Computational Statistics (2022) 37:591–611
https://doi.org/10.1007/s00180-021-01142-y

ORIGINAL PAPER

Stochastic functional linear models and Malliavin calculus

Ruzong Fan[1] · **Hong-Bin Fang**[1]

Received: 30 September 2020 / Accepted: 4 August 2021 / Published online: 27 August 2021

Abstract

In this article, we study stochastic functional linear models (SFLM) driven by an underlying square integrable stochastic process $X(t)$ which is generated by a standard Brownian motion. Utilizing the magnificent Itô integrals and Malliavin calculus, $X(t)$ is expanded into a summation of orthogonal multiple integrals, i.e., Wiener-Itô chaos expansions, which is the counterpart of the Taylor expansion of deterministic functions. Based on the expansion, we show that the fourth moments of linear functionals of underlying stochastic process $X(t)$ are bounded by the square of their second moments when $X(t)$ is a finite linear combination of multiple Itô integrals. Therefore, an optimal minimax convergence rate in mean prediction risk of SFLM is valid if eigenvalues of related linear operators are of order k^{-2r} by using results in literature when the underlying process $X(t)$ is a linear combination of multiple Itô integrals. A sufficient and necessary condition of finite fourth moment of random functions of multiple Itô integrals is proved, which is a key condition in methodology and convergence rates of functional linear regressions. Our results show that the optimal minimax convergence rate in mean prediction risk can be applied to the class of linear combination of multiple Itô integrals which are not necessarily Gaussian processes. Moreover, the sufficient and necessary condition of finite fourth moment for multiple Itô integrals can be directly applied to show methodology and convergence rates of functional linear models. Using the theory of stochastic analysis, one may construct a reproducing kernel Hilbert space (RKHS) associated with a square integrable stochastic process to facilitate analysis of functional data.

Keywords Itô integrals · Malliavin calculus · Wiener-Itô chaos expansions · Functional linear models · Minimax · Optimal convergence rate · Reproducing kernel Hilbert space · Smoothing splines · Sobolev space

✉ Ruzong Fan
rf740@georgetown.edu

[1] Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center, Washington, DC 20057, USA

1 Introduction

For high-dimension data such as those of next-generation sequencing, functional data analysis technique has been found to be very useful in compressing and drawing valid information in data analysis. In literature, various decomposition methods are proposed to study functional regressions such as functional principal component analysis (FPCA) (Hall and Hosseini-Nasab 2006; Wang et al. 2016; Yao et al. 2005a, b). Functional canonical correlation analysis (FCCA) is developed to measure functional correlation (Eubank and Hsing 2008; He et al. 2003, 2010). In addition, reproducing kernel Hilbert space (RKHS) is widely used in dissecting functional data (Hsing and Ren 2009; Yuan and Cai 2010).

In comparison, the progress of theoretical research using stochastic analysis to functional models is not very fast and mature. The sophisticated theory of stochastic processes is not fully utilized in the theoretical dissection of functional data. To our knowledge, the utilization of stochastic integrations in statistical analysis is still not beyond the well-respected work of Dr. Wahba about 50 years ago, i.e., single layer Itô integral of Brownian motion in splines (Kimeldorf and Wahba 1970a, b, 1971; Wahba 1978, 1990). It is well-known that there is a family of zero-mean Gaussian random variables associated with a positive-definite function in the continuous case by Kolmogorov consistency theorem (Kolmogorov 1933). Statisticians are familiar with RKHS associated with positive-definite function but not very much research is done using the theory of stochastic processes. The gap is due to a lack of communications between statistics and stochastic analysis (or statisticians and probabilists).

In stochastic analysis, there has been three major milestones in 20th century: Itô integrals, Doob-Meyer decomposition of super-martingales, and Malliavin calculus. Itô integrals and Doob-Meyer decomposition of super-martingales are successfully applied to statistics such as survival analysis. Malliavin calculus, however, is not widely known for statisticians. One core of Malliavin calculus is a Wiener-Itô chaos expansion of random functions, which is similar to the Taylor expansion of deterministic functions. Originally, the Malliavin calculus is developed as an infinite-dimensional differential calculus on the Wiener space to provide a probabilistic proof of Hörmander's hypoellipticity theorem (Hairer 2011; Hörmander 1967; Malliavin 1978). It defines derivatives of random variables on the Wiener space and then it develops related calculus, which is used to decompose square integrable random variables and square integrable stochastic processes into an orthogonal summations of multiple Itô integrals.

In the last 40 years, the Malliavin calculus has found many fruitful applications, such as application to show the existence and smoothness of densities of stochastic functionals of Gaussian processes and application to finance (Fournié et al. 1999, 2001; Nualart 2018; Ocone 1984; Ocone and Karatzas 1991). To study probabilistic approximations, Malliavin calculus is combined into Stein's methods to provide a complete characterization of Gaussian approximations and central limiting theorems for sequences of multiple stochastic integrals (Nourdin and Peccati 2009, 2012; Nualart and Peccati 2005; Nualart and Ortiz-Latorre 2008).

In this paper, we apply the Malliavin calculus and multiple Itô integrals to analyze functional data. We study stochastic functional linear models (SFLM) to connect a scalar random variable with a square integrable stochastic process $X(t)$ which is generated by a standard Brownian motion. The SFLM naturally explains that the variation of the scalar variable is intrinsically driven by the Brownian motion (or the stochastic process $X(t)$) and is influenced by errors. We show that the fourth moments of linear functionals of underlying stochastic process $X(t)$ are bounded by the square of their second moments when $X(t)$ is a finite linear combination of multiple Itô integrals. Therefore, an optimal minimax convergence rate in mean prediction risk of SFLM is valid if eigenvalues of related linear operators are of order k^{-2r}, $r > 0$, by using results in literature when $X(t)$ is a linear combination of multiple Itô integrals (Cai and Hall 2006; Cai and Yuan 2012; Yuan and Cai 2010; Crambes et al. 2009). We provide a sufficient and necessary condition for fourth moment to be finite, which is a key condition in methodology and convergence rates of functional linear regressions (Ferré and Yao 2003, 2005; Delaigle and Hall 2012; Hall and Horowitz 2007; Li and Hsing 2007, 2010).

The reminder of this article is organized as follows. In Sect. 2, we introduce the SFLM and establish a relation between the SFLM and minimization procedure to estimate the mean of the scalar random variable with a stochastic functional component. In Sect. 3, Itô integrals and Wiener-Itô chaos expansions of Malliavin calculus are briefly introduced. In Sect. 4, we provide theoretical results of stochastic functional linear models: we show an optimal minimax convergence rate in mean prediction risk in Sects. 4.1 and 4.2, we provide a sufficient and necessary condition of finite fourth moment with applications to functional inverse regression and rates of convergence. Section 5 presents results of simulation studies. We conclude with discussion and some remarks in Sect. 6.

2 Stochastic functional linear models

Let $W = W(t) = W(t, \omega), \omega \in \Omega, t \in [0, 1]$, be a standard Brownian motion or a one-dimensional Wiener process on a complete probability space $(\Omega, \mathcal{F}, P)$, where the σ-field $\mathcal{F}$ is given by $\mathcal{F} = \sigma\{W(t) : 0 \le t \le 1\}$. Let $x_0(t)$ be a square integrable function defined on $[0, 1]$, $X(t, \omega)$ be a square integrable stochastic process defined on $[0, 1]$ satisfying $\mathrm{E}X(t) = x_0(t)$ and $\int_0^1 \mathrm{E}X^2(t)dt < \infty$. For all $t \in [0, 1]$, assume that $X(t)$ is $\mathcal{F}$-measurable. Let Y be a scalar random variable generated by a functional linear model

$$\begin{aligned} Y &= \alpha + \int_0^1 X(t)\beta(t)dt + \varepsilon, \\ &= \alpha + \int_0^1 x_0(t)\beta(t)dt + \int_0^1 [X(t) - x_0(t)]\,\beta(t)dt + \varepsilon, \end{aligned} \tag{1}$$

where α is an intercept, $\beta(t) \in H = L^2([0, 1])$ is a square integrable regression coefficient function, and ε is an error term such that $\mathrm{E}\varepsilon = 0$ and $\mathrm{E}\varepsilon^2 = \sigma_e^2 < \infty$. Assume that ε is independent of $W(t), t \in [0, 1]$. Denote the covariance function of $X(t)$ by $C(s, t) = \mathrm{Cov}[X(s), X(t)]$.

