.845	57.17	0.78	1.893	1.282	8.905	0.99	1.891	1.161	13.60	0.99	2.088
200	1.827	94.62	0.55	1.339	1.270	8.414	1.00	1.338	1.096	2262	0.99	2.152
500	1.779	184.9	0.20	0.847	1.231	7.497	0.99	0.847	1.193	773.6	0.98	1.635
1000	1.750	319.3	0.02	0.599	1.214	6.484	1.00	0.599	1.201	196.6	0.98	1.203
2000	1.725	567.8	0.00	0.423	1.205	5.489	1.00	0.424	1.198	283.2	0.97	1.010
5000	1.689	1210	0.00	0.268	1.200	5.471	1.00	0.269	1.196	186.4	0.96	0.701

NOTE: X's are sampled from a pool of 4524 daily realizations of the S&P 500 indices. The true ATE is $\theta = 1.2$. For the propensity score estimation, Group A investors employ the kernel method with the L^2-norm, Group B investors employ the kernel method with the user-defined metric in (5.6), and Group C investors employ the parametric logit model in (5.8). "MSE" stands for the mean square error; "EC" and "AL" stand for the empirical coverage and the average length for the 95% confidence interval.

We first look at Table 3 for the first scenario. The biases and variances in this simulation generally exceed those in Section 5.1. This can be partly attributed to the Wiener processes, which exhibit a much higher variance—the right endpoint of each process has a variance of 100. Second, the left panel of this table shows that the L^2-norm employed by Group A investors fails to capture the link between the highly fluctuating X and the treatment D. There is a persistent bias of roughly 0.72, and MSEs are growing with the sample size. Third, Group C investors recognize that $TV_m(X)$ is crucial for treatment assignment, but they employ a misspecified logit model. For the sample size of $n = 100$, the parametric method seems to outperform in terms of MSE. However, this perception may be deceptive, as for small sample sizes, variance often dominates bias. With the sample size growing, the parametric estimator is evidently inconsistent, exhibiting a persistent bias of roughly 0.11. Finally, in contrast, Group B investors, applying the nonparametric kernel estimation and user-defined metric in (5.6), attain the best results. While there is still a discernible bias of approximately 0.03 for $n = 5000$, there is a clear trend indicating the bias is converging to zero. Also, the stable MSEs (rescaled by a factor of n) suggest a $\sqrt{n}$-consistent estimation of the ATE.

In terms of CI coverage, Group B consistently provides empirical coverage above the nominal 95%. Although signs of over-coverage are present, the size of CIs steadily decreases as the sample size increases, suggesting that the approach used in Group B yields informative CIs. In contrast, the empirical coverages in Groups C and A gradually fall below 95% and continue to decline as the sample size increases.

Comparing the results between Groups A and B highlights the critical importance of selecting an appropriate semi-norm when dealing with functional covariates, a notion that resonates with (Ferraty and Vieu 2006, chap. 13.6). Without an appropriate choice, the vast amount of information embedded in infinite-dimensional covariates cannot be effectively extracted by economic models. This significance parallels the importance of selecting the correct set of covariates in a linear regression context. In this particular example, the choice of the right semi-norm is evidently dependent on the knowledge of the regulator's objectives, typically guided by economic theory. Developing a data-driven approach for selecting the semi-norm is beyond the scope of this article. Nevertheless, it presents an interesting topic for future research. We refer to (Ferraty and Vieu 2006, chap. 13) for a relevant discussion.

Table 4, where we have changed the DGP of $X(t)$ from Wiener$[0, 1]$ to a random draw from a pool of 4524 daily realizations of the S&P 500 index, shows a pattern similar to that of Table 3, except that the parametric estimator starts to perform poorly from $n = 200$. Group B, in which the investors employ nonparametric kernel estimation and a user-defined metric as defined in (5.6), is the only group that delivers satisfactory results across all sample sizes. Although the empirical coverages in Groups B and C are both above the nominal 95% (up to $n = 5000$), the coverage in Group C exhibits a clear downward trend due to inconsistent estimation, and its CIs are considerably larger than those in Group B.

Overall, the simulation results support the implementation of our proposed method.

6. Empirical Application

In this section, we illustrate the usefulness of our approach. Maintaining a low unemployment rate, stabilizing inflation, and promoting rapid economic growth are the ultimate goals of central banks. How monetary policies, such as changes in the Federal Open Market Committee (FOMC) target rates, can affect macroeconomic targets, has been extensively studied in the literature; see, for example, Romer and Romer (1989), Bagliano and Favero (1998), and Christiano, Eichenbaum, and Evans (1999), and many others. Much of the work mentioned relies on structural models to identify the economic effects of monetary contraction and expansion. Alternatively, (Angrist, Jordà, and Kuersteiner 2018, AJK hereafter) proposed to use a potential outcome time series framework, based on the IPW estimator for treatment effects, to study the impulse response function of monetary policy (see also Angrist and Kuersteiner 2011). AJK (on p. 374) argued that the impulse response function can be defined as a marginalized version of ATEs at different horizons. Here, we build on the foundational method proposed by AKJ, adapting it within our framework by incorporating functional covariates.

