Statistica Sinica Preprint No: SS-2020-0480	
Title	Optimal Function-on-Function Regression With Interaction Between Functional Predictors
Manuscript ID	SS-2020-0480
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0480
Complete List of Authors	Honghe Jin, Xiaoxiao Sun and Pang Du
Corresponding Author	Xiaoxiao Sun
E-mail	xiaosun@arizona.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

Statistica Sinica

Optimal Function-on-Function Regression with Interaction between Functional Predictors

Honghe Jin, Xiaoxiao Sun and Pang Du

University of Georgia, University of Arizona and

Virginia Polytechnic Institute and State University

Abstract: We consider a functional regression model in the framework of reproducing kernel Hilbert spaces, where the interaction effect of two functional predictors, as well as their main effects, over the functional response is of interest. The regression component of our model is expressed by one trivariate coefficient function, the functional ANOVA decomposition of which yields the main and interaction effects. The trivariate coefficient function is estimated by optimizing a penalized least squares objective with a roughness penalty on the function estimate. The estimation procedure can be implemented easily using standard numerical tools. Asymptotic results for the proposed model, with or without functional measurement errors, are established under the reproducing kernel Hilbert space (RKHS) framework. Extensive numerical studies show the advantages of the proposed method over existing methods in terms of the prediction and estimation of the coefficient functions. An application to the histone modifications and gene expressions of a liver cancer cell line further demonstrates the better prediction accuracy of the proposed method over that of its competi-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

tors.

Key words and phrases: Functional ANOVA, Functional interaction, Function-on-Function regression, Measurement errors, Minimax convergence rate, Penalized least squares, Tensor product.

1. Introduction

Functional regression models, such as scalar-on-function, function-on-scalar, and function-on-function regression models, have attracted much attention (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006). In this article, we consider a second-order function-on-function regression model. For $1 \leq i \leq n$, the ith response function $Y_i(\cdot)$ is related to two independent functional predictors $X_i(\cdot)$ and $Z_i(\cdot)$ through

$$Y_i(t) = \int_{I_x} \int_{I_z} X_i(r) Z_i(s) \beta(t, r, s)\, dsdr + \epsilon_i(t), \quad t \in I_y. \tag{1.1}$$

Here, $\beta(\cdot, \cdot, \cdot) : I_y \times I_x \times I_z \to \mathbb{R}$ is an unknown trivariate coefficient function, and $\epsilon_i(\cdot)$, independent of $(X_i(\cdot), Z_i(\cdot))$, is an independent and identically distributed (i.i.d.) random error function with zero mean and bounded covariance function. We consider both the case without measurement errors when $X_i(\cdot)$ and $Z_i(\cdot)$ are observable, and the case with measurement errors.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

In the latter case, only surrogate processes of $X_i(\cdot)$ and $Z_i(\cdot)$, contaminated by measurement errors, are observable. Without loss of generality, we assume that I_y, I_x, and I_z are the interval $[0, 1]$, with the notion that their generalization to distinct and general compact intervals is trivial. Using a functional ANOVA decomposition, we show that model (1.1) can be written as the following function-on-function regression model:

$$\begin{aligned} Y_i(t) =& \alpha(t) + \int_{I_x} X_i(r)\beta_x(t,r)dr + \int_{I_z} Z_i(s)\beta_z(t,s)ds + \\ & \int_{I_x}\int_{I_z} X_i(r)Z_i(s)\beta_{xz}(t,r,s)\, dsdr + \epsilon_i(t). \end{aligned} \tag{1.2}$$

Here, $\alpha(\cdot)$: $I_y \to \mathbb{R}$ is the intercept function, $\beta_x(\cdot,\cdot)$ is the coefficient function corresponding to the main effect of $X_i(\cdot)$, $\beta_z(\cdot,\cdot)$ is the coefficient function corresponding to the main effect of $Z_i(\cdot)$, and $\beta_{xz}(\cdot,\cdot,\cdot)$ is the coefficient function corresponding to the interaction between $X_i(\cdot)$ and $Z_i(\cdot)$. Certain side conditions on β_x, β_z, and β_{xz}, introduced later, are needed to ensure the model identifiability. Our model is motivated by an epigenetics study on how the gene expression level is regularized by the modifications of two histones near the gene for the liver cancer cell line HepG2. The functional response is the normalized gene expression levels of each gene over genome coordinates. The density levels of two histone modifications over genome

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

coordinates are the two functional predictors.

Functional regression models with a scalar response have been studied extensively. Two main approaches are the functional principal component analysis (FPCA) approach (Yao et al., 2005a,b; Cai et al., 2006; Hall et al., 2007; Yao and Müller, 2010) and the roughness penalty approach (Ramsay and Silverman, 2005; Yuan et al., 2010; Cai and Yuan, 2012; Du and Wang, 2014; Usset et al., 2016). Recently, function-on-function regression models, especially those with a single functional predictor, have received considerable attention. The FPCA approach for function-on-function regression uses eigenfunctions in the Karhunen–Loève expansions of the response and predictor functions to represent the bivariate coefficient function (Yao et al., 2005a,b; Wu and Müller, 2011; Crambes et al., 2013). For the roughness penalty approach, Sun et al. (2018) extended the work of Cai and Yuan (2012) to model a functional response. The roughness penalty approach for function-on-function regression was also considered in Ramsay and Silverman (2005) and Ivanescu et al. (2015). Other methods include the Nadaraya–Watson approach (Ferraty et al., 2012), Bayesian mixed model approach (Meyer et al., 2015), and marginal approach (Lian, 2015). In addition to these linear models, various nonlinear function-on-function regression models with a single functional predictor have been developed.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

For example, Reimherr et al. (2018) provide a function-on-function extension of the scalar-on-function continuously additive model introduced by Müller et al. (2013). Matsui (2020) and Sun and Wang (2020) extended the functional quadratic regression model in Yao and Müller (2010) to add a quadratic term of $X(s)$ as a second functional predictor.

Function-on-function regression models with multiple functional predictors have been studied by numerous researchers. Ivanescu et al. (2015) considered an additive linear model with a mix of a multivariate scalar variable and two functional predictors. Scheipl et al. (2015) studied an additive mixed-effect model that includes interactions between scalar variables with a varying-coefficient function, smooth effect of scalar variables, and continuously additive model of a functional predictor. However, neither study considered interactions between two functional predictors or examined the theoretical properties of their estimators. Luo and Qi (2017) introduced a signal compression approach to additive function-on-function regression models. A selected number of leading eigenfunctions in the Karhunen–Loève expansion of the centered regression mean function are used to build up an approximate model, and thus reduce the problem to a principal component analysis type of constrained optimization. However, these works all ignore the interaction effects between functional predictors, which may

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

result in biased estimations of the coefficient functions and incorrect conclusions. On the other hand, the scarcity of functional models incorporating interactions may be attributed to the extra numerical and theoretical complexity induced by the additional dimension in the corresponding coefficient function.

To the best of our knowledge, the only existing function-on-function regression model that includes interactions between functional predictors is the signal compression approach (Luo and Qi, 2019), extended from Luo and Qi (2017). The approach is based on a truncated Karhunen–Loève expansion of the centered regression mean function, that is, the centered version of the double integral term in (1.1). The coefficient functions are estimated as a linear combination of the leading eigenfunctions in the truncated expansion using a constrained optimization procedure. Owing to their choice of expanding the regression mean function, rather than the functional predictors or the response, the selected leading eigenfunctions in the signal compression approach can possibly provide a more effective basis expansion system for the coefficient functions than that of the FPCA. Furthermore, their approach adds an adaptive roughness penalty to the objective functions to provide better control over the smoothness of the function estimates than that of the FPCA. However, such improvements over

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

the FPCA come with a price. First, the computation becomes more complicated. The constrained nonlinear optimization procedure requires the tuning of a roughness parameter, as well as specifying a truncation point of the Karhunen–Loève expansion. Second, these improvements make it difficult to study the asymptotic properties of the resulting estimates. Neither Luo and Qi (2017) nor Luo and Qi (2019) provide a rigorous asymptotic theory. Given the involvement of both a truncation point and a roughness penalty, it is not clear whether such a rigorous theoretical development is even realistic for their approach. Lastly, even with such improvements, our numerical experiments show that its performance is still not as good as the RKHS-based method proposed in this paper.