Given a sample of data $((X_1, Y_1), \dots, (X_n, Y_n)$ consisting of n independent copies of (X, Y), let us denote

$$\eta(X_i) = \alpha + \int_0^1 X_i(t)\beta(t)dt.$$

Thus, the model (1) can be re-written as $Y_i = \eta(X_i)+\varepsilon_i$, where ε_i are independent error terms of copies of ε, $i = 1, \dots n$. An estimate $\hat{\eta}_{n\lambda}(x)$ of $\eta(x)$ is obtained by finding $\hat{\alpha} \in R$ and $\hat{\beta}(t) \in H$ to minimize $\frac{1}{n}\sum_{i=1}^{n}[Y_i - \eta(X_i)]^2 + \lambda J(\beta) = \ell_n(\eta) + \lambda J(\beta)$, where $\ell_n(\eta) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \eta(X_i)]^2$, $\lambda > 0$ is a smoothing parameter, J is roughness penalty on β (Cai and Yuan 2012; Li and Hsing 2007; Yuan and Cai 2010). Therefore, $\hat{\eta}_{n\lambda}$ is a regularization estimate of η given by

$$\hat{\eta}_{n\lambda} = \arg\min_{\alpha\in R,\beta\in H}\left\{\frac{1}{n}\sum_{i=1}^{n}[Y_i - \eta(X_i)]^2 + \lambda J(\beta)\right\}. \tag{2}$$

Let (X_{n+1}, Y_{n+1}) be a new copy of (X, Y) which is independent of $((X_1, Y_1), \dots, (X_n, Y_n))$. Given the estimate $\hat{\eta}_{n\lambda}(x)$, the prediction accuracy can be measured by the excess risk (Cai and Yuan 2012)

$$\begin{aligned}\mathcal{E}(\hat{\eta}_{n\lambda}) &= \mathrm{E}^*\big[Y_{n+1} - \hat{\eta}_{n\lambda}(X_{n+1})\big]^2 - \mathrm{E}^*\big[Y_{n+1} - \eta(X_{n+1})\big]^2\\ &= \mathrm{E}^*\big[\hat{\eta}_{n\lambda}(X_{n+1}) - \eta(X_{n+1})\big]^2,\end{aligned} \tag{3}$$

where E^* represents an expectation taken over (X_{n+1}, Y_{n+1}). The relation (3) can be verified since $Y_{n+1} = \eta(X_{n+1}) + \varepsilon_{n+1}$ and ε_{n+1} is an error term which is independent of ε_i, $i = 1, \dots n$.

Let the penalty functional J be a squared semi-norm on H such that the null space $H_0 = \{\beta \in H : J(\beta) = 0\}$ is a finite subspace of H with orthonormal basis $\{\xi_1(t), \dots, \xi_N(t)\}$. Let H_K be the orthogonal complement of H_0 such that $H = H_0 \oplus H_K$. For $h \in H$, it can be uniquely decomposed into $h = h_0 + h_1$ such that $h_0 \in H_0$ and $h_1 \in H_K$ are projections onto H_0 and H_K. Let $K(\cdot, \cdot)$ be the reproducing kernel (RK) of H_K such that $J(h_1) = ||h_1||_K^2$.

Note $H_K \subset H$ is the RKHS of $K(s, t)$, i.e., H_K is the closed subspace of H formed by all finite linear combinations of form $\sum_i a_i K_{t_i}(t)$, where $K_{t_i}(t) = K(t_i, t)$. For $f(\cdot) = K_s(\cdot)$ and $g(\cdot) = K_t(\cdot)$ in H_K, their inner product is

$$\langle f(\cdot), g(\cdot)\rangle_K = \langle K_s(\cdot), K_t(\cdot)\rangle_K = K(s, t),$$

which leads to $\langle K_t(\cdot), f(\cdot)\rangle_K = f(t)$. The following representer theorem is from literature (Cai and Yuan 2012; Yuan and Cai 2010).

Theorem 2.1 *Assume that $K(s, t)$ is continuous and the ℓ_n depends on η only through $\eta(X_1), \eta(X_2), \dots, \eta(X_n)$. Then there exist $d = (d_1, \dots, d_N)$ and $c = (c_1, \dots, c_n)$ such that*

$$\hat{\beta}_{n\lambda}(t) = \sum_{k=1}^{N} d_k \xi_k(t) + \sum_{j=1}^{n} c_j L_K X_j(t),$$
$$\hat{\alpha}_{n\lambda} = \bar{Y} - \int_0^1 \bar{X}(t)\hat{\beta}_{n\lambda}(t)dt,$$

where $\bar{Y} = \frac{1}{n}\sum_{k=1}^{n} Y_i$, $\bar{X}(t) = \frac{1}{n}\sum_{k=1}^{n} X_k(t)$, *and* $L_K X_j(t)$ *is an operator defined via*

$$L_K h(t) = \int_0^1 K(t,s)h(s)ds, h \in H.$$

3 A brief introduction of Malliavin calculus

Let $H = L^2([0,1])$ be a Hilbert space defined by an inner product and an norm $\|\cdot\|_H$

$$\langle g, h\rangle_H = \int_0^1 g(t)h(t)dt \quad \text{and} \quad \|g\|_H^2 = \langle g, g\rangle_H.$$

For each $h \in H$, let us denote the Wiener integral of h by $W(h) = \int_0^1 h(t)dW_t$. Note that

$$\mathrm{E}[W(h)W(g)] = \langle g, h\rangle_H.$$

Hence, the mapping $h \longrightarrow W(h)$ can be extended to a linear isometry between the Hilbert space H and the Gaussian space $L^2(\Omega, \mathcal{F}, P)$.

For $n \geq 1$, let $H^{\otimes n} = L^2([0,1]^n)$ be a Hilbert space defined by an inner product

$$\langle g, h\rangle_{H^{\otimes n}} = \int_{[0,1]^n} g(t_1,\ldots,t_n)h(t_1,\ldots,t_n)dt_1\ldots dt_n.$$

Let $L_S^2([0,1]^n) \subset L^2([0,1]^n)$ be a space of symmetric square integrable Borel real functions. For any $h \in L_S^2([0,1]^n)$, we have

$$\|h\|_{H^{\otimes n}}^2 = n! \int_{\Delta_n} h^2(t_1,\ldots,t_n)dt_1\ldots dt_n,$$

where $\Delta_n = \{(t_1,\ldots,t_n) \in [0,1]^n : 0 \leq t_1 \leq \cdots \leq t_n \leq 1\}$. For $g \in L_S^2([0,1]^n)$, define its multiple stochastic integral as an Itô integral (Itô 1951)

$$I_n(g) = n! \int_0^1 \int_0^{t_n} \cdots \int_0^{t_2} g(t_1,\ldots,t_n)dW_{t_1}\ldots dW_{t_n}$$
$$= n! \int_{\Delta_n} g(t_1,\ldots,t_n)dW_{t_1}\ldots dW_{t_n}.$$

From Itô (1951) or Nualart (2006, 2018) or Nunno et al. (2009), we have for $g, h \in L_S^2([0,1]^n)$

$$\mathrm{E}[I_n(g)I_m(h)] = \begin{cases} 0 & \text{if } n \neq m \\ n!\langle g, h\rangle_{H^{\otimes n}} & \text{if } n = m \end{cases}. \tag{4}$$

If $g \in L^2([0,1]^n)$ is not necessarily symmetric, define its symmetrization as

$$\tilde{g}(t_1, \ldots, t_n) = \frac{1}{n!}\sum_{\sigma} g(t_{\sigma_1}, \ldots, t_{\sigma_n}),$$

where the sum is taken over all permutations σ of $(1, 2, \ldots, n)$. Then, we may define a stochastic integral $I_n(g) = I_n(\tilde{g})$. The following Wiener-Itô decomposition is well-known in the theory of Malliavin calculus (Itô 1951; Nunno et al. 2009; Oksendal 2003; Wiener 1938).

Lemma 3.1 *Let $F(\omega) \in L^2(\Omega, \mathcal{F}, P)$ be a square integrable random variable. There exists a constant f_0 and a sequence of symmetric square integrable functions $f_n(t_1, \ldots, t_n) \in L_S^2([0,1]^n)$, $n = 1, 2, \ldots$, such that*

$$F = f_0 + \sum_{n=1}^{\infty} I_n(f_n). \tag{5}$$

In addition, the sum of (5) converges in $L^2(\Omega, \mathcal{F}, dP)$ and

$$\mathrm{E}(F^2) = f_0^2 + \sum_{n=1}^{\infty} \mathrm{E}I_n^2(f_n) \;=\; f_0^2 + \sum_{n=1}^{\infty} n!\|f_n\|_{H^{\otimes n}}^2 \;<\; \infty.$$

One may note that the mean of the random variable F in Lemma 3.1 is equal to f_0 and the variance of F is $\sum_{n=1}^{\infty} n!\|f_n\|_{H^{\otimes n}}^2$.