AKJ studied a multinomial treatment effect model that involves five treatment statuses of the monetary policy, increasing the *federal funds target rate* by 0.25%, 0.5%, decreasing it by 0.25%, 0.5%, and no change. To align this with our framework, we simplify their approach by consolidating these categories into a binary treatment variable: treatment $D = 1$ if the target rate is decreased (expansion), and $D = 0$ if it is increased or unchanged (contraction). For Y, we use the outcome variables from the monthly dataset adopted in AJK, including the federal funds rate, indicators for inflation and industrial production,

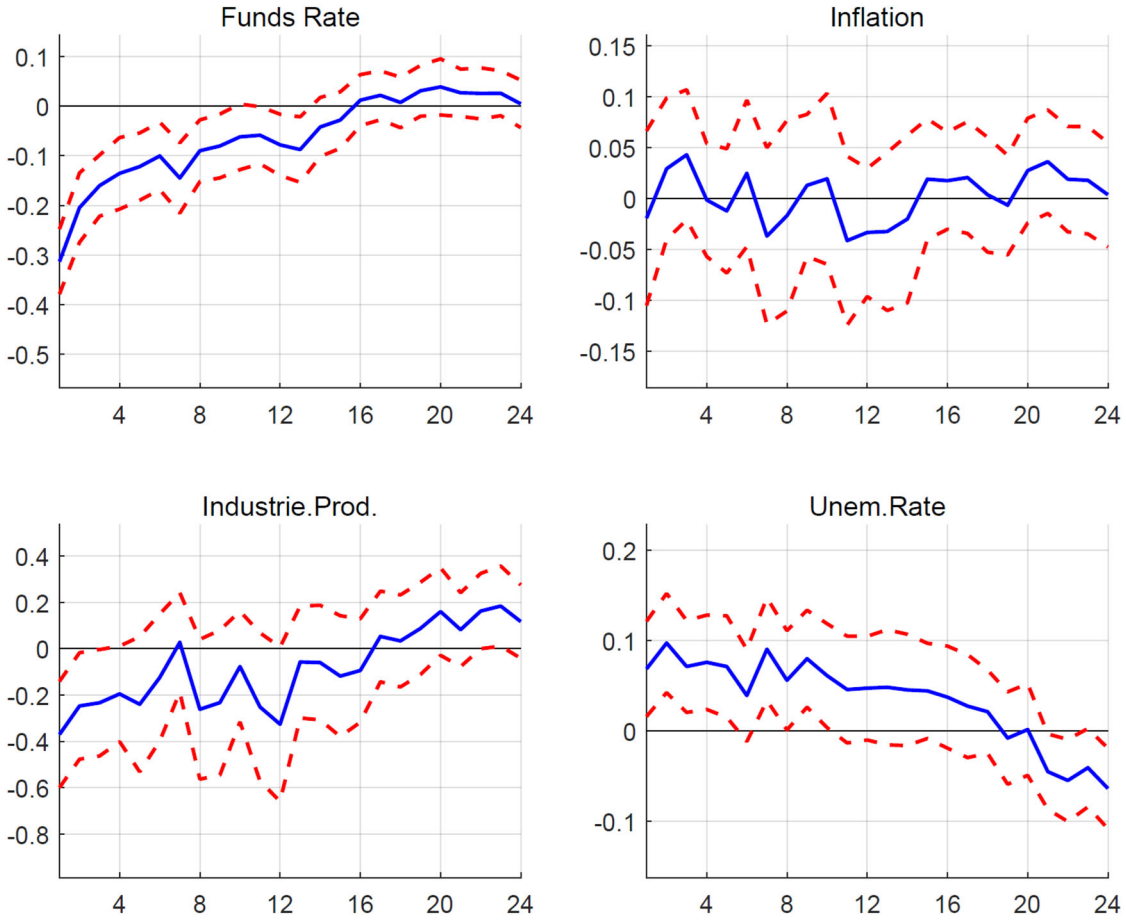

Figure 5. Estimated ATEs of target rate changes on funds rate, inflation, industrial production, and unemployment rate.
NOTE: The solid blue lines represent changes measured in percent of the level of inflation and industrial production and as point changes in the rates for the federal funds rate and the unemployment rate. Dashed lines indicate 95% confidence bands.

and unemployment rate.[7] Here, we focus on re-analyzing the setup in Section 4 of AKJ, where they studied the effects of the monetary policies before the 2008 financial crisis. For this period, the data start from August 1989 to July 2007 and contain 242 observations.