In this paper, we propose a penalized least squares method for the non-additive functional regression model (1.1). Under the RKHS framework, the trivariate coefficient function in (1.1) naturally splits into component coefficient functions through a functional ANOVA decomposition, including both the main effects and the interaction of the functional predictors. The estimators of the coefficient function and the component coefficient functions are obtained by optimizing a penalized least squares objective. The objective consists of the sum of the integrated mean square errors, a roughness penalty, and a smoothing parameter balancing the two parts. The

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

optimization only involves quadratic programming, and can be solved using a standard Newton–Raphson procedure. For its theoretical property, we show that the proposed method achieves the minimax convergence rate in mean prediction for a function-on-function regression model. In our simulation studies, we compare our method with those in Scheipl et al. (2015) and Luo and Qi (2019). In general, our method exhibits the best performance under a non-additive model setting such as (1.2), while being competitive under the setting of additive models. In the application to an epigenetic study on human liver cancer cells, our method also shows the best cross-validation prediction performance. In summary, the proposed method has the following advantages: (1) Theoretically, it achieves the optimal rate in the mean prediction for a function-on-function regression model, without or with functional measurement errors; (2) Computationally, the penalized least squares optimization is a quadratic programming problem and allows an easy implementation using standard optimization procedures; (3) Numerically, it demonstrates more accurate performance in the coefficient function estimation and the mean prediction, especially in the non-additive model setting; (4) The functional ANOVA decomposition under the RKHS framework provides easy access to the main and interaction effects of the functional predictors.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

The rest of the article is organized as follows. Section 2 introduces the penalized function-on-function regression model incorporating the interaction between the functional predictors. An easily implementable estimation algorithm is presented in Section 3. The optimal convergence rate in the mean prediction is presented in Section 4. Section 5 contains simulations that include comparisons with other methods and a numerical demonstration of the functional ANOVA decomposition. A real-data example is presented in Section 6. Section 7 concludes the article. The proof of the main theorem and other technical details are collected in the Supplementary Material.

2. Penalized Quadratic Function-on-Function Regression with Functional Interaction

Let $\{(X(r), Z(s), Y(t)) : r \in I_x, s \in I_z, t \in I_y\}$ be random processes defined on the domains I_x, I_z, $I_y \subseteq \mathbb{R}$, respectively. The observed data $(X_i(r), Z_i(s), Y_i(t))$, for $i = 1, ..., n$, are independent copies of the three random processes. We consider the quadratic function-on-function regression model in (1.1) and decompose β into the functional ANOVA form

$$\beta = \beta_\emptyset + \beta_t + \beta_r + \beta_s + \beta_{t,r} + \beta_{t,s} + \beta_{r,s} + \beta_{t,r,s}, \tag{2.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

where side conditions are imposed to ensure the estimability. The decomposition (2.3) allows us to rewrite the regression function $\int_{I_x}\int_{I_z} X_i(r)Z_i(s)\beta(t,r,s)\,dsdr$ in (1.1). The collective term $\int_{I_x}\int_{I_z} X_i(r)Z_i(s)(\beta_\emptyset+\beta_t+\beta_r+\beta_s+\beta_{r,s})dsdr$ does not depend on r or s, and represents the intercept function in model (1.2). The terms $\int_{I_x}\int_{I_z} X_i(r)Z_i(s)\beta_{t,r}dsdr$ and $\int_{I_x}\int_{I_z} X_i(r)Z_i(s)\beta_{t,s}dsdr$ are the main effects of $X_i(r)$ and $Z_i(s)$, respectively if we normalize $X_i(r)$ and $Z_i(s)$ to make $\int_{I_x} X_i(r)dr=\int_{I_z} Z_i(s)ds=1$. The interaction effect in (1.2) now corresponds to $\int_{I_x}\int_{I_z} X_i(r)Z_i(s)\beta_{t,r,s}dsdr$.

Suppose the coefficient function β in (1.1) is an unknown function in an RKHS $\mathcal{H}$, with the reproducing kernel K. We propose estimating β by minimizing the following penalized least squares:

$$\frac{1}{n}\sum_{i=1}^{n}\int_{I_y}\{Y_i(t)-\int_{I_x}\int_{I_z}\beta(t,r,s)X_i(r)Z_i(s)\,dsdr\}^2\,dt+\lambda J(\beta), \qquad (2.4)$$

where the roughness penalty $J(\beta)$ is a squared semi-norm on $\mathcal{H}$ quantifying the roughness of β, and $\lambda>0$ is the smoothing parameter balancing the trade-off between the goodness of fit and the smoothness of β. In this study, we consider a roughness penalty $J(\beta)$, with the exact form and interpretation of its components presented in the Supplementary Material. Note that the procedure works similarly for other common penalties, such as those

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

presented in Wahba (1990) and Gu (2013).

The functional ANOVA decomposition (2.3) of β is derived from the decomposition of the RKHS $\mathcal{H}$. Based on the tensor sum decomposition, $\mathcal{H}$ can be decomposed as $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0 = \{\beta \in \mathcal{H} : J(\beta) = 0\}$ with the reproducing kernel K_0, $\mathcal{H}_1$ is its orthogonal complement with the reproducing kernel K_1, and $K = K_0 + K_1$. Using the tensor product decomposition, one can decompose $\mathcal{H} = \mathcal{H}^y \otimes \mathcal{H}^x \otimes \mathcal{H}^z$, where $\mathcal{H}^y$, $\mathcal{H}^x$, and $\mathcal{H}^z$ are marginal subspaces of $\mathcal{H}$ for Y, X, and Z with reproducing kernels K^y, K^x, and K^z respectively, and $K(t_1, r_1, s_1; t_2, r_2, s_2) = K^y(t_1, t_2)K^x(r_1, r_2)K^z(s_1, s_2)$. We then have the following tensor sum decompositions of the marginal subspaces $\mathcal{H}^y = \mathcal{H}^y_0 \oplus \mathcal{H}^y_1$, $\mathcal{H}^x = \mathcal{H}^x_0 \oplus \mathcal{H}^x_1$, and $\mathcal{H}^z = \mathcal{H}^z_0 \oplus \mathcal{H}^z_1$, with the reproducing kernels $K^y = K^y_0 + K^y_1$, $K^x = K^x_0 + K^x_1$, and $K^z = K^z_0 + K^z_1$, respectively. Combining the tensor sum and the tensor product decomposition, we have

$$\begin{aligned}
\mathcal{H} &= (\mathcal{H}^y_0 \oplus \mathcal{H}^y_1) \otimes (\mathcal{H}^x_0 \oplus \mathcal{H}^x_1) \otimes (\mathcal{H}^z_0 \oplus \mathcal{H}^z_1) \\
&= (\mathcal{H}^y_0 \otimes \mathcal{H}^x_0 \otimes \mathcal{H}^z_0) \oplus (\mathcal{H}^y_0 \otimes \mathcal{H}^x_1 \otimes \mathcal{H}^z_0) \oplus (\mathcal{H}^y_0 \otimes \mathcal{H}^x_0 \otimes \mathcal{H}^z_1) \\
&\oplus (\mathcal{H}^y_0 \otimes \mathcal{H}^x_1 \otimes \mathcal{H}^z_1) \oplus (\mathcal{H}^y_1 \otimes \mathcal{H}^x_0 \otimes \mathcal{H}^z_0) \oplus (\mathcal{H}^y_1 \otimes \mathcal{H}^x_1 \otimes \mathcal{H}^z_0) \\
&\oplus (\mathcal{H}^y_1 \otimes \mathcal{H}^x_0 \otimes \mathcal{H}^z_1) \oplus (\mathcal{H}^y_1 \otimes \mathcal{H}^x_1 \otimes \mathcal{H}^z_1) \\
&= \mathcal{H}_0 \oplus \mathcal{H}_1,
\end{aligned} \tag{2.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

where $\mathcal{H}_0 = \mathcal{H}_0^y \otimes \mathcal{H}_0^x \otimes \mathcal{H}_0^z$, and $\mathcal{H}_1$ collects the remaining terms in (2.5). The corresponding kernels, denoted by K_0 and K_1, are defined as $K_0 = K_0^y K_0^x K_0^z$ and

$$K_1 = K_0^y K_1^x K_0^z + K_0^y K_0^x K_1^z + K_0^y K_1^x K_1^z + K_1^y K_0^x K_0^z + K_1^y K_1^x K_0^z + K_1^y K_0^x K_1^z + K_1^y K_1^x K_1^z.$$

