A robust spatial autoregressive scalar-on-function regression with t-distribution

Tingting Huang, Gilbert Saporta, Huiwen Wang, Shanshan Wang

A Robust Spatial Autoregressive Scalar-on-Function Regression with t-distribution

Tingting Huang [1,2], Gilbert Saporta [3]
Huiwen Wang [1,4], Shanshan Wang [1,2]

1. School of Economics and Management, Beihang University, Beijing, China
2. Beijing Key Laboratory of Emergence Support Simulation Technologies for City Operations, Beijing, China
3. CEDRIC CNAM, 292 rue St Martin, 75141 Paris Cedex 03, France
4. Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, China

Address for correspondence: Shanshan Wang, School of Economics and Management, Beihang University, Xueyuan Road No.37, Haidian District, Beijing, China
E-mail: sswang@buaa.edu.cn
Phone: (+86) 010 82339337
Fax: (+86) 010 82328037

This research was financially supported by the National Natural Science Foundation of China under grant nos. 71420107025 and 11701023.

A Robust Spatial Autoregressive Scalar-on-Function Regression with t-distribution

Abstract

Modelling functional data in the presence of spatial dependence is of great practical importance as exemplified by applications in the fields of demography, economy and geography, and has received much attention recently. However, for the classical scalar-on-function regression (SoFR) with functional covariates and scalar responses, only a relatively few literature is dedicated to this relevant area, which merits further research. We propose a robust spatial autoregressive scalar-on-function regression (RSSoFR) by incorporating a spatial autoregressive parameter and a spatial weight matrix into the SoFR to accommodate spatial dependencies among individuals. The t-distribution assumption for the error terms makes our model more robust than the classical spatial autoregressive models under normal distributions. We estimate the model by firstly projecting the functional predictor onto a functional space spanned by an orthonormal functional basis and then presenting an expectation-maximization (EM) algorithm. Simulation studies show that our estimators are efficient, and are superior in the scenario with spatial correlation and heavy tailed error terms. A real weather dataset demonstrates the superiority of our model to the SoFR in the case of spatial dependence.

Keywords: EM algorithm; FPCA; Functional linear model; Spatial autoregressive model; Spatial dependence; t-distribution

1 INTRODUCTION

Functional data are high-dimensional structured data that vary over a continuous domain. Examples are datasets recorded densely over time, space or time-space, like weather data, stock market data, trajectory data, diffusion tensor imaging (DTI) data and mass spectrometry data. Functional data analysis (FDA), as a new area of statistics, develops statistical methods to analyze information within functional data and has been applied in many subject areas, such as biology, medical sciences, meteorology, econometrics, finance, chemometrics and geophysics (Ramsay and Silverman (2002, 2005)). One of the most important tools in FDA is functional regression, including scalar-on-function regression (SoFR), function-on-scalar regression and function-on-function regression. In particular, the first one, i.e., the classical SoFR (Ramsay and Dalzell (1991); Hastie and Mallows (1993); Hall and Horowitz (2007)) with scalar responses and functional covariates, is of great interest and has been extensively studied. Let Y be a centered scalar response variable, and $X(t)$ be a second-order stochastic process on a compact interval Γ, $E(X(t)) = 0$ and $E(\int_\Gamma X^2(t)dt) < \infty$. A SoFR associates the functional predictor $X(t)$ to the scalar response Y by

$$Y = \int_\Gamma X(t)\beta(t)dt + \epsilon, \tag{1.1}$$

where $\beta(t)$ is the unknown slope function, and ϵ is the random error term, which is independent of $X(t)$ and has zero mean and finite variance.