As for the propensity score function, AKJ implemented an (ordered) Probit model, choosing covariates based on the Taylor rule (see Section 3.2 of AKJ for details). In contrast, we specify our propensity score function nonparametrically, incorporating a functional covariate X and a conventional vector covariate Z. We estimate the propensity score by (3.2), which is replicated below:

$$\hat{p}(X_i, Z_i) = \frac{\sum_{j\neq i}^{n} D_j K(\|X_j - X_i\|/h) K_z((Z_j - Z_i)/h_z)}{\sum_{j\neq i}^{n} K(\|X_j - X_i\|/h) K_z((Z_j - Z_i)/h_z)}. \tag{3.2}$$

The vector of covariates Z includes the market expectation, which is an adjusted difference between the futures contract price and the current target rate (see AJK and their supplemental data appendix for more details of this measure), inflation, changes of unemployment rates, lagged inflations, lagged changes of unemployment rates, amounting to a five-dimensional vector. Thus, we heuristically choose the bandwidth $h_z = r_z \cdot n^{-\frac{1}{d+4}}$, where $d = 5$ and r_z is the sample range of $\{Z_i\}_{i=1}^n$.

To acknowledge the influence of historical federal funds rates on current decisions, AKJ also incorporate a scalar variable, which indicates the change in the previous month's target rates, in the propensity score model. In contrast, we include a functional covariate X, which is a curve that records the entire *federal funds effective rates* (FFER) from the previous month.[8] The FFER can be regarded as the real-world realization of the targets set by the FOMC. Since it is reasonable to expect that the FOMC would take into account both the level and the change of the historical FFER when deciding target rates, this motivates us to use the Sobolev norm, $\|x\|_{W^{k,p}} = \left(\sum_{0\leq\alpha\leq k} \int |\partial^\alpha x|^p dx\right)^{1/p}$ for $k = 1$ and $p = \infty$, to measure the distance between different realizations of the functional covariate X. We chose the bandwidth $h = r_x \cdot n^{-\frac{1}{5}}$, where r is the sample range of $\{\sup_t X_i(t)\}_{i=1}^n$.

The estimation results are summarized in Table 5 and Figure 5. They generally replicate Table 4 and Figure 4 of AKJ, using our approach based on the nonparametrically estimated

[7] We are grateful to Joshua Angrist for sharing the data and codes on his website, *https://economics.mit.edu/people/faculty/josh-angrist/angrist-data-archive.*

[8] The data for the FFER is downloaded from *https://fred.stlouisfed.org*, and they are sampled on a daily basis.

Table 5. Estimated ATEs at horizons 6, 12, 18, and 24 months based on data from August 1989 to July 2007.

Month	Funds rate	Inflation	Indust.Prod	Unem.Rate
6	−0.100 [−0.168, −0.032]	0.025 [−0.047, 0.097]	−0.124 [−0.398, 0.150]	0.039 [−0.012, 0.090]
12	−0.078 [−0.140, −0.016]	−0.033 [−0.096, 0.030]	−0.326 [−0.659, 0.007]	0.047 [0.105, −0.015]
18	0.008 [−0.043, 0.058]	0.004 [−0.053, 0.060]	0.034 [−0.165, 0.232]	0.021 [−0.025, 0.068]
24	0.005 [−0.043, 0.052]	0.004 [−0.047, 0.055]	0.117 [−0.042, 0.275]	−0.064 [−0.108, −0.019]

NOTE: Reported values represent changes, measured in percent for the levels of inflation and industrial production, and as point changes for the federal funds rate and the unemployment rate, each accompanied by their 95% confidence intervals.

propensity score with functional covariates. Specifically, Table 5 presents the estimated ATEs of monetary expansion policies on the FFER, inflation, industrial production, and the unemployment rate, measured at 6, 12, 18, and 24 months; the corresponding CIs are also presented. Meanwhile, Figure 5 displays the curves of these ATEs over the horizon from 1 to 24 months, along with their corresponding 95% confidence bands. There are several differences between our reports and those of AKJ: we report the *level* impulse responses rather than the *cumulated* impulse responses reported in AKJ; we present 95% confidence intervals instead of standard deviations. The estimation results show that monetary expansion (lowering the federal target rate) has an immediate effect on reducing the FEER, but this impact gradually diminishes over time. The expansion policy also increases industrial production and decreases the unemployment rate in the long run.

Supplementary Materials

The supplementary materials provide the proofs of the theoretical results in the main text and related lemmas.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Kurisu acknowledges financial support by JSPS KAKENHI Grant Numbers 20K13468 and 23K12456.