With the aid of the tensor product and tensor sum decompositions, we now represent the coefficient function β using the basis functions of $\mathcal{H}^y$, $\mathcal{H}^x$, and $\mathcal{H}^z$. Let the basis functions be $\{\psi_j^y, j = 1, \ldots, N_y\}, \{\psi_j^x, j = 1, \ldots, N_x\}$, and $\{\psi_j^z, j = 1, \ldots, N_z\}$ for $\mathcal{H}_0^y$, $\mathcal{H}_0^x$, and $\mathcal{H}_0^z$, respectively, and $\phi_i^y(t) = \int_{I_y} K_1^y(t,u) Y_i(u)\, du$, $\phi_i^x(r) = \int_{I_x} K_1^x(r,u) X_i(u)\, du$, and $\phi_i^z(s) = \int_{I_z} K_1^z(s,u) Z_i(u)\, du$, for $i = 1, \ldots, n$, for $\mathcal{H}_1^y$, $\mathcal{H}_1^x$, and $\mathcal{H}_1^z$, respectively, where $N_y = \dim(\mathcal{H}_0^y)$, $N_x = \dim(\mathcal{H}_0^x)$, and $N_z = \dim(\mathcal{H}_0^z)$. Let $\psi^y = (\psi_1^y, \ldots, \psi_{N_y}^y)^T$, $\psi^x = (\psi_1^x, \ldots, \psi_{N_x}^x)^T$, $\psi^z = (\psi_1^z, \ldots, \psi_{N_z}^z)^T$, $\phi^y = (\phi_1^y, \ldots, \phi_n^y)^T$, $\phi^x = (\phi_1^x, \ldots, \phi_n^x)^T$, and $\phi^z = (\phi_1^z, \ldots, \phi_n^z)^T$. The coefficient function estimate $\hat{\beta}$ is of the form

$$\hat{\beta}(t,r,s) = \{\mathbf{d}_y^T \psi^y(t) + \mathbf{c}_y^T \phi^y(t)\}\{\mathbf{d}_x^T \psi^x(r) + \mathbf{c}_x^T \phi^x(r)\}\{\mathbf{d}_z^T \psi^z(s) + \mathbf{c}_z^T \phi^z(s)\}, \tag{2.6}$$

where $\mathbf{d}_y, \mathbf{c}_y, \mathbf{d}_x, \mathbf{c}_x, \mathbf{d}_z$, and $\mathbf{c}_z$ are coefficient vectors. The major advantage of this representation is that although we only include the quadratic term in

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

the quadratic function-on-function regression model, we can still study the main effects of X and Z on different subspaces in (2.5). For instance, the main effects of X and Z are estimated on $\mathcal{H}_1^x \otimes \mathcal{H}_0^z \otimes \mathcal{H}_0^y$ and $\mathcal{H}_0^x \otimes \mathcal{H}_1^z \otimes \mathcal{H}_0^y$, respectively.

3. Estimation

Based on the representation of β in (2.6), we transform the penalized least squares (2.4) into a form that can be easily optimized using standard numerical procedures. Plugging (2.6) into (2.4), we have

$$\begin{aligned}\int_{I_x}\int_{I_z} X(r)Z(s)\beta(t,r,s)drds = &\int_{I_x}\int_{I_z} X(r)Z(s)\mathbf{d}^T\psi(t,r,s)drds \\ &+ \int_{I_x}\int_{I_z} X(r)Z(s)\mathbf{c}^T\phi(t,r,s)drds,\end{aligned} \tag{3.7}$$

where $\psi = (\psi_1, \ldots, \psi_N)^T = (\psi_1^y\psi_1^x\psi_1^z, \ldots, \psi_{N_y}^y\psi_{N_x}^x\psi_{N_z}^z)^T$, $\phi = (\phi_1, \ldots, \phi_M)^T = (\phi_1^y\psi_1^x\psi_1^z, \ldots, \phi_n^y\phi_n^x\phi_n^z)^T$, and $\mathbf{d}$ and $\mathbf{c}$ are the corresponding coefficient vectors. Different selections of kernel functions usually yield different basis functions ψ and ϕ. We use the tensor product cubic spline as an example, where we have $N_y = N_x = N_z = 2$, and the basis functions $\psi_1^*(r) = 1$, $\psi_2^*(r) = r - 0.5$, with $*$ running through the index set $\{y, x, z\}$. Similarly, the basis functions ϕ are some scaled version of Bernoulli polynomi-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

als; see, for example, Gu (2013). Then, the dimension of the null space $\mathcal{H}_0$ is $N = N_y N_x N_z = 8$, and the number of basis functions in $\mathcal{H}_1$ is $M = (n+N_y)(n+N_x)(n+N_z) - N_y N_x N_z = n^3 + 6n^2 + 12n$.

The integral in (2.4) can be computed using a Gaussian quadrature; see, for example, Abramowitz and Stegun (1970), with weights $(w_1, \ldots, w_T)$ and knots $(t_1, \ldots, t_T)$: $\int_{I_y} \eta(t)\, dt = \sum_j w_j \eta(t_j)$. Let W be a diagonal matrix repeating $\text{diag}(w_1, \ldots, w_T)$ n times, S be an $nT \times N$ matrix with the $((i-1)T + j, k)th$ entry $\int_{I_x} \int_{I_z} X_i(r) Z_i(s) \psi_k(t_j, r, s) dr ds$, and R be an $nT \times M$ matrix with the $((i-1)T + j, l)th$ entry $\int_{I_x} \int_{I_z} X_i(r) Z_i(s) \phi_l(t_j, r, s) dr ds$, where $i = 1, \ldots, n$, $j = 1, \ldots, T$, $k = 1, \ldots, N$, and $l = 1, \ldots, M$. Let Σ be an $M \times M$ matrix with the $(i, j)th$ entry $\langle \phi_i(t, r, s), \phi_j(t, r, s) \rangle_{\mathcal{H}_1}$. Then, the objective function (2.4) is reduced to

$$\min_{\mathbf{d}, \mathbf{c}} \frac{1}{n} (Y_w - S_w \mathbf{d} - R_w \mathbf{c})^T (Y_w - S_w \mathbf{d} - R_w \mathbf{c}) + \lambda \mathbf{c}^T \Sigma \mathbf{c},$$

where $Y_w = W^{\frac{1}{2}} Y$, $S_w = W^{\frac{1}{2}} S$, $R_w = W^{\frac{1}{2}} R$. We use the generalized cross-validation (GCV) method to select the smoothing parameter λ (Wahba, 1975).

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

4. Optimal Rate for Mean Prediction

In this section, we show that our penalized least squares estimator achieves the minimax rate of convergence for the mean prediction of a quadratic function-on-function regression model. The result is shown for the model with functional measurement errors, and is also valid for the model without functional measurement errors as a corollary. The complete proof, including all the lemmas, is collected in the Supplementary Material S3.