Regarding model (1.1), there has been a great deal of literature dedicated to its extensions, see Morris (2015) and Reiss et al. (2017) for reviews. According to the roles the functional predictor $X(t)$ plays in regression, the variants of the SoFR can be categorized into three groups, linear functional predictor regression, nonlinear functional predictor regression and nonparametric functional regression. The first group involves the generalized SoFR (Marx and Eilers (1999); James (2002)), the multi-level SoFR (Crainiceanu et al. (2009); Goldsmith et al. (2012)), the functional mixture regression (Fang et al. (2011)) and the partial SoFR (Shin (2009)). For the second team,

multivariate nonlinear regressions with scalar regressors have been extended to the functional predictor case, such as the functional quadratic regression (Yao and Müller (2010)), the single-functional index model (Ait-Saïdi et al. (2008)), the multiple-index model with functional covariates (James and Silverman (2005)) and the continuously additive model (Müller et al. (2013)). The last group, the nonparametric paradigm of the SoFR, is initially studied by Ferraty and Vieu (2006). A key assumption in the aforementioned literature is that individuals are mutually independent, which may crumble. Yet researches concerning the SoFR with spatial dependence are relatively rare.

In reality, however, spatially correlated data are very common. For example, unemployment data (Topa (2001)) and housing prices data show spatial correlation, annual precipitation of a city relates to it's neighboring cities'. The characteristic of these data is that the dependency of two spatial units is determined by the distance between them. This distance can be either Euclidean or more general (Isard et al. (1970)), for example, social distance, policy distance and economic distance. In spatial statistics and spatial economics, an abundance of publications are devoted to analyze data with spatial dependence (Cressie and Wikle (2015); Schabenberger and Gotway (2017); Anselin (1998); Lesage and Pace (2009); Wang et al. (2019)). Discrete entities (areal data) and continuous surfaces (point-referenced data) are the two main research objects (a distinction between them refers to Anselin (2002)). For the last few years, there are many works fitting spatially correlated functional data, to name a few, Zhang et al. (2011), Menafoglio and Secchi (2017), Nerini et al. (2010), Zhang et al. (2016), Giraldo et al. (2017) and Aguilera-Morillo et al. (2017). We find these papers are either interested in point-referenced data (for which kriging methods are commonly used) or related to regressions with functional responses. Articles concerning fitting functional regressors on scalar responses for discrete entities are relatively scanty. In this study, our interest lies in the SoFR for areal data.

To better illustrate our motivation for implementing this research and its significance, here is an application instance. We collected weather data from the China Meteorolog-

ical Yearbook covering the period between 2005 and 2007. These data record monthly mean temperatures and monthly total precipitation in 34 major cities in China. Our aim is to investigate the effect of temperature on precipitation over these three years. The SoFR (1.1) is popularly used to study this problem (see Ramsay and Silverman (2005)). In the data preprocessing, we average the monthly temperatures and the monthly precipitation over 3 years to get their means, and add up the mean monthly total precipitation over 12 months to obtain the mean annual rainfall. The scalar response is the logarithm of the mean annual total precipitation, and the functional covariate is the mean monthly temperature. After applying the SoFR directly to the weather data, we want to examine whether there is spatial dependence among the residuals. The Moran's I test statistic (Cliff and Ord (1972)) is employed (in Section 5 we explained why the weather dataset can be regarded as areal data instead of point-referenced data). We find the resulting value of the Moran's I statistic is 0.48, and the p-value is smaller than 0.001, which indicates significant correlation exists among the residuals of the SoFR.

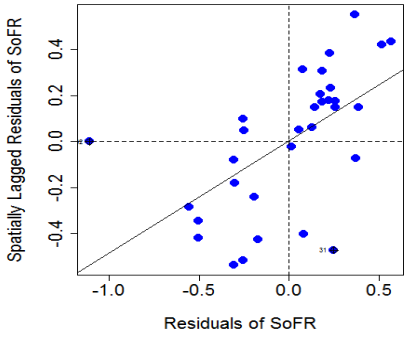

Figure 1: The Moran Scatterplot of the residuals of the SoFR.

For further illustration, we display the Moran Scatterplot of the residuals in Figure 1, from which it can be seen that there is approximately linear relationship between the residuals and the spatially lagged residuals (i.e., $\boldsymbol{\epsilon}$ and $n\boldsymbol{W}\boldsymbol{\epsilon}$, which are defined the same as those in (2.1)) of the SoFR, which means the SoFR may be not appropriate for spatially dependent data. This motivates the incorporation of spatial correlation in the analyses of the SoFR. A detailed analysis of this weather dataset and the construction

of the weight matrix $\boldsymbol{W}$ can be found in Section 5.