Lemma 3.2 *Let $X(t)$ be a measurable stochastic process and $x_0(t)$ be a square integrable function on $[0,1]$ such that $\int_0^1 \mathrm{E}X^2(t)dt < \infty$ and $\mathrm{E}X(t) = x_0(t)$, i.e., $X(t)$ is square integrable with a mean function $x_0(t)$. There exists a sequence of deterministic measurable kernels $x_n(t_1, \ldots, t_n, t) \in L^2([0,1]^{n+1})$ on $[0,1]^{n+1}$, $n = 1, 2, \ldots$, such that*

$$X(t) = x_0(t) + \sum_{n=1}^{\infty} I_n(x_n(\cdot, t)) \tag{6}$$

and all $x_n(t_1, \dots, t_n, t)$ *are symmetric with respect to the variables* $t_1, \dots, t_n$. *In addition, the sum of (6) converges in* $L^2(\Omega \times [0, 1], dP \times dt)$ *and*

$$\begin{aligned} \mathrm{E}\|X\|_H^2 &= \int_0^1 \mathrm{E}X^2(t)dt \\ &= \int_0^1 x_0^2(t)dt + \sum_{n=1}^{\infty} \int_0^1 \mathrm{E}I_n^2(x_n(\cdot, t))dt \\ &= \int_0^1 x_0^2(t)dt + \sum_{n=1}^{\infty} n!\|x_n\|_{H^{\otimes(n+1)}}^2. \end{aligned}$$

4 Theoretical results

4.1 Optimal rates of convergence

Let $K(s, t)$ be the RK of the penalty functional $J(\beta)$ introduced in the Sect. 2. By the spectral theorem, there exists a complete orthonormal system $\{\phi_1, \phi_2, \dots\} \in H_K$ such that $L_K\phi_i = \kappa_i\phi_i$, where $\kappa_i \geq 0$ is eigenvalue corresponding to ϕ_i, $i = 1, 2 \dots$. Moreover, $\kappa_1 \geq \kappa_2 \geq \cdots > 0$. Let $L_{K^{1/2}}$ be a linear operator defined by $L_{K^{1/2}}\phi_i = \sqrt{\kappa_i}\phi_i$.

For the RK $K(s, t)$ and the covariance function $C(s, t)$, define a linear operator $L_{K^{1/2}CK^{1/2}}$ by $L_{K^{1/2}CK^{1/2}}h = L_{K^{1/2}}(L_C(L_{K^{1/2}}h))$, where L_C is an operator defined via

$$L_C h(t) = \int_0^1 C(t, s)h(s)ds, \, h \in H.$$

By the spectral theorem, there exists a complete orthonormal system $\{\psi_1, \psi_2, \dots\}$ such that $L_{K^{1/2}CK^{1/2}}\psi_i = s_i\psi_i$, where $s_i \geq 0$ is eigenvalue corresponding to ψ_i, $i = 1, 2 \dots$. Moreover, $s_1 \geq s_2 \geq \cdots > 0$.

For two positive sequences a_k and b_k, $a_k \asymp b_k$ means that a_k/b_k is bounded away from 0 and ∞ as $k \to \infty$. In Cai and Yuan (2012), an optimal minimax convergence rate in mean prediction risk is proved under an assumption that the fourth moments of linear functionals of $X(t)$ are bounded by the square of their second moments. In the following theorem, we show the result without the assumption if $X(t)$ is a linear combination of multiple Itô integrals.

Theorem 4.1 *Assume that* $X(t) = x_0(t) + \sum_{n=1}^{N} I_n(x_n(\cdot, t))$, $N \geq 1$, *is a square integrable stochastic process on* $(\Omega, \mathcal{F}, P)$, $\mathcal{F} = \sigma\{W(t) : 0 \leq t \leq 1\}$. *Let* $\mathcal{E}(\hat{\eta}_{n\lambda})$ *be the mean prediction risk defined by the relation (3). Suppose the eigenvalues* $\{s_k : k \geq 1\}$ *of the linear operator* $L_{K^{1/2}CK^{1/2}}$ *satisfy* $s_k \asymp k^{-2r}$ *for some constant* $0 < r < \infty$, *then*

$$\lim_{A\to\infty} \lim_{n\to\infty} \sup_{\beta\in H(K)} P\left\{\mathcal{E}(\hat{\eta}_{n\lambda}) \geq An^{-\frac{2r}{2r+1}}\right\} = 0,$$

when λ is of order $n^{-2r/(2r+1)}$.

Proof By Theorem 2 of Cai and Yuan (2012), we only need to show that there exists a constant C_N such that

$$\mathrm{E}\left(\int_0^1 X(t)f(t)dt\right)^4 dt \le C_N\left[\mathrm{E}\left(\int_0^1 X(t)f(t)dt\right)^2\right]^2, \tag{7}$$

for any square integrable function $f(t) \in H$. The proof of the inequality (7) is provided in section A of the Supplementary Materials by three lemmas, i.e., Lemma A.1, Lemma A.2, and Lemma A.3. The first two Lemma A.1 and Lemma A.2 are used to prove Lemma A.3 which shows the inequality (7). ♣

Remark 4.1 The Theorem 2.1 is slightly different from that in Cai and Yuan (2012), since we use the framework in Yuan and Cai (2010). The two papers use slightly different penalties. In Yuan and Cai (2010), in keeping with splines, the penalty is a squared semi-norm on the RKHS, i.e., it does not penalize constant or linear terms of H_0 as represented by the $\{\xi_i, \ldots, \xi_N(t)\}$. Cai and Yuan (2012) uses a penalty that is the squared norm on the RKHS so it penalizes everything. As the un-penalized subspace H_0 is finite dimensional, it does not lead to any difference in the results. The similar frameworks as Yuan and Cai (2010) are used in Du and Wang (2014) and Sun et al. (2018). Therefore, the difference of the two frameworks is that a finite dimensional null space H_0 is included in our framework as Yuan and Cai (2010), while $H_0 = \emptyset$ is empty in Cai and Yuan (2012).

Remark 4.2 The relation (7) states that the fourth moments of linear functionals of $X(t)$ are bounded by the square of their second moments and so have bounded kurtosis. When $N = 1$, the stochastic process $X(t) = x_0(t) + I_1(x_1(\cdot, t))$ is a Gaussian process and the inequality (7) is an equality and $C_1 = 3$. When $N > 1$, $X(t) = x_0(t) + \sum_{n=1}^{N} I_n(x_n(\cdot, t))$ is not a Gaussian process and $c_N > 3$. To show the inequality (7), we will use the theory of Itô integrals and Malliavin calculus to show the three lemmas in the Supplementary Materials.

4.2 Sufficient and necessary condition of finite fourth moments

The following Theorem provides a sufficient and necessary condition for fourth moment to be finite, which is a key condition in methodology and convergence rates of functional linear regressions (Delaigle and Hall 2012; Hall and Horowitz 2007).

Theorem 4.2 *Assume that $X(t) = I_n(x_n(\cdot, t)), n \ge 1$. Then X has finite fourth moment in that $\mathrm{E}\int_0^1 X^4(t)dt < \infty$ if and only if*

$$\int_0^1 \|x_n(\cdot, t)\|^4_{H^{\otimes n}} dt = \int_0^1 \left[\int_{[0,1]^n} x_n^2(t_1, \ldots, t_n, t)dt_1 \ldots dt_n\right]^2 dt < \infty.$$

The proof of this Theorem is similar to that of Lemma A.2. For completion of presentation, we provide the details in section B of the **Supplementary Materials**.

Li and Hsing (2007) requires a finite fourth moment of m-th derivative $X^{(m)}$. The following result provides a necessary and sufficient condition for finite fourth moment of $X^{(m)}$ which can be proved as Theorem 4.2.

Theorem 4.3 *Assume that $X(t) = I_n(x_n(\cdot, t)), n \geq 1$, and $x_n(t_1, \ldots, t_n, t)$ belongs to the Sobolev space of order m*

$$\mathcal{W}_2^{n,m} = \left\{ g \in H^{\otimes(n+1)}) : g \text{ is } m\text{-times differentiable with respect to } t \text{ and } \frac{\partial^m g}{\partial t^m} \in H^{\otimes^{n+1}} \right\}.$$

Then $X^{(m)}(t) = I_n\left(\frac{\partial^m x_n}{\partial t^m}(\cdot, t)\right)$ has finite fourth moment in that $\mathrm{E}\int_0^1 \left[X^{(m)}(t)\right]^4 dt < \infty$ if and only if

$$\int_0^1 \left\| \frac{\partial^m x_n}{\partial t^m}(\cdot, t) \right\|_{H^{\otimes n}}^4 dt = \int_0^1 \left[\int_{[0,1]^n} \left(\frac{\partial^m x_n}{\partial t^m}(t_1, \ldots, t_n, t) \right)^2 dt_1 \ldots dt_n \right]^2 dt < \infty.$$

In the Wiener-Itô chaos expansion (6), $X(t)$ is decomposed into a summation of an infinite orthogonal Itô integrals. In practice, a finite number of terms should be enough to model functional data. The following theorem extends the results of Theorems 4.2 and 4.3.