4.1 Notation and Excess Risk

We start by defining some notation and the excess risk. Recall that the reproducing kernel K of $\mathcal{H}$ can be viewed as the kernel function of a positive–definite and compact linear operator L_K, defined by

$$L_K(f)(t_1, s_1, r_1) = \int\int\int K(t_1, s_1, r_1; t_2, s_2, r_2) f(t_2, s_2, r_2) dt_2 ds_2 dr_2.$$

It is easy to show that $L_{K^{1/2}}(\mathcal{L}_2) = \mathcal{H}(K)$ and

$$\langle f, g\rangle_{\mathcal{H}} = \langle L_{K^{-1/2}} f, L_{K^{-1/2}} g\rangle_{\mathcal{L}_2} = \langle f, L_{K^{-1}} g\rangle_{\mathcal{L}_2}. \tag{4.8}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

Because one can move between the $\mathcal{H}$ and $\mathcal{L}_2$ inner products, as in (4.8), we work with $\mathcal{L}_2$ instead of $\mathcal{H}$ for the minimization problem (2.4). Following the functional measurement errors setup in Jadhav and Ma (2020), we assume the observable surrogate processes are U_i and V_i, such that $U_i(\cdot) = X_i(\cdot) + \tau_i(\cdot)$ and $V_i(\cdot) = Z_i(\cdot) + \upsilon_i(\cdot)$, where $\tau_i(\cdot)$ and $\upsilon_i(\cdot)$ are functional measurement error processes independent of X_i, Z_i, and the random error processes ϵ_i. Let β_0 be the true coefficient function, and $\hat{\beta}_{n\lambda}$ be the estimator as the minimizer of (2.4) when there exist functional measurement errors. Then, there exist $f_0, \hat{f} \in \mathcal{L}_2$ such that $\beta_0 = L_{K^{1/2}} f_0$ and $\hat{\beta}_{n\lambda} = L_{K^{1/2}} \hat{f}_\lambda$. For a given reproducing kernel K and a covariance function C of $U(r)V(s)$, we further define the new kernel function T as

$$T\left((t_1, r_1, s_1), (t_2, r_2, s_2)\right) = \int\int\int\int\int K^{1/2}\left((t_1, r_1, s_1), (w, u_1, v_1)\right) C\left((u_1, v_1), (u_2, v_2)\right) K^{1/2}\left((w, u_2, v_2), (t_2, r_2, s_2)\right) du_1 dv_1 du_2 dv_2 dw,$$

which, by the spectral theorem, has the decomposition $\sum_{k=1}^{\infty} \rho_k \zeta_k(t_1, s_1, r_1)\zeta_k(t_2, s_2, r_2)$. Then, we have $L_{K^{1/2}CK^{1/2}}(\zeta_k) = \rho_k \zeta_k$, for $k = 1, 2, \cdots$, where $L_{K^{1/2}CK^{1/2}}(f) = L_{K^{1/2}}(L_C(L_{K^{1/2}}(f)))$. We are interested in recovering the functional $\eta_\beta(\cdot, \cdot, t)$: $\mathcal{L}_2 \times \mathcal{L}_2 \to \mathbb{R}$, where $\eta_\beta(U, V, t) = \int_{I_x}\int_{I_z} U(r)V(s)\beta(t, r, s)drds$. Let $(U_{n+1}, V_{n+1}, Y_{n+1})$ be a new observation that has the same distribution as, and is independent

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

of (U_i, V_i, Y_i), for $i = 1, \cdots, n$. The prediction accuracy can be measured by the excess risk

$$\begin{aligned}\mathcal{E}_n^m(\hat{\beta}_{n\lambda}) &= \int_{I_y} \mathbb{E}^* \left\{ \eta_{\hat{\beta}_{n\lambda}}(U_{n+1}, V_{n+1}, t) - \eta_{\beta_0}(U_{n+1}, V_{n+1}, t) \right\}^2 dt \\ &= \langle T^{1/2}(\hat{f} - f_0), T^{1/2}(\hat{f} - f_0) \rangle_{\mathcal{L}_2} = \left\| T^{1/2}(\hat{f} - f_0) \right\|_{\mathcal{L}_2}^2,\end{aligned}$$

where $\mathbb{E}^*$ represents the expectation taken over $(U_{n+1}, V_{n+1}, Y_{n+1})$ only.

4.2 Asymptotic Results

We now show that the proposed method achieves the minimax rate of convergence for the mean prediction in terms of the excess risk $\mathcal{E}_n^m(\cdot)$. We make the following assumptions.

(C1) The reproducing kernel K is symmetric, positive definite, and square integrable.

(C2) Assume that there exists a constant $c > 0$ for any $f \in L_2([0,1]^3)$ such that

$$\int \mathbb{E} \left\{ \int \int U(s)V(r)f(t,r,s)dsdr \right\}^4 dt \leq c \int \left[\mathbb{E} \left\{ \int \int U(s)V(r)f(t,r,s)dsdr \right\}^2 \right]^2 dt.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

(C3) Assume the eigenvalues $\{\rho_k : k \geq 1\}$ of the linear operator $L_{K^{1/2}CK^{1/2}}$ satisfy $\rho_k \asymp k^{-2\omega}$, for some constants $0 < \omega < \infty$.

(C4) For all $i = 1, \cdots, n$, there exists a constant $M > 0$ such that $\mathbb{E}(\epsilon_i(t)^2) \leq M < \infty$.

(C5) For all $1 \leq i \leq n$ and $k \geq 1$, we assume that $\langle X_i \upsilon_i, L_{K^{1/2}}\zeta_k\rangle^2_{\mathcal{L}_2} \leq \rho_k, \langle \tau_i Z_i, L_{K^{1/2}}\zeta_k\rangle^2_{\mathcal{L}_2} \leq \rho_k$, and $\langle \tau_i \upsilon_i, L_{K^{1/2}}\zeta_k\rangle^2_{\mathcal{L}_2} \leq \rho_k$.

Assumption (C1) guarantees the spectral decomposition of the kernel via Mercer's theorem. Assumption (C2) is a common assumption in the asymptotic analysis of functional linear models (Cai and Yuan, 2012). It assumes that the fourth moment is bounded by a constant multiplying the square of the second moment. We only make the assumptions for the observed surrogate variables, because it is realistic to verify them. The convergence rate in Theorem 1 is determined by the decay rate of the kernel function T, which is assumed in Assumption (C3). The kernel function T is the interaction between the reproducing kernel K and the covariance function C of the two functional predictors. If K and C are perfectly aligned, it is easy to show that the constant r in the rate depends jointly on the decay rates of K and C. In our setting, r is jointly determined by K and C, as well as the alignment between K and C, in a complicated

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

way. Assumption (C4) assumes that the variance of the error terms is bounded. Assumption (C5) ensures that the signal-to-noise ratios (SNRs) of the functional predictors against the functional measurement errors are sufficiently large to recover the true predictor functions.

Theorem 1. *If Assumptions (C1) to (C5) hold,*

$$\lim_{A\to\infty}\lim_{n\to\infty}\sup_{\beta\in\mathcal{H}(K)}\mathbb{P}\{\mathcal{E}_n^m \geq An^{-2\omega/(2\omega+1)}\} = 0,$$

provided that $\lambda \asymp n^{-2\omega/(2\omega+1)}$.

The rate $n^{-2\omega/(2\omega+1)}$ in Theorem 1 matches the minimax lower bound of the convergence rate for the prediction risk under a general function-on-function regression model, as established in Theorem 3 of Sun et al. (2018). Therefore, the rate of the prediction risk for our penalized least squares estimator is the minimax rate.

The model without measurement errors can be considered as a special case with zero measurement errors, and its asymptotic result can be proved similarly to Theorem 1, with slightly weaker assumptions. Therefore, we present the result as Corollary 1, with the proof omitted.