When the predictor is scalar, instead of functional, linear regression models including spatial autoregressive (SAR) model, spatial error model (SEM) and spatial Durbin model (SDM) are frequently used to accommodate spatial dependence for areal data (Lesage and Pace (2009)). It is straightforward to borrow concepts from the spatial models for modelling functional data with lattice structure. As the SAR model is representative among these spatial models and has been widely studied, we mainly focus on the SAR model and formulate our new model. Applications of the SAR model can be seen in Case (1991), Topa (2001), and Olubusoye et al. (2016), among others. Estimation methods for the SAR model are described in Ord (1975), Lee (2004), Kelejian and Prucha (2001), Lee (2007) and Lesage and Pace (2009). In an SAR model, a spatial weight matrix is employed to quantify adjacent relations among the observations, and an unknown spatial autoregressive parameter ρ is used to reflect the strength of spatial dependence. We take the advantages of the SAR model and incorporate the spatial correlation into the SoFR model using a spatial autocorrelation parameter and a weight matrix.

We noticed that in the above spatial models, the residual terms are generally presumed to follow normal distributions. This assumption is not realistic, as in applied problems we often expect tails of the error distribution are longer than the those of the normal distributions. Besides, when using the ordinary least squares (OLS) or maximum likelihood estimation method to estimate the spatial models, the estimated parameters can be largely affected by atypical points in dataset. Thus there is a need to consider non-normality error term in the spatial linear models. Fortunately, t-distribution, which has thick tails, provides an alternative. So we assume the residuals follow t-distributions in our model. The proposed new model is named the robust spatial autoregressive scalar-on-function regression (RSSoFR).

In this article, we presented a robust spatial functional linear model (RSSoFR), which extends the independent SoFR to the spatial scenario where the responses are spatially correlated. And by supposing the error terms follow t-distributions, the new model is

more robust than the classical spatial models under normal distribution assumptions. To estimate the RSSoFR, we firstly project the functional predictor onto a smaller functional space spanned by orthonormal basis functions, then propose an expectation-maximization (EM) algorithm to handle the spatial parameters. Two functional bases, the functional partial-least squares (FPLS) basis and the functional principal component (FPC) basis are considered. We also use a spatial cross-validation approach to select models. The simulation results show that our estimation method performs better than the SoFR with t-distributions when the spatial dependencies are present, and the spatial functional linear model (SSoFR) with normal distributions when the error terms have thick tails. We also use the RSSoFR to analyze a real weather dataset of China, which proved to have better fit and prediction results than the SoFR.

Recently, Pineda-Ríos et al. (2019) put forward a functional SAR model with functional predictors, scalar responses and spatially dependent errors, which accounts for the spatial dependence among disturbances. Under the normality assumption, they developed the least squares and maximum likelihood as the estimation methods of the parameters. The differences between Pineda-Ríos et al. (2019)'s paper and ours exist in three folded. First, the two proposed models apply different ways to accommodate the spatial dependence in regression; Second, the former assumes normality for the noise term while the latter presumes t-distribution to deal with the possible non-normality and thick tails; Finally, we proposed an EM algorithm to implement estimation.

The article is organized as follows. In Section 2, we formulate the new model. The proposed estimation method is constructed based on orthonormal functional basis and the EM algorithm in Section 3. The finite-sample performances of the proposed estimators are evaluated through simulation studies in Section 4. Finally, in Section 5, we use a real dataset to document the usefulness of this methodology. We conclude the article with a discussion in Section 6.

2 MODEL SPECIFICATION

2.1 Spatial autoregressive (SAR) model

Ord (1975) proposed a spatial autoregressive (SAR) model with parsimonious parameters. Assume $\{(x_i, y_i)\}_{i=1}^n$ are observed from n spatial units on a lattice, and denote $\boldsymbol{x} = \big(x_1, x_2, \cdots, x_n\big)'$, $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$. The SAR model is,

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{x}\beta + \boldsymbol{\epsilon}, \qquad \epsilon_i \sim N(0, \sigma^2) \tag{2.1}$$

where $\boldsymbol{W} = (w_{ii'})_{n\times n}$ is a pre-specified spatial weight matrix, in which $w_{ii'}$ represents the weight between units i and i', and $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ is the noise term, which is independent of $\boldsymbol{x}$, independently and identically follow normal distributions $N(0, \sigma^2)$. ρ and β are the parameters to be estimated. Here, the scalar parameter $\rho \in (-1, 1)$.