Theorem 4.4 *Assume that $X(t) = \sum_{n=1}^N I_n(x_n(\cdot, t))$, $N \geq 1$. Then,*

1. *X has finite fourth moment in that $\mathrm{E}\int_0^1 X^4(t)dt < \infty$ if for all $n = 1, \ldots, N$,*

$$\int_0^1 \|x_n(\cdot, t)\|_{H^{\otimes n}}^4 dt = \int_0^1 \left[\int_{[0,1]^n} x_n^2(t_1, \ldots, t_n, t) dt_1 \ldots dt_n \right]^2 dt < \infty.$$

2. *Assume that for all $n = 1, \ldots, N$, $x_n(t_1, \ldots, t_n, t)$ belongs to the Sobolev space $\mathcal{W}_2^{n,m}$ of order m. Then $X^{(m)}(t) = \sum_{n=1}^N I_n\left(\frac{\partial^m x_n}{\partial t^m}(\cdot, t)\right)$ has finite fourth moment in that $\mathrm{E}\int_0^1 \left[X^{(m)}(t)\right]^4 dt < \infty$ if for all $n = 1, \ldots, N$,*

$$\int_0^1 \left\| \frac{\partial^m x_n}{\partial t^m}(\cdot, t) \right\|_{H^{\otimes n}}^4 dt = \int_0^1 \left[\int_{[0,1]^n} \left(\frac{\partial^m x_n}{\partial t^m}(t_1, \ldots, t_n, t) \right)^2 dt_1 \ldots dt_n \right]^2 dt < \infty.$$

Proof We will show the first conclusion since the second one is implied by the first. Note that $X^4(t)$ consists of five type terms $I_n^4(x_n(\cdot, t))$, $I_n^3(x_n(\cdot, t))I_m(x_m(\cdot, t))$, $I_n^2(x_n(\cdot, t))I_m^2(x_m(\cdot, t))$, $I_n^2(x_n(\cdot, t))I_m(x_m(\cdot, t))I_k(x_k(\cdot, t))$, $I_n(x_n(\cdot, t))I_m(x_m(\cdot, t))$

$I_k(x_k(\cdot,t))I_\ell(x_\ell(\cdot,t))$, $n, m, k, \ell = 1, \ldots, N$, and n, m, k, ℓ are all different. Therefore, we only need to show that $\int_0^1 \mathrm{E} I_n^4(x_n(\cdot,t))dt < \infty$, $n = 1, \ldots, N$, implies that the rest four terms have finite moment, which is implied by

$$\begin{aligned}
&\int_0^1 \mathrm{E}\left[I_n^2(x_n(\cdot,t))I_m^2(x_m(\cdot,t))\right]dt \\
&\quad \le \left(\int_0^1 \mathrm{E} I_n^4(x_n(\cdot,t))dt \int_0^1 \mathrm{E} I_m^4(x_m(\cdot,t))dt\right)^{1/2} \quad < \quad \infty, \\
&\int_0^1 \mathrm{E}\left[I_n^3(x_n(\cdot,t))I_m(x_m(\cdot,t))\right]dt \\
&\quad \le \left(\int_0^1 \mathrm{E} I_n^4(x_n(\cdot,t))dt \int_0^1 \mathrm{E}\left[I_n^2(x_n(\cdot,t))I_m^2(x_m(\cdot,t))\right]dt\right)^{1/2} \quad < \quad \infty, \\
&\int_0^1 \mathrm{E}\left[I_n^2(x_n(\cdot,t))I_m(x_m(\cdot,t))I_k(x_k(\cdot,t))\right]dt \\
&\quad \le \left(\int_0^1 \mathrm{E} I_n^4(x_n(\cdot,t))dt \int_0^1 \left[\mathrm{E} I_m^2(x_m(\cdot,t))I_k^2(x_k(\cdot,t))\right]dt\right)^{1/2} \quad < \quad \infty, \\
&\int_0^1 \mathrm{E}\left[I_n(x_n(\cdot,t))I_m(x_m(\cdot,t))I_k(x_k(\cdot,t))I_\ell(x_\ell(\cdot,t))\right]dt \\
&\quad \le \left(\int_0^1 \mathrm{E}\left[I_n^2(x_n(\cdot,t))I_m^2(x_m(\cdot,t))\right]dt \int_0^1 \left[\mathrm{E} I_k^2(x_k(\cdot,t))I_\ell^2(x_\ell(\cdot,t))\right]dt\right)^{1/2} \\
&\quad < \quad \infty.
\end{aligned}$$

♣

Theorem 4.5 *Assume that* $X(t) = \sum_{n=1}^\infty I_n(x_n(\cdot,t))$. *Then,*

1. *X has finite fourth moment in that* $\mathrm{E}\int_0^1 X^4(t)dt < \infty$ *if*

$$\begin{aligned}
&\sup_{N\ge1}\left(N^3 \sum_{n=1}^N \int_0^1 \|x_n(\cdot,t)\|_{H^{\otimes n}}^4 dt\right) \\
&\quad = \left[\sup_{N\ge1} N^3 \sum_{n=1}^N \int_0^1 \left(\int_{[0,1]^n} x_n^2(t_1,\ldots,t_n,t)dt_1\ldots dt_n\right)^2 dt\right] < \infty.
\end{aligned}$$

2. *Assume that for all* $n = 1, 2, \ldots$, $x_n(t_1,\ldots,t_n,t)$ *belongs to the Sobolev space* $\mathcal{W}_2^{n,m}$ *of order m. Then* $X^{(m)}(t) = \sum_{n=1}^\infty I_n\left(\frac{\partial^m x_n}{\partial t^m}(\cdot,t)\right)$ *has finite fourth moment in that* $\mathrm{E}\int_0^1 \left[X^{(m)}(t)\right]^4 dt < \infty$ *if*

$$\sup_{N\geq 1}\left(N^3\sum_{n=1}^{N}\int_0^1\left\|\frac{\partial^m x_n}{\partial t^m}(\cdot,t)\right\|_{H^{\otimes n}}^4 dt\right)$$
$$=\sup_{N\geq 1}\left[N^3\sum_{n=1}^{N}\int_0^1\left(\int_{[0,1]^n}\left(\frac{\partial^m x_n}{\partial t^m}(t_1,\ldots,t_n,t)\right)^2 dt_1\ldots dt_n\right)^2 dt\right]<\infty.$$

Proof Note that

$$\left(\sum_{n=1}^{N} I_n(x_n(\cdot,t))\right)^4 = \left[\left(\sum_{n=1}^{N} I_n(x_n(\cdot,t))\right)^2\right]^2 \leq \left[N\sum_{n=1}^{N} I_n^2(x_n(\cdot,t))\right]^2 \leq N^3\sum_{n=1}^{N} I_n^4(x_n(\cdot,t)).$$

Taking expectation of the above relation and letting $N\to\infty$ will show the theorem. ♣

In functional sliced inverse regression and convergence rates of functional linear models, a finite fourth Hilbertian norm moment in that $\mathrm{E}\,\|X\|_H^4<\infty$ is required (Ferré and Yao 2003, 2005; Li and Hsing 2007, 2010). The following corollary shows that the finite fourth moment of $\mathrm{E}\int_0^1 X^4(t)dt<\infty$ implies $\mathrm{E}\,\|X\|_H^4<\infty$.

Corollary 4.1 *Assume that $X(t)$ is a stochastic process defined on* $[0,1]$. *Then X has finite fourth moment in that* $\mathrm{E}\int_0^1 X^4(t)dt<\infty$ *implies* $\mathrm{E}\|X\|_H^4<\infty$.

Proof Note that the conclusion holds since

$$\mathrm{E}\|X\|_H^4=\mathrm{E}\left[\int_0^1 X^2(t)dt\right]^2 \;\leq\; \mathrm{E}\int_0^1 X^4(t)dt.$$

♣

Corollary 4.2 *Assume that $X(t)$ is a stochastic process defined on* $[0,1]$, *and $X^{(m)}(t)$ exists. Then $X^{(m)}(t)$ has finite fourth moment* $\mathrm{E}\int_0^1\left[X^{(m)}(t)\right]^4 dt<\infty$ *implies* $\mathrm{E}\|X^{(m)}\|_H^4<\infty$.