(C2′) Assume that there exists a constant $c > 0$, for any $f \in L_2([0,1]^3)$,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

such that

$$\int \mathbb{E}\left\{\int\int X(s)Z(r)f(t,r,s)dsdr\right\}^4 dt \leq c\int\left[\mathbb{E}\left\{\int\int X(s)Z(r)f(t,r,s)dsdr\right\}^2\right]^2 dt.$$

Corollary 1. *Let $\mathcal{E}_n$ be the excess risk for the model without functional measurement errors. If Assumptions (C1), (C2′), (C3), and C(4) hold,*

$$\lim_{A\to\infty}\lim_{n\to\infty}\sup_{\beta\in\mathcal{H}(K)}\mathbb{P}\{\mathcal{E}_n \geq An^{-2\omega/(2\omega+1)}\} = 0,$$

provided that $\lambda \asymp n^{-2\omega/(2\omega+1)}$.

5. Numerical Experiments

In this section, we study the numerical performance of the proposed approach by means of simulations. We first compare the proposed method with two other existing methods, and then demonstrate how the functional ANOVA works between models (1.1) and (1.2).

We simulated data according to model (1.1) under three scenarios with different coefficient functions, as specified below:

Scenario 1: $\beta(t,r,s) = (r-0.5)^2\exp(t) - t\{(s-0.5)^2 - 0.3s\}$

Scenario 2: $\beta(t,r,s) = \sin(t)\cos(\pi r^3) + \cos(t)\sin(2\pi s)$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

Scenario 3: $\beta(t,r,s) = \exp(-t)\{s^2 + r^2\cos(\pi s)\}$.

When t is fixed, the coefficient function β is additive in r and s in Scenarios 1 and 2, but contains an interaction term between r and s in Scenario 3. For each scenario, the functional predictors X_i and Z_i were generated from $X_i(r) = \sum_{k=1}^{50}(-1)^{k+1}k^{-1}W_{ik}\vartheta_1(r,k)$ and $Z_i(s) = \sum_{k=1}^{50}(-1)^{k+1}k^{-1}W_{ik}\vartheta_2(s,k)$, where the constants W_{ik} were generated from the uniform distribution $U(-\sqrt{3},\sqrt{3})$, $\vartheta_1(r,k) = 1$ if $k = 1$, and $\vartheta_1(r,k) = \sqrt{2}\cos\{(k-1)\pi r\}$ otherwise, and $\vartheta_2(s,k) = s^2$ if $k = 1$, and $\vartheta_2(s,k) = \frac{1-\cos((k-1)\pi s)}{\{(k-1)\pi\}^2}$ otherwise. We generated $n = 50$ samples, each with 20 time points on the interval $[0,1]$. The random errors $\epsilon(t)$ were from a normal distribution with a zero mean and a constant variance σ^2. The SNRs $\sum_i \int \mathrm{E}\{\eta_i^2(t)\}\,dt / \sum_i \int \mathrm{E}\{\epsilon_i^2(t)\}\,dt$ were set to 1, 5, and 10 in each scenario, where $\eta_i(t) = \int_0^1 \int_0^1 X_i(r)Z_i(s)\beta(t,r,s)\,drds$.

5.1 Comparisons with Other Methods

We compare our method, *QFFR*, with two function-on-function regression models for multiple functional predictors, namely the additive function-on-function regression method of Scheipl et al. (2015), denoted by *pffr*, and the signal compression function-on-function regression method of Luo and Qi (2019), denoted by *FRegSig*. For *pffr* and *FRegSig*, we used their respective R packages in all the simulations in this section and the real-data analysis

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

	SNR	*QFFR*	*pffr*	*FRegSig*
Scenario 1	1	2.529(1.480)	**1.096**(1.023)	4.736(14.83)
	5	1.288(0.510)	**0.357**(0.287)	0.815(1.468)
	10	0.747(0.419)	**0.157**(0.163)	0.327(0.758)
Scenario 2	1	**0.208**(0.148)	5.565(16.44)	1.641(1.951)
	5	**0.469**(0.291)	375.7(190.7)	3.414(30.87)
	10	**0.225**(0.361)	382.7(398.0)	1.652(1.741)
Scenario 3	1	**0.613**(1.026)	5.066(7.862)	2.149(43.98)
	5	**0.0398**(0.0669)	0.376(1.063)	0.603(0.819)
	10	**0.0198**(0.0316)	0.286(0.462)	0.831(1.253)

Table 1: The averages (standard deviations) of the MISEs ($\times 10^{-3}$) under the three scenarios. The best result is shown in boldface.

in the next section.

The prediction accuracy was quantified by the mean integrated squared error MISE$= \frac{1}{\breve{n}} \sum_{i=1}^{\breve{n}} \int_0^1 \{\eta_i(t) - \hat{\eta}_i(t)\}^2 \, dt$, where $\breve{n}$ is the sample size of the test data, and $\hat{\eta}_i(t) = \int_0^1 \int_0^1 X_i(r) Z_i(s) \hat{\beta}(t, r, s) \, dr ds$. We repeated the simulation 100 times for the test data of size $\breve{n} = 50$. The prediction performance of *QFFR*, *pffr*, and *FRegSig* in terms of the MISE is summarized in Table 1. We also show the true coefficient functions and their estimates from the three methods when t is fixed at 0.48 in Figure 1. As expected, the MISEs in Table 1 decrease as the SNR increases for each method in each scenario, except for the *pffr* in Scenario 2. Specifically, in Scenario 1, where the true model is additive, *pffr* performs best with the smallest MISEs in all the scenarios. This is not surprising, because *pffr* is designed

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

for the additive model, whereas the interaction terms in *QFFR* could lead to a slight overfitting problem. Note too that the difference between the MISEs of the two methods is small. Figure 1 confirms that *QFFR* does a decent job in terms of estimation and is comparable to *pffr*, but that the *FRegSig* estimate is a little off. For Scenario 2, the true model is also additive, but *pffr* clearly had difficulty in this scenario. The *pffr* estimate shown in Figure 1 for Scenario 2 is completely off the target, as shown in Table 1. The sinusoidal functions might be challenging for *pffr*. On the other hand, the proposed *QFFR* method had the best performance, both numerically and visually. In Scenario 3, the true model was non-additive. Thus, *pffr* was not expected to perform well, because it assumed an additive structure. Despite this, *pffr* still outperformed *FRegSig* when the SNRs were high. On the other hand, *QFFR* was clearly the best, both numerically and visually, in this non-additive scenario.

To assess the effect of the error distributions, we performed the same simulations using heavy-tailed error distributions. In particular, we considered errors from t-distributions with different SNRs. The results, collected in the Supplementary Material, show similar patterns to the findings with normally distributed errors.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

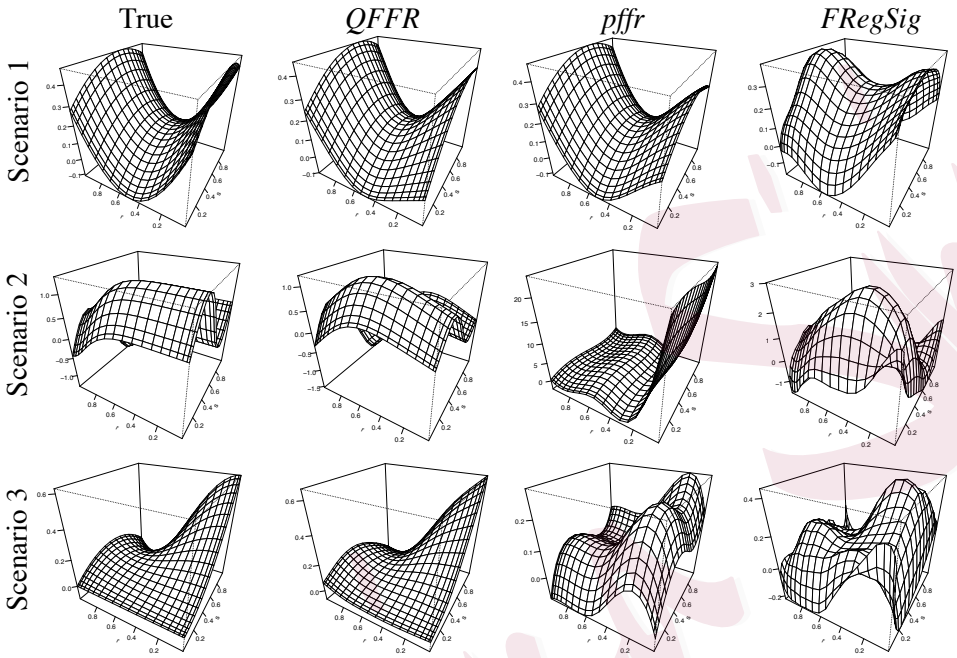

Figure 1: Perspective plots of the true coefficient functions $\beta(t, r, s)$ (left panels) and their estimates by *QFFR*, *pffr*, and *FRegSig* at $t = 0.48$ for the three scenarios.