In model (2.1), the spatial weight matrix $\boldsymbol{W}$ is exogenous, with each entry $w_{ii'}$ assigned a value according to the contiguity or distances between units i and i' in different contexts. For spatial contiguity matrices, the value of $w_{ii'}$ is binary,

$$w_{ii'} = \begin{cases} 1, & i \textit{ and } i' \textit{ are neighboring} \\ 0, & i \textit{ and } i' \textit{ are not neighboring} \end{cases}.$$

The assessment of neighbors relies on a known map indicating spatial arrangement of points, which can be regular or irregular. On a regular grid, units are neighbors if they share a border (rook case), a vertex (bishop case), or share either a border or a vertex (queen case) (for more details refer to Anselin (1998)). In an irregularly spaced case, units are also neighboring when they have common edges. In the case of knowing geographical locations of spatial units, $w_{ii'}$ can be functions of distance $d_{ii'}$ between two points. Popular choices are the inverse distance (IV) and the negative exponential (NE) of distance,

$$w_{ii'}^{IV} = \frac{1}{d_{ii'}}, \qquad w_{ii'}^{NE} = e^{-d_{ii'}}. \tag{2.2}$$

When the units are areal data, boundary length information can be also used to form $w_{ii'}$, see for example Decey (1968).

We mention some more general spatial matrices. In a social network, $w_{ii'}$ is normally set to 1 if persons i and i' are friends, 0 otherwise. Note that $\boldsymbol{W}$ is not necessarily a symmetric matrix. For example on a social media Sina Weibo, person i is a follower of person i' but i' does not follow i, we have $w_{ii'} = 1,\ w_{i'i} = 0$. Some economic factors such as the GDP (Gross Domestic Product) and income can be also used to establish $\boldsymbol{W}$, for example, Case et al. (1993) used $w_{ii'} = \frac{1}{|INC_i - INC_{i'}|}$, where INC_i is per capita income in state i. More information for $\boldsymbol{W}$ refers to Isard et al. (1970) and Anselin (1998).

In general, we standardize the spatial matrix $\boldsymbol{W}$ to be a row-normalized matrix; in this matrix, the summation of the row elements is unity, and the entries on the diagonal are zeros. About the interpretations of model (2.1), refer to Anselin (2002).

2.2 Robust spatial autoregressive scalar-on-function regression (RSSoFR)

Following Qu and Lee (2015), consider the spatial processes located on an unevenly spaced lattice $D \subseteq R^d$, $d \geq 1$. And we observe $\{(x_i(t), y_i)\}_{i=1}^n$ from n spatial units on D. Here, $x_i(t)$s are square integrable second-order stochastic processes defined on a compact set Γ. Without loss of generality, we presume Γ is the unit, i.e. $t \in [0, 1]$.

Denote $\boldsymbol{x}(t) = \big(x_1(t), x_2(t), \cdots, x_n(t)\big)'$, we formulate the RSSoFR as

$$\boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \int_0^1 \boldsymbol{x}(t)\beta(t)dt + \boldsymbol{\epsilon}, \qquad \epsilon_i \sim t(\nu) \tag{2.3}$$

where $\boldsymbol{W}$ is pre-defined as the $\boldsymbol{W}$ in model (2.1), $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ independently and identically follow t-distributions with freedom ν. Here, ν, $\beta(t)$ and $\rho \in [0, 1)$ are the parameters to be estimated.

As mentioned previously, ρ is a scale parameter that reflects strength of the impacts

from neighbours. Greater values of ρ indicate that y_i is more strongly affected by its neighbours. The $\boldsymbol{W}$ matrix can be pre-formed as the $\boldsymbol{W}$ pre-specified in model (2.1). We use the inverse distance to construct $\boldsymbol{W}$ in Section 5 for real data analysis. The proposed RSSoFR is more general than the SoFR and the SAR model.