In functional inverse regressions, three conditions are required (Ferré and Yao 2003, 2005; Hsing and Ren 2009; Li and Hsing 2010):

1. $\mathrm{E}\,\|X\|_H^4 < \infty$.
2. For any function $f, \beta_1, \ldots, \beta_K \in H$, there exist constants $c_0, \ldots, c_K$ such that

$$\mathrm{E}\left[\langle f, X\rangle_H \mid \langle \beta_1, X\rangle_H, \cdots, \langle \beta_K, X\rangle_H\right] = c_0 + \sum_{k=1}^{K} \langle \beta_k, X\rangle_H. \tag{8}$$

3. X has an elliptically contoured distribution.

If $X(t)$ is a Gaussian process given by $X(t) = x_0(t) + I_1(x_1(\cdot, t))$, the linear span (8) is valid due to that the projections of X are jointly normal and X has an elliptically contoured distribution. In addition, the characteristic function is given by

$$\mathrm{E}\left[\exp\left(i\langle \beta, X\rangle_H\right)\right] = \exp\left[i\langle \beta, x_0\rangle_H - \int_{[0,1]} \left(\int_0^1 x_1(t_1, t)\beta(t)dt\right)^2 dt_1\right].$$

Based on the relation (4) and Wiener-Itô chaos expansion (6), the covariance function of a square integrable stochastic process $X(t)$ is

$$\begin{aligned} C(s,t) &= \mathrm{Cov}[X(t), X(s)] \\ &= \mathrm{E}\,[X(t) - x_0(t)][X(s) - x_0(s)] \\ &= \sum_{n=1}^{\infty} n!\langle x_n(\cdot, t), x_n(\cdot, s)\rangle_{H^{\otimes n}} \\ &= \sum_{n=1}^{\infty} n! \int_{[0,1]^n} x_n(t_1, \ldots, t_n, t) x_n(t_1, \ldots, t_n, s) dt_1 \ldots dt_n. \end{aligned} \tag{9}$$

From the relation (9), the variance of random term $\int_0^1 X(t)\beta(t)dt$ is given by

$$\begin{aligned} \mathrm{Var}\left(\int_0^1 X(t)\beta(t)dt\right) &= \int_0^1 \int_0^1 \beta(s)C(s,t)\beta(t)dsdt \\ &= \sum_{n=1}^{\infty} n! \int_{[0,1]^n} \left(\int_0^1 x_n(t_1, \ldots, t_n, t)\beta(t)dt\right)^2 dt_1 \ldots dt_n. \end{aligned}$$

In Delaigle and Hall (2012), a different version of finite fourth moment is used, which is $\mathrm{E}\left[\int_0^1 \int_0^1 X(s)C(s,t)X(t)dsdt\right]^2 < \infty$. The following corollary shows that the finite fourth moment of $\mathrm{E}\|X\|_H^4 < \infty$ implies it.

Corollary 4.3 *Assume that $X(t, \omega)$ is a stochastic process defined on* $[0, 1]$. *Then X has finite fourth moment in that* $\mathrm{E}\,\|X\|_H^4 < \infty$ *implies* $\mathrm{E}\left[\int_0^1 \int_0^1 X(s)C(s,t)X(t)dsdt\right]^2 < \infty$.

Proof Note that

$$\int_0^1 \text{Var}[X(t)]dt = \int_0^1 C(t,t)dt \;=\; \sum_{n=1}^{\infty} n!\|x_n\|^2_{H^{\otimes(n+1)}} \;<\; \infty.$$

Thus, $C(s,t)$ is square integrable since -

$$\begin{aligned}\int_0^1\int_0^1 C^2(s,t)dt &= \int_0^1\int_0^1 [\text{Cov}[X(t),X(s)]]^2\,dsdt \\ &\le \int_0^1\int_0^1 \text{Var}[X(s)]\text{Var}[X(t)]dsdt \\ &= \left[\sum_{n=1}^{\infty} n!\|x_n\|^2_{H^{\otimes(n+1)}}\right]^2 < \infty. \end{aligned} \tag{10}$$

Therefore, we have

$$\begin{aligned}&\text{E}\left[\int_0^1\int_0^1 X(s)C(s,t)X(t)dsdt\right]^2 \\ &\le \text{E}\left[\left(\int_0^1\int_0^1 (X(s)X(t))^2\,dsdt\right)\left(\int_0^1\int_0^1 C^2(s,t)dsdt\right)\right] \\ &= \text{E}\;\|X\|^4_H \int_0^1\int_0^1 C^2(s,t)dsdt \;<\; \infty.\end{aligned}$$

♣

5 Simulations studies

For simplicity, the intercept α is taken as 0 in the stochastic functional linear model (1). The interval [0, 1] is partitioned by a vector of 100 equally spaced points to represent a predictor curve. Such as Yuan and Cai (2010), the true slope function $\beta(t)$ is given by $\beta(t) = \sum_{k=1}^{50} 4(-1)^{k+1}k^{-2}\phi_k(t)$, where $\phi_1(t) = 1$ and $\phi_{k+1}(t) = \sqrt{2}\cos(k\pi t)$ for $k \ge 1$. For the stochastic process $X(t)$ in the model (1), we simulate data using four processes:

- A standard Brownian motion, i.e., $X(t) = W(t)$. The covariance function of $X(t)$ is $C(s,t) = \min(s,t)$.
- An Ornstein-Uhlenbeck process, which is given by the following stochastic differential equation (Bishwal 2007; Bouleau and Lepingle 1992; Iacus 2008)

$$dX(t) = -\theta_1 X(t)dt + \theta_2 dW(t), \tag{11}$$

  where $\theta_1 = 1$ and $\theta_2 = 5$. The covariance function of $X(t)$ is $C(s,t) = \theta_2^2\exp(-\theta_1|t-s|)/(2\theta_1)$.

- A non-Gaussian process $X(t) = \int_0^t dW(v) \int_0^v (\exp(-\gamma v) + \exp(-\gamma u))\, dW(u)$, $\gamma = 0.75$. If $t \le s$, the covariance function of $X(t)$ is given by (Pavliotis 2014; Ikeda and Watanabe 1989)

$$\begin{aligned} C(s,t) &= \int_0^t dv \int_0^v \left[\exp(-\gamma v) + \exp(-\gamma u)\right]^2 du \\ &= \frac{1}{2}\int_0^t dv \int_0^t \left[\exp(-2\gamma v) + 2\exp(-\gamma v - \gamma u) + \exp(-2\gamma u)\right] du \\ &= \int_0^t dv \int_0^t \exp(-2\gamma v) du + \left[\int_0^t \exp(-\gamma u) du\right]^2 \\ &= \frac{t}{2\gamma}\left(1 - \exp(-2\gamma t)\right) + \frac{1}{\gamma^2}\left(1 - \exp(-\gamma t)\right)^2. \end{aligned} \tag{12}$$

- A process of Yuan and Cai (2010), i.e., $X(t) = \sum_{k=1}^{50} \zeta_k Z_k \phi_k(t)$, where $\zeta_k = (-1)^{k+1} k^{-1}$, Z_k is independently sampled from a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$. One may note that the process is used in literature of functional linear models (Cai and Yuan 2012; Du and Wang 2014; Sun et al. 2018). The covariance function of $X(t)$ is $C(s,t) = \sum_{k=1}^{50} k^{-2}\phi_k(s)\phi_k(t)$.

Correspondingly, we consider four reproducing kernels as follows: $K(s,t) = \min(s,t)$ of Brownian motion, $K(s,t) = \theta_2^2 \exp(-\theta_1|t-s|)/(2\theta_1)$ of Ornstein-Uhlenbeck process, $K(s,t) = \int_0^t ds \int_0^s \left[\exp(-\gamma s) + \exp(-\gamma u)\right]^2 du$ of non-Gaussian process, and $K(s,t) = B_2(s)B_2(t)/4 - B_4(|s-t|)/24$ of cubic spline, where $B_m(\cdot)$ is the m-th Bernoulli polynomial (Gu 2013; Wang 2011). By Theorem 4.1, the processes of Brownian motion, Ornstein-Uhlenbeck process, and non-Gaussian integral $X(t) = \int_0^t ds \int_0^s (\exp(-\gamma s) + \exp(-\gamma u))\, dW(u)$ have a minimax rate of convergence $n^{-2r/(2r+1)}$, and so does the process of Yuan and Cai (2010).

For the standard Brownian motion $W(t)$, the Karhunen-Loève expansion is given by

$$W(t) = \sum_{k=1}^{\infty} \theta_k Z_k \psi_k(t), \tag{13}$$

where $\theta_k = \frac{2}{(2k-1)\pi}$, $\{Z_k : k \ge 1\}$ are identically and independently distributed (iid) $N(0,1)$ random variables, and $\psi_k(t) = \sqrt{2}\sin\left(\frac{1}{2}(2k-1)\pi t\right)$ [Iacus 2008; Pavliotis 2014]. Therefore, the covariance function of $W(t)$ is $C(s,t) = \sum_{k=1}^{\infty} \theta_k^2 \psi_k(s)\psi_k(t)$. Therefore, it is worthy of noting that both covariance functions of $W(t)$ and the process of Yuan and Cai (2010) are given by trigonometric functions in the order of k^{-2}. The covariance function of $W(t)$ is condensed as $C(s,t) = \min(s,t)$ while the covariance function of the process of Yuan and Cai (2010) is a truncated summation.