We also conducted a simulation study to illustrate the finite-sample properties of the proposed estimator. The SNR was set as five in Scenarios 1 to 3, and the sample size varied from 20 to 1000. The results (log(MISE)) of the proposed method *QFFR* and *pffr* and *FRegSig* are shown in Figure 2. The proposed method converges as the sample size increases. We observed similar patterns for *FRegSig* in the three scenarios. The method *pffr* performed best in Scenario 1, but failed to converge in Scenarios 2 and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

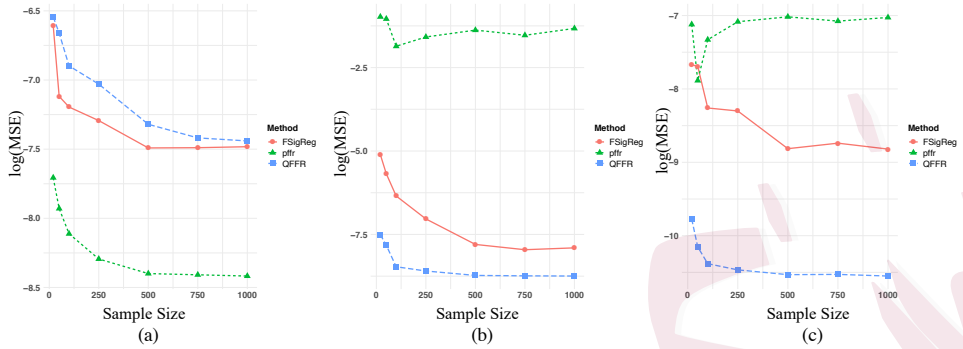

Figure 2: MISEs of the three methods in (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3

3.

In summary, *QFFR* is a competitive method in the scenarios of additive models, and the best method in the non-additive scenario. While *FRegSig* delivers satisfactory performance in all the scenarios, it is much less accurate than *QFFR*. Because all three methods employ roughness penalties on the function estimates, a possible reason for the lower accuracy of *FRegSig* might be that its principal component type of optimization is numerically more challenging than the penalized least squares used in the other two methods.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

5.2 Functional Measurement Errors

In this section, we show the estimation results when the covariate functions contain normal distributed errors under the same experiment setting. In particular, we first generated $X_i(r)$ and $Z_i(s)$ as before. Then, we added the measurement errors and applied the methods for the comparison with $U_i(r) = X_i(r) + \tau_i(r)$ and $V_i(s) = Z_i(s) + \upsilon_i(s)$, where $\tau_i(r_j) \sim N(0, \tau^2)$ and $\upsilon_i(s_k) \sim N(0, \upsilon^2)$, for $i = 1, ..., n$, $j = 1, ..., J$, and $k = 1, ..., K$. The response function Y_i was generated using (1.1). In the simulation, we set the SNR of the response to 10 and the SNR of the covariates

$$\text{SNR}_{XZ} = \frac{\sum_i \int \text{E}\{X_i^2(r)\}\, dr}{\sum_i \tau_i^2} = \frac{\sum_i \int \text{E}\{Z_i^2(s)\}\, ds}{\sum_i \upsilon_i^2}$$

to 2, 5, and 10.

In Table 2, we present the MISEs of the three methods, namely *QFFR*, *pffr*, and *FRegSig*, with their standard deviations under the three scenarios. The distributions of the MISEs in Table 2 are similar to those in Table 1. We expected the MISEs to decrease as the SNR_{XZ} increased. The proposed method *QFFR* performed best in Scenarios 2 and 3. Comparing the results in the scenarios with and without measurement errors, the MISEs increase significantly under the scenarios with measurement errors. However, when

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

	SNR_{XZ}	*QFFR*	*pffr*	*FRegSig*
Scenario 1	2	9.123(2.001)	**4.551**(1.221)	10.757(6.830)
	5	1.488(0.475)	**1.205**(0.332)	3.075(2.500)
	10	0.998(0.502)	**0.875**(0.267)	1.588(0.917)
Scenario 2	2	**3.822**(0.971)	7.355(3.487)	9.956(6.751)
	5	**1.074**(0.291)	4.357(1.335)	2.462(1.490)
	10	**0.569**(0.401)	97.35(78.00)	1.998(0.812)
Scenario 3	2	**4.332**(2.083)	4.353(2.105)	8.414(3.025)
	5	**0.475**(0.241)	3.257(2.809)	2.782(0.723)
	10	**0.180**(0.099)	4.201(0.462)	1.055(0.801)

Table 2: The averages (standard deviations) of the MISEs ($\times 10^{-3}$) under the three scenarios with measurement errors. The best result is shown in boldface.

the SNR_{XZ} is sufficiently large, for example, $\text{SNR}_{XZ} = 10$, the prediction performance is close to that when there were no measurement errors.

5.3 Functional ANOVA Demonstration

We now demonstrate numerically that model (1.1) can be transformed to model (1.2) using the functional ANOVA. To show the transformation, we used the non-additive model from Scenario 3 when there are no measurement errors, and set the SNR to two. Following the functional ANOVA decomposition in (2.3), we estimated the main and interaction coefficient functions in the corresponding function spaces in (2.5). For instance, the main effect coefficient function for the functional predictor $X(t)$ was estimated in the function subspace $\mathcal{H}_1^y \otimes \mathcal{H}_1^x \otimes \mathcal{H}_0^z$ by using the kernel $K_1^y K_1^x K_0^z$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

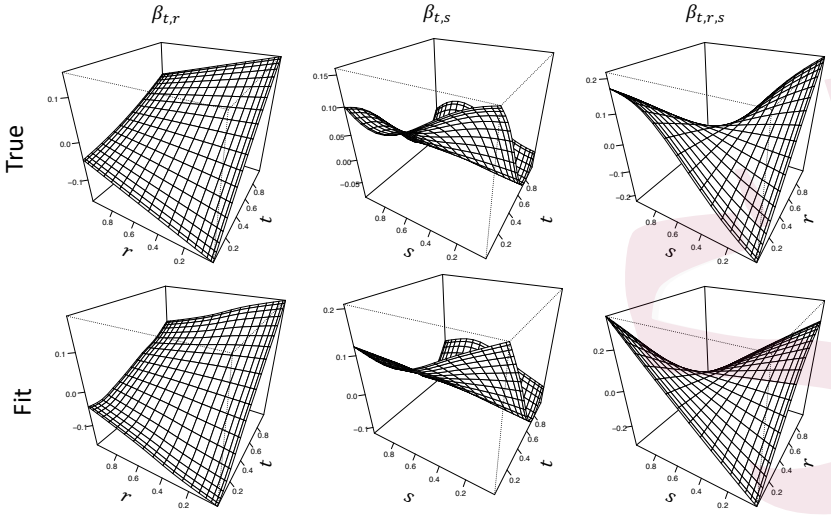

Figure 3: The functional ANOVA of coefficient functions. The first row shows the decomposition of the true function, and those in the second row are the fitted coefficient functions.