- When $\rho = 0$, the RSSoFR reduces to an SoFR.

- When $\boldsymbol{x}(t)$ is free of t, the RSSoFR degenerates into an SAR model with t-distributions.

To provide better insight onto the new model, we reformulate equation (2.3) as the following equivalent expression,

$$\boldsymbol{y} = (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1} \int_0^1 \boldsymbol{x}(t)\beta(t)dt + (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}\boldsymbol{\epsilon},$$

which shows how $\boldsymbol{y}$ is generated. We can know the mean of $\boldsymbol{y}$ given $\boldsymbol{x}(t)$ is $E\big(\boldsymbol{y}|\boldsymbol{x}(t)\big) = (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1} \int_0^1 \boldsymbol{x}(t)\beta(t)dt$. Thus the spatial process of $\boldsymbol{y}$ is not stationary, i.e., the mean of $\boldsymbol{y}$ depends on the spatial units' locations through the functional covariate $\boldsymbol{x}(t)$. The error terms $(\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}\boldsymbol{\epsilon}$ shows the residuals of y_is are spatially correlated. As each ϵ_i follows a t-distribution, the MLE method can not be used. We put forward an EM algorithm to estimate the parameters in (2.3).

3 ESTIMATION METHOD

In this section, we first expand the functional predictor by orthogonal functional basis, then present an EM algorithm to obtain the estimators of the spatial autocorrelation parameter ρ, the slope function $\beta(t)$, and the freedom parameter ν in the RSSoFR (2.3).

3.1 Expand the functional predictor by orthonormal basis

Note that, before estimating the parameters of our model, data representation methods, such as smoothing and interpolation, should be used to convert discretely recorded data $x_i(t_j)$ to curves $x_i(t)$. And $\boldsymbol{x}(t), \boldsymbol{y}$ are centered in advance.

A standard method to handle the functional predictor in the SoFR is expressing the curves by a linear combination of orthonormal basis functions. We also firstly expand the functions in (2.3) so that the functional term $\int_0^1 \boldsymbol{x}(t)\beta(t)dt$ can be represented by finite real vectors. Suppose $\{\phi_j(t)\}_{j=1}^{\infty}$ is a standard orthogonal basis of L^2 space, which is composed of square integrable functions. Then $x_i(t)$ and $\beta(t)$ can be written as $x_i(t) = \sum_{j=1}^{\infty} a_{ij}\phi_j(t)$ and $\beta(t) = \sum_{j=1}^{\infty} b_j\phi_j(t)$, where $a_{ij} = \int_0^1 x_i(t)\phi_j(t)dt$, $b_j = \int_0^1 \beta(t)\phi_j(t)dt$. Moreover, $\int_0^1 x_i(t)\beta(t)dt = \sum_{j=1}^{\infty} a_{ij}b_j$ by orthogonality. Denote $\boldsymbol{a}_j = (a_{1j}, a_{2j}, \ldots, a_{nj})'$, we can rewrite model (2.3) as

$$\boldsymbol{y} = \rho\boldsymbol{W}\boldsymbol{y} + \sum_{j=1}^{\infty} \boldsymbol{a}_j b_j + \boldsymbol{\epsilon}, \tag{3.1}$$

which is easier to cope with under the numerical vectors framework. In practice, the estimation of $\beta(t)$ is an ill-posed problem. Regularization procedure is thus needed. Some authors add a penalty term in the objective function to put constraints on $\hat{\beta}(t)$ (Cardot et al. (2003); Crambes et al. (2009)). Here, we project $\boldsymbol{x}(t)$ and $\beta(t)$ onto a finite dimensional space spanned by m basis functions $\{\phi_j(t)\}_{j=1}^{m}$. Therefore, expression (3.1) has the truncated form

$$\boldsymbol{y} \approx \rho\boldsymbol{W}\boldsymbol{y} + \sum_{j=1}^{m} \boldsymbol{a}_j b_j + \boldsymbol{\epsilon}. \tag{3.2}$$

Regarding the orthonormal functional basis $\phi_j(t)$s, we can use the Fourier basis, the FPC basis (Dauxois et al. (1982); James et al. (2000); Li and Hsing (2010)) and the FPLS basis (Preda and Saporta (2005); Preda et al. (2007); Preda and Saporta (2007); Aguilera et al. (2010); Delaigle and Hall (2012)). Here, we both introduce the FPC basis and the FPLS basis in the process of estimation. As these two bases are adaptive

to properties of data, whereas other bases independent of data will make it hard to decide which m terms should be included in the truncated RSSoFR (3.2).