To make a comparison with the existing methods in functional data analysis, we analyze the data using function *fRegress* in *fda* to estimate the function $\beta(t)$ by two spline bases: B-spline basis and Fourier basis (Ramsay and Silverman 2005; Ramsay

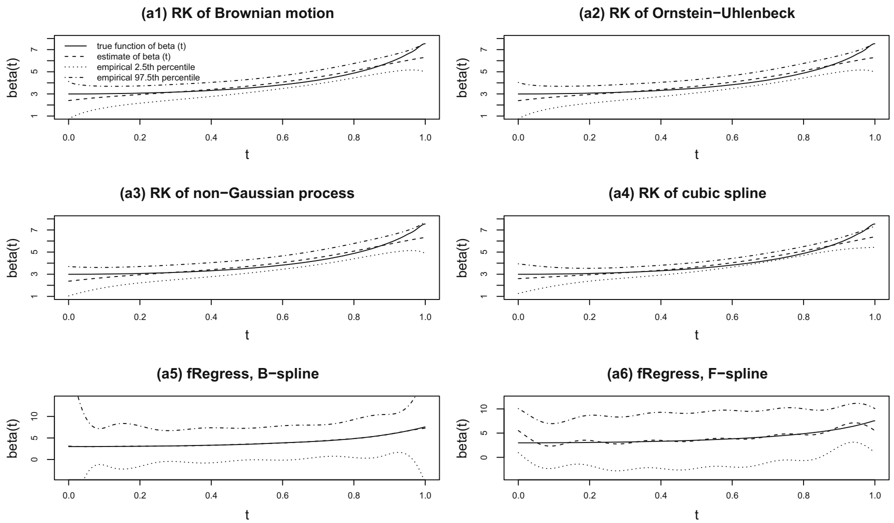

Fig. 1 True coefficient function $\beta(t)$ and its mean estimates as well as means of 95% confidence limits based on 1000 data replicates when $X(t)$ is a standard Brownian motion. In plot (a1), the reproducing kernel (RK) of Brownian motion is used; in plot (a2), the RK of Ornstein-Uhlenbeck process is used; in plot (a3), the RK of non-Gaussian is used; and in plot (a4), the RK of cubic spline is used. The function *fRegress* in *fda* is used to estimate the function $\beta(t)$ by two spline bases: B-spline basis in plot (a5) and Fourier basis in plot (a6)

et al. 2009). One may want to note that the function *fRegress* in *fda* is similar to function *fregre.basis* in *fad.usc* (Febrero-Bande and Oviedo de la Fuente 2012).

In Figs. 1, 2, 3, and 4, we show the true coefficient function $\beta(t)$ and its mean estimates as well as means of 95% confidence limits based on 1000 data replicates when $X(t)$ is the standard Brownian motion, Ornstein-Uhlenbeck process, non-Gaussian process, and the process of Yuan and Cai (2010), respectively. In each figure, the RKs $K(s,t)$ of Brownian motion, Ornstein-Uhlenbeck process, non-Gaussian process, and cubic spline are used to build RKHS. For each data set, an estimate of $\beta(t)$ and its 95% point-wise confidence intervals are calculated on a grid of $t \in [0, 1]$. The confidence intervals are derived in the same way of section 4 of Du and Wang (2014). Then, the mean of $\beta(t)$ and its mean confidence intervals are calculated based on the 1000 data replicates. It can be seen that the estimated coefficient functions track the true function well in each figure. In addition, the estimated coefficient functions are similar for the four RKs. The estimations using function *fRegress* in plots (a5) and (a6) of Figs. 1, 2, 3, and 4 are close to the true function $\beta(t)$ but the confidence intervals are wide except those in Fig. 2.

To assess the mean prediction accuracy, we generate an additional $n^* = 200$ independent predictor curves $X_1^*, \ldots X_{200}^*$ for each data replicate. Then, we calculate a mean squared error $MSE = \frac{1}{n^*}\sum_{i=1}^{n^*}\left(\eta_{\hat{\beta}}(X_i^*) - \eta_{\beta}(X_i^*)\right)$, where $\hat{\beta}$ is the estimator obtained from the data replicate. Figure 5 gives box plots of $log_{10}(MSE)$ based on the 1000 data replicates. For the four processes, the box plots of $log_{10}(MSE)$ show similar results for the four RKs of Brownian motion, Ornstein-Uhlenbeck pro-

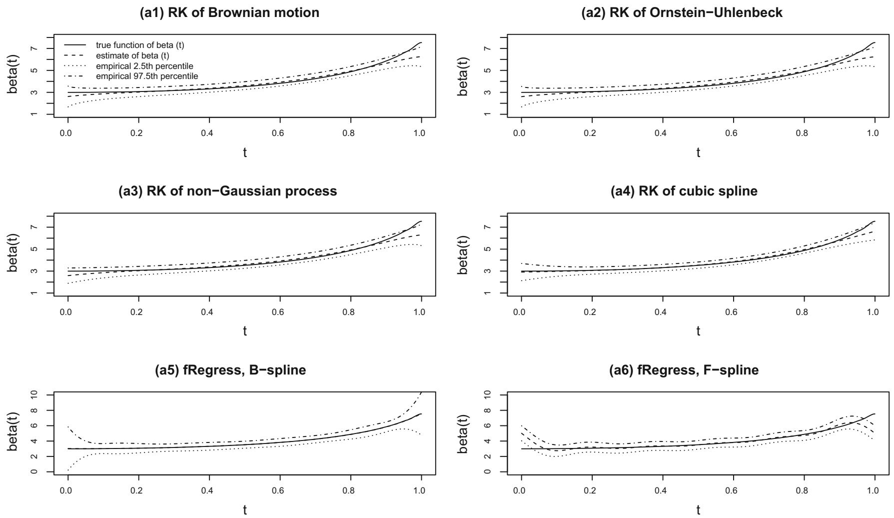

Fig. 2 True coefficient function $\beta(t)$ and its mean estimates as well as means of 95% confidence limits based on 1000 data replicates when $X(t)$ is an Ornstein-Uhlenbeck process. In plot (a1), the RK of Brownian motion is used; in plot (a2), the RK of Ornstein-Uhlenbeck process is used; in plot (a3), the RK of non-Gaussian is used; and in plot (a4), the RK of cubic spline is used. The function *fRegress* in *fda* is used to estimate the function $\beta(t)$ by two spline bases: B-spline basis in plot (a5) and Fourier basis in plot (a6)

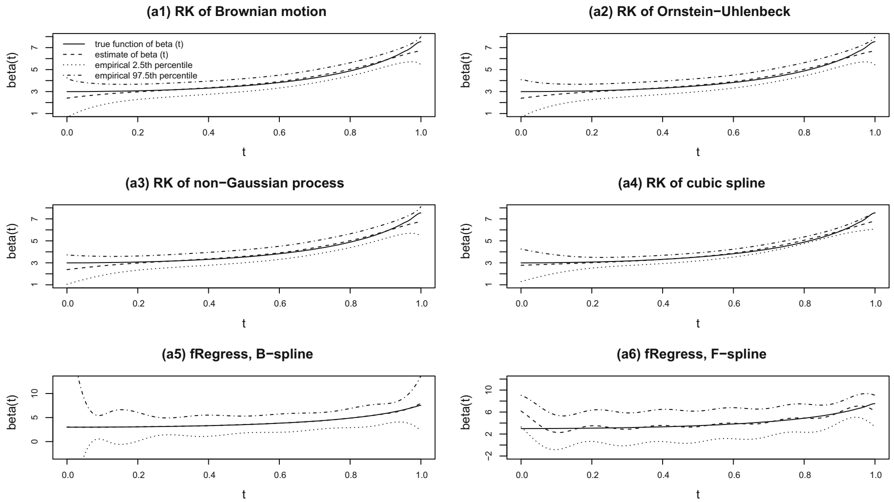

Fig. 3 True coefficient function $\beta(t)$ and its mean estimates as well as means of 95% confidence limits based on 1000 data replicates when $X(t)$ is a non-Gaussian process. In plot (a1), the RK of Brownian motion is used; in plot (a2), the RK of Ornstein-Uhlenbeck process is used; in plot (a3), the RK of non-Gaussian is used; and in plot (a4), the RK of cubic spline is used. The function *fRegress* in *fda* is used to estimate the function $\beta(t)$ by two spline bases: B-spline basis in plot (a5) and Fourier basis in plot (a6)