References

Angrist, J. D., Jordà, Ò., and Kuersteiner, G. M. (2018), "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited," *Journal of Business & Economic Statistics*, 36, 371–387. [11]

Angrist, J. D., and Kuersteiner, G. M. (2011), "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score," *Review of Economics and Statistics*, 93, 725–747. [11]

Bagliano, F. C., and Favero, C. A. (1998), "Measuring Monetary Policy with VAR Models: An Evaluation," *European Economic Review*, 42, 1069–1112. [11]

Benatia, D., Carrasco, M., and Florens, J.-P. (2017), "Functional Linear Regression with Functional Response," *Journal of Econometrics*, 201, 269–291. [2]

Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. (2007), "Non-Parametric Confidence Bands in Deconvolution Density Estimation," *Journal of the Royal Statistical Society*, Series B, 69, 483–506. [9]

Boente, G., and Vahnovan, A. (2017), "Robust Estimators in Semi-Functional Partial Linear Regression Models," *Journal of Multivariate Analysis*, 154, 59–84. [1]

Cattaneo, M. D. (2010), "Efficient Semiparametric Estimation of Multi-Valued Treatment Effects under Ignorability," *Journal of Econometrics*, 155, 138–154. [7]

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999), "Monetary Policy Shocks: What Have We Learned and to What End?" in *Handbook of Macroeconomics* (vol. 1), ed. J. B. Taylor, and M. Woodford, pp. 65–148, Amsterdam, The Netherlands: Elsevier Science B.V. [11]

Ciarleglio, A., Petkova, E., Ogden, T., and Tarpey, T. (2018), "Constructing Treatment Decision Rules based on Scalar and Functional Predictors When Moderators of Treatment Effect Are Unknown," *Journal of the Royal Statistical Society*, Series C, 67, 1331–1356. [2]

Farrell, M. H. (2015), "Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations," *Journal of Econometrics*, 189, 1–23. [4,7]

Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2010), "Rate of Uniform Consistency of Nonparametric Estimates with Functional Variables," *Journal of Statistical Planning and Inference*, 140, 335–352, [1,4,5]

Ferraty, F., and Nagy, S. (2022), "Scalar-on-Function Local Linear Regression and Beyond," *Biometrika*, 109, 439–455. [4]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis*, New York: Springer. [1,2,3,4,5,11]

Florens, J.-P., and Van Bellegem, S. (2015), "Instrumental Variable Estimation in Functional Linear Models," *Journal of Econometrics*, 186, 465–476. [2]

Geenens, G. (2011), "Curse of Dimensionality and Related Issues in Nonparametric Functional Regression," *Statistics Surveys*, 5, 30–43. [3]

Hahn, J. (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331. [6]

Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [1]

Hsing, T., and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Chichester: Wiley. [1]

Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference*, Cambridge: Cambridge University Press. [1]

Jiang, J., Kelly, B., and Xiu, D. (2023) "(Re') Imag (in) ing Price Trends," *The Journal of Finance*, 78, 3193–3249. [2]

Li, Q., Racine, J. S., and Wooldridge, J. M. (2009), "Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data," *Journal of Business & Economic Statistics*, 27, 206–223. [5,7]

Kokoszka, P., and Reimherr, M. (2017), *Introduction to Functional Data Analysis*, Boca Raton, FL: CRC Press. [1]

Lin, Z., Kong, D., and Wang, L. (2023), "Causal Inference on Distribution Functions," *Journal of the Royal Statistical Society*, Series B, 85, 378–398. [1]

Miao, R., Xue, W., and Zhang, X. (2020), "Average Treatment Effect Estimation in Observational Studies with Functional Covariates," working paper. [1]

Ramsey, J. O., and Silverman, B. W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer. [1,2]

Romer, C. D., and Romer, D. H. (1989), "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, 4, 121–170. [11]

Singh, R., Xu, L., and Gretton, A. (2024), "Kernel Methods for Causal Functions: Dose, Heterogeneous and Incremental Response Curves," *Biometrika*, 111, 497–516. [1,5]

van der Vaart, A. W., and van Zanten, J. H. (2007), "Bayesian Inference with Rescaled Gaussian Process Priors," *Electronic Journal of Statistics*, 1, 433–448. [5]

Wong, R. K., and Chan, K. C. G. (2018), "Kernel-based Covariate Functional Balancing for Observational Studies," *Biometrika*, 105, 199–213. [1]

Zhao, Q. (2019), "Covariate Balancing Propensity Score by Tailored Loss Functions," *Annals of Statistics*, 47, 965–993. [1]

Zhao, Y., Luo, X., Lindquist, M., and Caffo, B. (2018), "Functional Mediation Analysis with an Application to Functional Magnetic Resonance Imaging Data," arXiv:1805.06923. [2]

Zhou, D.-X. (2002), "The Covering Number in Learning Theory," *Journal of Complexity*, 18, 739–767. [5]