3.1.1 *Functional principal component (FPC) basis*

Let $K(s,t)$ denote the covariance function of $X(t)$, i.e. $K(s,t) = \mathrm{Cov}(X(t), X(s))$. By Mercer's theorem, the spectral decomposition of $K(s,t)$ is then $K(s,t) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(s)$ $\varphi_j(t)$, where $\lambda_1 > \lambda_2 > \cdots > 0$ are eigenvalues and $\{\varphi_j(t)\}_{j=1}^{\infty}$ are the corresponding orthogonal eigenfunctions. According to the Karhunen-Loève expansion, $X(t)$ can be expanded as $X(t) = \sum_{j=1}^{\infty} a_j \varphi_j(t)$, where a_js are uncorrelated random variables with mean zero and variance $E(a_j^2) = \lambda_j$, $a_j = \int_0^1 X(t)\varphi_j(t)dt$.

For the observations $\{y_i, x_i(t)\}_{i=1}^n$, the empirical version of $K(s,t)$ is $\hat{K}(s,t) = \frac{1}{n}\sum_{i=1}^n x_i(s)$ $x_i(t)$. Moreover, it can be shown that $\hat{K}(s,t) = \sum_{j=1}^n \hat{\lambda}_j \hat{\varphi}_j(s)\hat{\varphi}_j(t)$, where $\hat{\lambda}_j$ and $\hat{\varphi}_j(t)$ are the estimators of $\varphi_j(t)$ and λ_j, respectively. For the ith observation $x_i(t)$, the estimator of a_{ij} is then $\hat{a}_{ij} = \int_0^1 x_i(t)\hat{\varphi}_j(t)$, and $x_i(t)$ can be written as $x_i(t) = \sum_{j=1}^n \hat{a}_{ij}\hat{\varphi}_j(t)$. Similarly, based on the estimated orthonormal functional basis $\{\hat{\varphi}_j(t)\}_{j=1}^n$, $\beta(t)$ has the expansion $\beta(t) = \sum_{j=1}^n \tilde{b}_j \hat{\varphi}_j(t)$ with $\tilde{b}_j = \int_0^1 \beta(t)\hat{\varphi}_j(t)dt$. Therefore, the sample counterpart of (3.2) is

$$\boldsymbol{y} \approx \rho \boldsymbol{W}\boldsymbol{y} + \sum_{j=1}^{m} \hat{\boldsymbol{a}}_{\boldsymbol{j}} \tilde{b}_j + \boldsymbol{\epsilon}. \tag{3.3}$$

where $\hat{\boldsymbol{a}}_{\boldsymbol{j}} = (\hat{a}_{1j}, \hat{a}_{2j}, \cdots, \hat{a}_{nj})'$.

3.1.2 *Functional partial-least squares (FPLS) basis*

In the analysis of numerical data, the partial-least squares (PLS) regression is an efficient alternative to the principal component regression (PCR) by taking the response variable $\boldsymbol{y}$ onto account. Like the PCR, the PLS regression also provides a set of orthogonal basis functions. De Jong (1993) has proved that with the same number of components, the PLS regression has a better fit than the PCR. Especially when a large proportion of the variations of $x_i(t)$s do not explain the responses $\boldsymbol{y}$, the PLS regression has competitive advantages over the FPC regression.

Here, to form the FPLS basis functions, we first neglect the autoregressive term $\rho W\boldsymbol{y}$ of (2.3), which means building the basis without considering spatial correlation. Then, use an iterative process introduced by Preda and Saporta (2005) to get the basis. Supposing we construct m FPLS basis functions, the main steps are,

(1) Begin from $l = 1$. Also write $\boldsymbol{x}_l(t) = \boldsymbol{x}(t),\ \boldsymbol{y}_l = \boldsymbol{y}$.