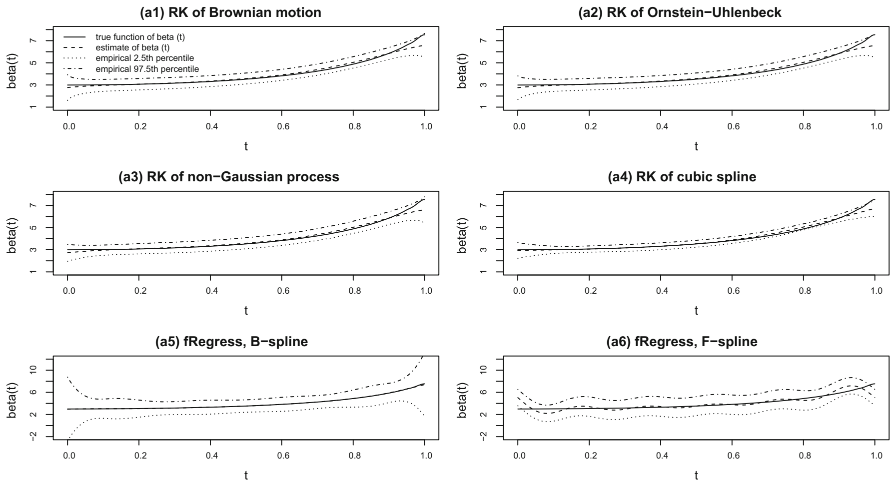

Fig. 4 True coefficient function $\beta(t)$ and its mean estimates as well as means of 95% confidence limits based on 1000 data replicates when $X(t)$ is the process in Yuan and Cai (2010). In plot (a1), the RK of Brownian motion is used; in plot (a2), the RK of Ornstein-Uhlenbeck process is used; in plot (a3), the RK of non-Gaussian is used; and in plot (a4), the RK of cubic spline is used. The function *fRegress* in *fda* is used to estimate the function $\beta(t)$ by two spline bases: B-spline basis in plot (a5) and Fourier basis in plot (a6)

cess, non-Gaussian process, and cubic spline. The mean squared errors using function *fRegress* are higher than those of the four RKs.

For each data set, an estimation error (EE) is calculated as the integrated squared error of the estimate, i.e., $\int_0^1 \left(\hat{\beta}(t) - \beta(t)\right)^2 dt$. Figure 6 shows box plots of $log_{10}(EE)$ based on the 1000 data replicates. The box plots of $log_{10}(EE)$ show that results for the four RKs are similar and the estimation errors using function *fRegress* are higher.

In summary, the results in the six figures show that the results by the four processes or RKs are better than those of function *fRegress* since the variations from the four RKs are smaller. In addition, the estimations of the function $\beta(t)$ shown in Figs. 1, 2, 3, and 4 are similar by the four RKs and function *fRegress*.

6 Discussion and concluding remarks

In functional regressions, various decomposition methods are proposed to approximate the stochastic processes $X(t)$ such as FPCA. In addition, FCCA is developed to measure functional correlation. In this article, we decompose the stochastic process $X(t)$ via Wiener-Itô chaos expansions in Malliavin calculus. Based on the expansions, we show that the fourth moments of linear functionals of $X(t)$ are bounded by the square of their second moments when the underlying stochastic process is a finite linear combination of multiple Itô integrals. Moreover, a sufficient and necessary condition of finite fourth moment is provided.

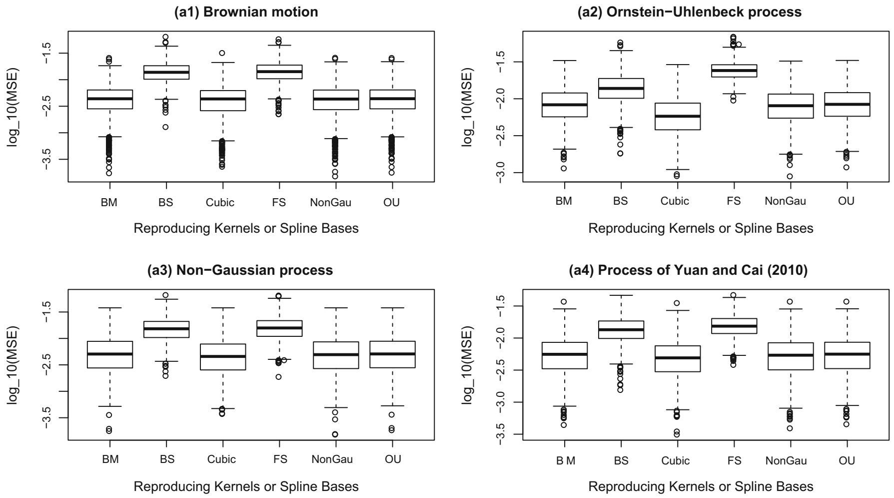

Fig. 5 Box plots of $log_{10}(MSE)$ based on 1000 data replicates. In plot (a1), $X(t)$ are simulated by Brownian motion; in plot (a2), $X(t)$ are simulated by Ornstein-Uhlenbeck process; in plot (a3), $X(t)$ are simulated by the non-Gaussian process; and in plot (a4), $X(t)$ are simulated by the process in Yuan and Cai (2010). In each plot, box plots of $log_{10}(MSE)$ are shown for four RKs and two spline bases. The four RKs are from Brownian motion (BM), Ornstein-Uhlenbeck (OU) process, non-Gaussian (NonGau) process, and cubic spline (Cubic). The two spline bases are B-spline basis (BS) and Fourier basis (FS) used in *fRegress*

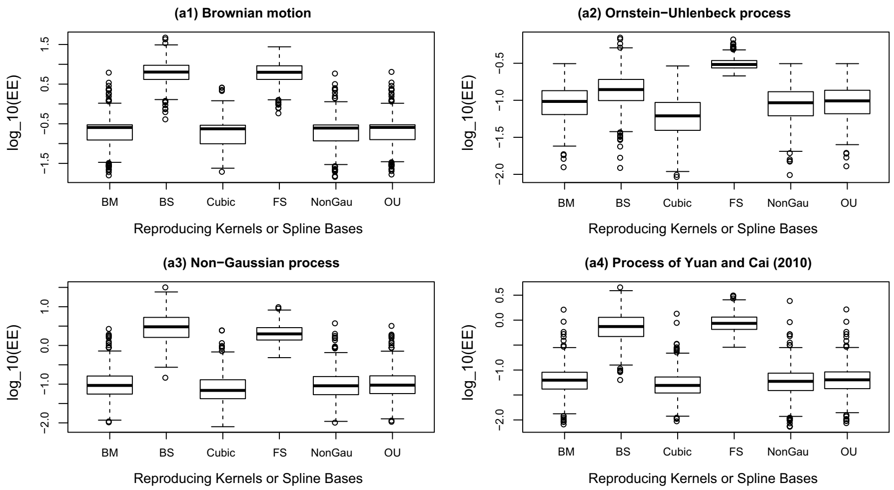

Fig. 6 Box plots of $log_{10}(EE)$ based on 1000 data replicates. In plot (a1), $X(t)$ are simulated by Brownian motion; in plot (a2), $X(t)$ are simulated by Ornstein-Uhlenbeck process; in plot (a3), $X(t)$ are simulated by the non-Gaussian process; and in plot (a4), $X(t)$ are simulated by the process in Yuan and Cai (2010). In each plot, the box plots of $log_{10}(EE)$ are shown for four RKs and two spline bases. The four RKs are from Brownian motion (BM), Ornstein-Uhlenbeck (OU) process, non-Gaussian (NonGau) process, and cubic spline (Cubic). The two spline bases are B-spline basis (BS) and Fourier basis (FS) used in *fRegress*

When the underlying process $X(t)$ is Gaussian, Cai and Hall (2006) establishes an early version of convergence rate of functional linear models. Since then, condition (7) is assumed to prove the optimal minimax convergence rate in mean prediction risk (Cai and Yuan 2012; Yuan and Cai 2010). It can be seen that the condition (7) mimicks Gaussian processes. Therefore, the minimax rate is basically established for the Gaussian processes in literature.

In this paper, we show that the condition (7) is satisfied by a linear combination of multiple Itô integrals. Therefore, the optimal minimax convergence rate in mean prediction risk is valid if eigenvalues of related linear operators are of order k^{-2r}. Our results show that the optimal minimax convergence rate in mean prediction risk can be directly applied to the class of linear combinations of multiple Itô integrals. Note that multiple Itô integrals are not necessarily Gaussian processes. Our work makes it clear that the optimal minimax convergence rate holds for a linear combination of multiple Itô integrals which are driven by the standard Brownian motion. The result shows that multiple Itô integrals can be readily applied to analyze functional data.