The details of the kernels and the inner products are listed in the Supplementary Material S2. Then, fitting model (1.2) is equivalent to estimating the coefficient vectors corresponding to the basis functions in the subspaces. The three-component coefficient functions for the main and interaction effects in model (1.2) of $\beta(t, r, s) = \exp(-t)\{s^2 + r^2 \cos(\pi s)\}$ are shown in the first row of Figure 3, and the fitted ones are shown in the second row. To present the results graphically, we set $t = 0.48$ for the three-dimensional function $\beta_{t,r,s}$. Clearly, the proposed method can successfully recover all the true component coefficient functions using the functional ANOVA.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

6. Application: Epigenetic Study for Liver Cancer Cells

In this section, we examine how two types of histone modifications, H3K4me2 and H3K9me3, affect the gene expression in the human liver cancer cell line (HepG2) (Consortium et al., 2012). We calculated the values of the fold change over the control for the histone modifications levels on the promoter region, that is, 1000 bp upstream or downstream of where a gene's transcription begins. To study the liver cancer tumorigenesis, we selected 2,475 liver cancer related genes (Lee et al., 2011). There were 2,142 genes left after filtering out those genes with low histone modifications and gene expression levels. We removed genes when the maximum expression value or maximum histone density level was zero. The tight clustering method was applied to the gene expression, H3K4me2, and H3K9me3 levels (Tseng and Wong, 2005). A tight cluster that contained 59 genes with similar expression patterns was selected. The selected gene names can be found in the Supplementary Material S5. The resulting data are displayed in Figure 4.

	1st Qu.	Median	Mean (SD)	3rd Qu.
QFFR	0.462	**0.503**	**0.487** (**0.077**)	**0.522**
pffr	0.417	0.515	0.577 (0.227)	0.742
FRegSig	**0.405**	0.565	0.559 (0.382)	0.761

Table 3: The leave-one-out cross-validation scores for *QFFR*, *pffr*, and *FRegSig* in the liver cancer cell line data. The best result on each metric is shown in boldface.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

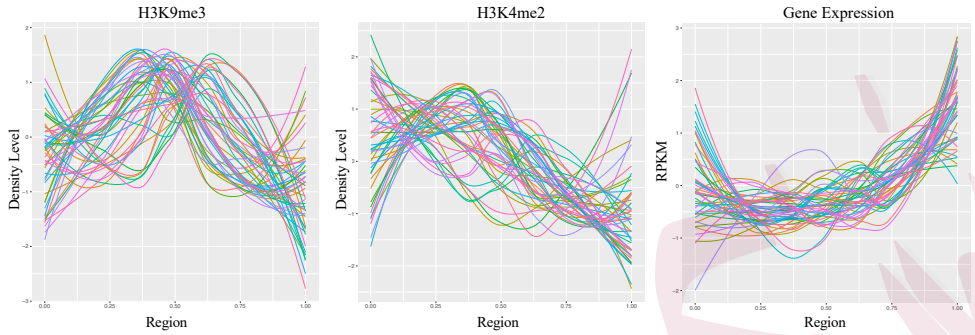

Figure 4: Histone modifications and gene expression levels for liver cancer cell line data. The x-axes represent genome coordinates normalized into $[0, 1]$. The y-axes denote normalized density levels of histone modifications (left two panels) and reads per kilobase of transcript, per million mapped reads (RPKM) for gene expression (rightmost panel).

We applied *pffr*, *FRegSig*, and the proposed method *QFFR* to the data. Because we were interested in studying how histone modifications regulate gene activities on the promoter region, we used the gene expression levels as the functional response and the density levels of the two histone modifications as the functional predictors. Our proposed method *QFFR* performed better than the other two methods in terms of the prediction accuracy, measured using the leave-one-out cross-validation score, as shown in Table 3. The cross-validation score is defined by $\frac{\sum_{i=1}^{n}\int\{Y_i(t)-\hat{Y}^{-i}(t)\}^2\,dt}{\sum_{i=1}^{n}\int\{Y_i(t)\}^2\,dt}$, where $\hat{Y}^{-i}(t)$ is the predicted value from the model fitted after deleting the ith gene in the training data.

In Figure 5, we show the coefficient functions estimated by *QFFR*, *pffr*,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

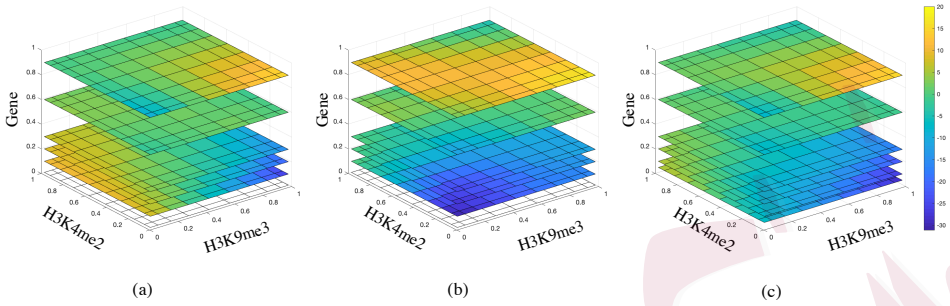

Figure 5: Fitted coefficient functions by (a) *QFFR*, (b) *pffr*, and (c) *FRegSig*.

and *FRegSig*. Each slice represents the estimated coefficient function at a fixed region of genes. The figure shows how H3K9me3 and H3K4me2 regulate (positively or negatively) gene activities over the promoter region. The coefficient function fitted by *pffr* displays a clear pattern of additive effects by H3K4me2 and H3K9me3. That is, the predicted gene expression level is a simple addition of the separate effects from the two types of histone modifications; see panel (b) of Figure 5. Both *FRegSig* and *QFFR* recover the interaction effect of similar patterns between H3K4me2 and H3K9me3 successfully. However, the interaction effect from our model is more significant than that in *FRegSig*. For example, according to *FRegSig*, the two histone modifications had only negative effects on the gene activities at slices from 0 to 0.2 along the gene axis; see panel (c) of Figure 5. On the other hand, the *QFFR* coefficient estimate reveals more dynamic changes

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

in these slices; see panel (a) of Figure 5. Based on the results in Table 3, *QFFR* may have estimated the interaction effect more accurately than did *FRegSig*.

In Figure 6, we show the functional ANOVA decomposition of the coefficient function $\hat{\beta}$ estimated by *QFFR*. In panels (a) and (b), we show the main effects of the two types of histone modifications, namely, H3K4me2 and H3K9me3, respectively. The histone modifications have varied modification effects on the gene expression over the promoter regions. The regulation effects are increasing from the start to the end points of the promoter region; see Panel (a). The regulation patterns for H3K9me3 are clearly different from the patterns for H3K4me2; see Panel (b). When they work jointly, the H3K9me3 regulation effect is more active at the end point of the promoter region for H3K4me2 compared with its effect at the start point of the promoter region; see Panel (c).