(2) Obtain a square integrable weight function $\omega_l(t)$ evaluated by $\omega_l(t) = \frac{E[\boldsymbol{y}_l\boldsymbol{x}_l(t)]}{\|E[\boldsymbol{y}_l\boldsymbol{x}_l(t)]\|}$, which maximizes $\underset{\|\omega_l(t)\|=1}{Cov}\big(\boldsymbol{y}_l, \int_0^1 \boldsymbol{x}_l(t)\omega_l(t)dt\ \big)$, where $\|\cdot\|$ is the norm, i.e., $\|\omega_l(t)\| = \sqrt{\int_0^1 \omega_l^2(t)dt}$.

(3) Regress $\boldsymbol{x}_l(t)$ and $\boldsymbol{y}_l$ on $\boldsymbol{a}_l$, separately, where $\boldsymbol{a}_l = \int_0^1 \boldsymbol{x}_l(t)\omega_l(t)dt$. That is $\boldsymbol{x}_l(t) = p_l(t)\boldsymbol{a}_l + \boldsymbol{\epsilon}_l^x(t)$ and $\boldsymbol{y}_l = q_l\boldsymbol{a}_l + \boldsymbol{\epsilon}_l^y$, where $p_l(t) = \frac{E[\boldsymbol{x}_l(t)\boldsymbol{a}_l]}{\|\boldsymbol{a}_l\|^2},\ q_l = \frac{E[\boldsymbol{y}_l\boldsymbol{a}_l]}{\|\boldsymbol{a}_l\|^2}$.

(4) Stop when $l = m$. Otherwise, set $\boldsymbol{x}_{l+1}(t) = \boldsymbol{\epsilon}_l^x(t),\ \boldsymbol{y}_{l+1} = \boldsymbol{\epsilon}_l^y$, and back to step 2.

The set of weight functions $\{\omega_j(t)\}_{j=1}^m$ is the FPLS basis. And the orthogonal components $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ are the regressors $\boldsymbol{a}_j$ of the truncated model (3.2). In practice, we compute $E[\boldsymbol{y}_l\boldsymbol{x}_l(t)]$, $E[\boldsymbol{x}_l(t)\boldsymbol{a}_l]$ and $E[\boldsymbol{y}_l\boldsymbol{a}_l]$ by their empirical versions. Denote $\hat{\boldsymbol{a}}_j$ as the estimator of $\boldsymbol{a}_j$, we have the same expression as (3.3), which is the truncated empirical form of (3.1).

3.2 An EM algorithm for the truncated RSSoFR

In this subsection we focus on (3.3) and propose an EM algorithm to estimate ρ, $\boldsymbol{b} = (\tilde{b}_1, \tilde{b}_2, \cdots, \tilde{b}_m)'$ and ν. Defining $\boldsymbol{A} = (\hat{a}_{ij})_{n\times m}$, the expression (3.3) can be written as

$$\boldsymbol{y} \approx \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{A}\boldsymbol{b} + \boldsymbol{\epsilon}. \tag{3.4}$$

It is not easy to directly write out the probability density function of $\boldsymbol{y}$ based on the noise term as $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ independently and identically follow t-distributions. However, we know the fact that a t-distribution can be regarded as a scale mixture of normal distributions, i.e., if u_i follows a gamma distribution $f(u_i) = \frac{1}{\Gamma(\frac{\nu}{2})}(\frac{\nu}{2})^{\frac{\nu}{2}}(u_i)^{\frac{\nu}{2}-1}e^{-\frac{\nu}{2}u_i}$, and $\epsilon_i \mid u_i \sim N(0, \frac{\sigma^2}{u_i})$, then the marginal density $f(\epsilon_i)$ of ϵ_i is a t-distribution with

freedom ν and scale parameter σ,

$$f(\epsilon_i) = \frac{\Gamma(\frac{\nu+1}{2})\sigma^{-1}}{(\pi\nu)^{\frac{1}{2}}\Gamma(\frac{\nu}{2})\{1+\frac{\epsilon^2}{\sigma^2\nu}\}^{\frac{1}{2}(\nu+1)}}.$$

Thus by introducing latent variables $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)'$, which are independent of $\boldsymbol{x}(t)$ and independently and identically distributed (i.i.d.), we put forward an EM algorithm to estimate (3.4).