The literature to study functional linear models using FPCA and FCCA and RKHS is large. We don't enumerate the applications of sufficient and necessary condition of finite fourth moment in details. Nevertheless, readers can apply the Wiener-Itô decomposition to dissect functional linear models in various situations. One should bear in mind that the sufficient and necessary condition of finite fourth moment can be readily verified by elementary calculus in applications which facilitates data analysis.

To prove central limiting theorems for sequences of multiple Itô integrals to converge to a standard normal, a key condition is that the fourth moments converge to a constant 3 which provides a complete characterization of Gaussian approximations (Nourdin and Peccati 2009, 2012; Nualart 2009; Nualart and Peccati 2005; Nualart and Ortiz-Latorre 2008). Interestingly, finite fourth moment condition is required to achieve the optimal minimax convergence rate in mean prediction risk of functional linear regressions. This paper provides a characterization of the finite fourth moment condition that can be easily verified by ordinary calculus techniques. The sufficient and necessary condition of finite fourth moment of multiple Itô integrals can be directly applied to show methodology and convergence rates of functional linear models.

Computational aspects: Our derivations and results are based on Itô integrals of the standard Brownian motions. Since Brownian motions and Itô integrals are well-studied in stochastic analysis, we can use them to facilitate computing and data analysis. For instance, the RKs of Brownian motion and Ornstein-Uhlenbeck process can be used to build RKHS. Moreover, we use a non-Gaussian process to build a RK and related RKHS. By simulation studies, we show their RKs can be used to estimate true function well. In terms of computational complexity, the calculation is pretty fast. The simulation of the 6 figures in the paper can be done in less than 3 hours on a personal computer. In addition, we have developed R codes to implement the proposed methods, which are available upon request.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00180-021-01142-y.

Acknowledgements This study is supported by the U.S. National Science Foundation grant DMS-1915904 (Ruzong Fan). Two anonymous reviewers and Dr. Usha Sita Govindarajulu provide very good and insightful comments for us to improve the manuscript. We thank Dr. Nualart for the kind explanation of our questions on Malliavin calculus, Dr. Pang Du and Dr. Xiaoxiao Sun for providing us simulation codes of their work, and Dr. Ming Yuan for discussion about optimal convergence rates in mean prediction risk.

References

Bishwal JPN (2007) Parameter estimation in stochastic differential equations. Springer, Berlin
Bouleau N, Lepingle D (1992) Numerical methods for stochastic process. Wiley, New York
Cai TT, Hall P (2006) Prediction in functional linear regression. Ann Stat 34:2159–2179
Cai TT, Yuan M (2012) Minimax and adaptive prediction for functional linear regression. J Am Stat Assoc 107:1201–1216
Crambes C, Kneip A, Sarda P (2009) Smoothing splines estimators for functional linear regression. Ann Stat 37:35–72
Delaigle A, Hall P (2012) Methodology and theory for partial least squares applied to functional data. Ann Stat 40:322–352
Du P, Wang X (2014) Penalized likelihood functional regression. Stat Sinica 24:1017–1041
Eubank RL, Hsing T (2008) Canonical correlation for stochastic processes. Stoch Process Their Appl 118:1634–1661
Febrero-Bande M, Oviedo de la Fuente M (2012) Statistical computing in functional data analysis: the R package fda.usc. J Stat Softw 51(4):1–28
Ferré L, Yao A (2003) Functional sliced inverse regression analysis. Statistics 37:475–488
Ferré L, Yao A (2005) Smoothed functional sliced inverse regression. Stat Sinica 15:665–685
Fournié E, Lasry JM, Lebouchoux J, Lions PL, Touzi N (1999) Applications of Malliavin Calculus to Monte Carlo methods in finance. Finance Stoch 3:391–412
Fournié E, Lasry JM, Lebouchoux J, Lions PL (2001) Applications of Malliavin Calculus to Monte Carlo methods in finance II. Finance Stoch 5:201–236
Gu C (2013) Smoothing spline ANOVA models, 2nd edn. Springer, Berlin
Hairer M (2011) On Malliavins proof of Hörmanders theorem. Bulletin des Sciences Mathématiques 135:650–666
Hall P, Horowitz JL (2007) Methodology and convergence rates for functional linear regression. Ann Stat 35:70–91
Hall P, Hosseini-Nasab M (2006) On properties of functional principal components analysis. J R Statist Soc B 68:109–126
He G, Muller HG, Wang JL (2003) Functional canonical analysis for square integrable stochastic processes. J Multiv Anal 85:54–77
He G, Muller HG, Wang JL, Yang W (2010) Functional linear regression via canonical analysis. Bernoulli 16:705–729
Hörmander L (1967) Hypoelliptic second order differential equations. Acta Mathematica 119:147–171
Hsing T, Ren H (2009) An RKHS formulation of the inverse regression dimension reduction problem. Ann Stat 37:726–755
Iacus SF (2008) Simulation and inference for stochastic differential equations with R examples. Springer, Berlin
Ikeda N, Watanabe S (1989) Stochastic differential equations and diffusion processes, 2nd edn. North-Holland Publishing Co., Amsterdam
Itô K (1951) Multiple Wiener integral. J Math Soc Jpn 3:157–169
Kimeldorf GS, Wahba G (1970) A correspondence between Bayesian estimation on stochastic process and smoothing by splines. Ann Math Stat 41:495–502
Kimeldorf GS, Wahba G (1970) Spline functions and stochastic processes. Sankhya: The Indian Journal of Statistics. Series A 32:173–180
Kimeldorf GS, Wahba G (1971) Some results on Tchebycheffian spline functions. J Math Anal Appl 33:82–95
Kolmogorov A (1933) The consistency-condition is well satisfied by virtue of the property of multivariate Gaussian distribution. Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin

Li Y, Hsing T (2007) On rates of convergence in functional linear regression. J Multivar Anal 98:1782–1804
Li Y, Hsing T (2010) Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. Ann Stat 38:3028–3062
Malliavin P (1978) Stochastic calculus of variation and hypoelliptic operators. Proceedings of the International Symposium on Stochastic Differential equation, Kyoto 1976. Wiley, New York-Chichester-Brisbane, pages 195-263
Nourdin I, Peccati G (2009) Steins methods on Wiener chaos. Probab Theory Related Fields 145:75–118
Nourdin I, Peccati G (2012) Normal approximations with Malliavin calculus?: from Steins method to Universality. Cambridge University Press, Cambridge
Nualart D (2006) The Malliavin calculus and related topics. Springer, Berlin
Nualart D (2009) Malliavin calculus and its applications. CBMS Regional Conference Series in Mathematics, volume 110
Nualart D (2018) Introduction to Malliavin calculus. Cambridge University Press, Cambridge
Nualart D, Peccati G (2005) Central limit theorems for sequences of multiple stochastic integrals. Ann Probab 33:177–193
Nualart D, Ortiz-Latorre S (2008) Central limit theorems for multiple stochastic integrals and Malliavin calculus. Stoch Process Their Appl 118:614–628
Nunno GD, Oksendal B, Proske F (2009) Malliavin calculus for Lévy processes with applications to finance. Universitext. Springer, Berlin
Ocone DL (1984) Malliavin Calculus and stochastic integral representation of diffusion processes. Stochastics 12:161–185
Ocone DL, Karatzas I (1991) A generalized Clark representation formula, with application to optimal portfolios. Stoch Stoch Rep 34:187–220
Oksendal B (2003) Stochastic differential equations. Universitext. Springer, Berlin
Pavliotis GA (2014) Stochastic processes and applications, diffusion processes, the Fokker-Planck and Langevin equations. Springer, New York
Ramsay JO, Hooker G, Graves S (2009) Functional data analysis with R and Matlab. Springer, New York
Ramsay JO, Silverman BW (2005) Functional Data Analysis, 2nd edn. Springer, New York
Stroock DW (1987) Homogeneous chaos revisited. Seminaire de Probabilités XXI, Lecture Notes in Mathematics 1247:1–8
Sun XX, Du P, Wang X, Ma P (2018) Optimal penalized function-to-function regression under a reproducing kernel Hilbert space framework. J Am Stat Assoc 113:1601–1611
Wahba G (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. J R Stat Soc B 40:364–372
Wahba G (1990) Spline models for observational data. Society for Industrial Mathematics
Wang Y (2011) Smoothing splines, methods and applications. CRC Press, A Chapman & Hall Book
Wang JL, Chiou JM, Muller HG (2016) Review of functional data analysis. Ann Rev Stat Its Appl 3:257–295
Wiener N (1938) The homogeneous chaos. Am J Math 60(4):897–936
Yao F, Muller HG, Wang JL (2005) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100:577–590
Yao F, Muller HG, Wang JL (2005) Functional linear regression analysis for longitudinal data. Ann Stat 33:2873–2903
Yuan M, Cai TT (2010) A reproducing kernel Hilbert space approach to functional linear regression. Ann Stat 38:3412–3444