7. Conclusion

In this paper, we propose a penalized least squares method to estimate the coefficient function in a quadratic function-on-function regression model, including the interaction between the functional predictors. Using the functional ANOVA decomposition under an RKHS framework, the trivariate co-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

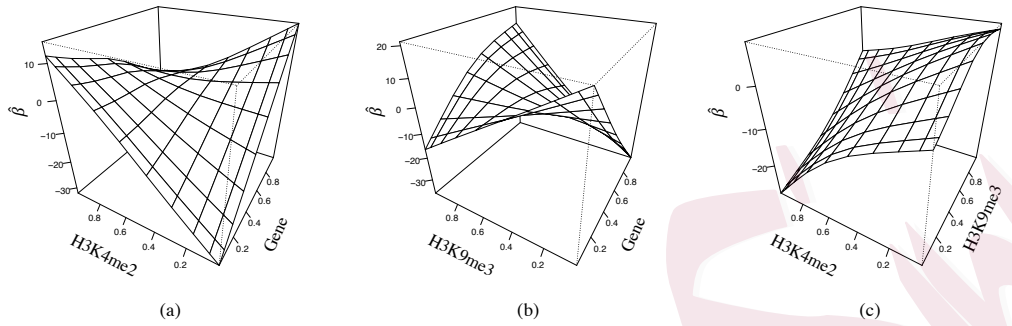

Figure 6: Functional ANOVA decomposition of the coefficient function estimated by *QFFR*. (a) The main effect of H3K4me2, (b) the main effect of H3K4me2, and (c) the interaction effect between H3K4me2 and H3K9me3 at the midpoint of the promoter region of the genes.

efficient function can naturally split into the main and interaction effects of the functional predictors, the separate estimates of which are readily available from the trivariate coefficient function estimate. We prove theoretically that the estimator for the model with or without functional measurement errors achieves the optimal convergence rate in the mean prediction. Simulation studies demonstrate that the proposed method has numerical advantages over existing methods. An application to liver cancer cell line data shows how two histone modifications at the promoter region jointly regulate gene activities.

The proposed *QFFR* method allows easy extensions in several aspects. For example, separate penalties on different components of the trivariate co-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

efficient function can be enforced by introducing multiple smoothing parameters in the roughness penalty; see, for example, Section 2.4 in Gu (2013). In addition, the model here contains only two functional predictors, but it can be extended easily to models with more than two functional predictors, although the computation will become more complicated if three-way or higher-order interactions are included. Furthermore, the method can be extended to model a categorical response function, such as binary or count response functions.

The number of sampling points per curve is an interesting topic in functional regression models (Cai and Yuan, 2011; Zhang and Wang, 2016). In this study, we focus on the case when curves are densely sampled such that the asymptotic property of the estimator is the same as the case when the curves are completely known. However, when the curves are sparsely sampled, with sampling points randomly dispersed across the curve domain, a common approach is to use the principal component analysis through conditional expectation (PACE) method of Yao et al. (2005a) to first perform presmoothing of the curves; see, for example, Sun et al. (2018).

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

Supplementary Material

In the online Supplementary Material, we introduce an example of the roughness penalty in S1 and the corresponding reproducing kernels in S2. S3 provides the proof of Theorem 1. The simulation results with random errors from heavy-tailed distributions are shown in S4. In S5, we list the gene names in the real-data example. A link to the R code of the proposed method can be found in S6.

Acknowledgments

The authors would like to thank the anonymous reviewers, the associate editor, and editor for their constructive comments and suggestions.

Funding

Du's research was supported by the U.S. National Science Foundation under grants DMS-1620945 and DMS-1916174. Sun's research was supported by the 18th Mile TRIF Funding from the University of Arizona and the U.S. National Institutes of Health under the grant 1U19AG065169-01A1.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

References

Abramowitz, M. and I. A. Stegun (1970). Handbook of mathematical functions with formulas, graphs, and mathematical tables, Volume 55. US Government printing office.

Cai, T. T., P. Hall, et al. (2006). Prediction in functional linear regression. Ann. Stat. 34(5), 2159–2179.

Cai, T. T. and M. Yuan (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. Ann. Stat. 39(5), 2330–2355.

Cai, T. T. and M. Yuan (2012). Minimax and adaptive prediction for functional linear regression. J. Amer. Statist. Assoc. 107(499), 1201–1216.

Consortium, E. P. et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414), 57.

Crambes, C., A. Mas, et al. (2013). Asymptotics of prediction in functional linear regression with functional outputs. Bernoulli 19(5B), 2627–2651.

Du, P. and X. Wang (2014). Penalized likelihood functional regression. Statist. Sin., 1017–1041.

Ferraty, F., I. Van Keilegom, and P. Vieu (2012). Regression when both response and predictor are functions. J. Multivariate Anal. 109, 10–28.

Ferraty, F. and P. Vieu (2006). Nonparametric functional data analysis: theory and practice. Springer Science & Business Media.

Gu, C. (2013). Smoothing spline ANOVA models, Volume 297. Springer Science & Business

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

Media.

Hall, P., J. L. Horowitz, et al. (2007). Methodology and convergence rates for functional linear regression. Ann. Stat. 35(1), 70–91.

Ivanescu, A. E., A.-M. Staicu, F. Scheipl, and S. Greven (2015). Penalized function-on-function regression. Computational Statistics 30(2), 539–568.

Jadhav, S. and S. Ma (2020). Functional measurement error in functional regression. Canad. J. Stat. 48(2), 238–258.

Lee, L., K. Wang, G. Li, Z. Xie, Y. Wang, J. Xu, S. Sun, D. Pocalyko, J. Bhak, C. Kim, et al. (2011). Liverome: a curated database of liver cancer-related gene signatures with self-contained context information. In BMC Genomics, Volume 12, pp. S3. BioMed Central.

Lian, H. (2015). Minimax prediction for functional linear regression with functional responses in reproducing kernel hilbert spaces. J. Multivariate Anal. 140, 395–402.

Luo, R. and X. Qi (2017). Function-on-function linear regression by signal compression. J. Amer. Statist. Assoc. 112(518), 690–705.

Luo, R. and X. Qi (2019). Interaction model and model selection for function-on-function regression. J. Comput. Graph. Stat. 28(2), 309–322.

Matsui, H. (2020). Quadratic regression for functional response models. Econometrics and Statistics 13, 125–136.

Meyer, M. J., B. A. Coull, F. Versace, P. Cinciripini, and J. S. Morris (2015). Bayesian function-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

on-function regression for multilevel functional data. Biometrics 71(3), 563–574.

Müller, H.-G., Y. Wu, and F. Yao (2013). Continuously additive models for nonlinear functional regression. Biometrika 100(3), 607–622.

Ramsay, J. O. and B. W. Silverman (2005). Functional data analysis. Springer-Verlag.

Reimherr, M., B. Sriperumbudur, B. Taoufik, et al. (2018). Optimal prediction for additive function-on-function regression. Electron. J. Stat. 12(2), 4571–4601.

Scheipl, F., A.-M. Staicu, and S. Greven (2015). Functional additive mixed models. J. Comput. Graph. Stat. 24(2), 477–501.

Sun, X., P. Du, X. Wang, and P. Ma (2018). Optimal penalized function-on-function regression under a reproducing kernel Hilbert space framework. J. Amer. Statist. Assoc., 1–11.

Sun, Y. and Q. Wang (2020). Function-on-function quadratic regression models. Comput. Stat. Data Anal. 142, 106814.

Tseng, G. C. and W. H. Wong (2005). Tight clustering: a resampling-based approach for identifying stable and tight patterns in data. Biometrics 61(1), 10–16.

Usset, J., A.-M. Staicu, and A. Maity (2016). Interaction models for functional regression. Comput. Stat. Data Anal. 94, 317–329.

Wahba, G. (1975). Smoothing noisy data with spline functions. Numer. Math. 24(5), 383–393.

Wahba, G. (1990). Spline models for observational data, Volume 59 of CBMS-NSF Regional Conf. Ser. in Appl. Math. Philadelphia: SIAM.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0480

REFERENCES

Wu, S. and H.-G. Müller (2011). Response-adaptive regression for longitudinal data. Biometrics 67(3), 852–860.

Yao, F. and H.-G. Müller (2010). Functional quadratic regression. Biometrika 97(1), 49–64.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc. 100(470), 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. Ann. Stat., 2873–2903.

Yuan, M., T. T. Cai, et al. (2010). A reproducing kernel Hilbert space approach to functional linear regression. Ann. Stat. 38(6), 3412–3444.

Zhang, X. and J.-L. Wang (2016). From sparse to dense functional data and beyond. Ann. Stat. 44(5), 2281–2321.