Given $\boldsymbol{u}$, (3.4) can be expressed as

$$\boldsymbol{y} \mid \boldsymbol{u} \approx (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}(\boldsymbol{A}\boldsymbol{b} + \sigma\boldsymbol{U}^{-\frac{1}{2}}\boldsymbol{\eta})$$

$$\boldsymbol{\eta} \sim N(\boldsymbol{0}, \boldsymbol{I}_n), \qquad \boldsymbol{U} = diag(\boldsymbol{u})$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)'$. Then the complete log-likelihood function for $\boldsymbol{y}$ and $\boldsymbol{u}$ is

$$\begin{aligned} \ln L(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{u}) &= \ln f(\boldsymbol{y}|\boldsymbol{u}) + \sum_{i=1}^{n} \ln f(u_i) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{y}}| - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{y}})'\boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{\boldsymbol{y}}) + \sum_{i=1}^{n} \ln f(u_i), \end{aligned} \tag{3.5}$$

where $\boldsymbol{\theta} = (\rho, \boldsymbol{b}, \sigma^2, \nu)$, $\boldsymbol{\mu}_{\boldsymbol{y}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}}$ are the mean and the covariance of $\boldsymbol{y}$, respectively,

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{y}} &= (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}\boldsymbol{A}\boldsymbol{b}, \\ \boldsymbol{\Sigma}_{\boldsymbol{y}} &= \left((\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}\right)' \cdot \sigma^2 \cdot \boldsymbol{U}^{-1} \cdot (\boldsymbol{I}_n - \rho\boldsymbol{W})^{-1}. \end{aligned}$$

Denote the constant term in (3.5) by C, the complete log-likelihood function is

$$\begin{aligned} \ln L(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{u}) =& C + \frac{1}{2}\sum_{i=1}^{n} \ln u_i - \frac{n}{2}\ln\sigma^2 + \ln|\boldsymbol{I}_n - \rho\boldsymbol{W}| - \frac{\boldsymbol{\epsilon}'\boldsymbol{U}\boldsymbol{\epsilon}}{2\sigma^2} \\ &+ \sum_{i=1}^{n}\left\{(\frac{\nu}{2}-1)\ln u_i - \frac{\nu}{2}u_i - \ln\Gamma(\frac{\nu}{2}) + \frac{\nu}{2}\ln\frac{\nu}{2}\right\}. \end{aligned} \tag{3.6}$$

Where $\boldsymbol{\epsilon} = \boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{A}\boldsymbol{b}$. Observing the above, if we known $\boldsymbol{U}$, the estimators of $\rho, \boldsymbol{b}, \sigma^2$ are directly obtained by the quasi-maximum likelihood estimation method

introduced by Lee (2004). And the freedom ν can be also estimated by maximizing the last term in (3.6), i.e.

$$\sum_{i=1}^{n}\Big\{(\frac{\nu}{2}-1)\ln u_i-\frac{\nu}{2}u_i-\ln\Gamma(\frac{\nu}{2})+\frac{\nu}{2}\ln\frac{\nu}{2}\Big\}.$$

However, $\boldsymbol{u}$ is missing data, and next we present the EM algorithm.

In E-step, given the kth estimate of $\boldsymbol{\theta}$, the computation of $E\big(\ln L(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})\big)$ comes down to compute $E(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})$ and $E(\sum_{i=1}^{n}\ln u_i|\boldsymbol{y},\boldsymbol{\theta}^{(k)})$. Note that the fact that the conditional distribution of u_i given ϵ_i is also a gamma distribution $G(\nu_1,\nu_2)$, where $\nu_1=\frac{1}{2}(\nu+1)$, $\nu_2=\frac{1}{2}(\nu+\frac{\epsilon_i^2}{\sigma^2})$, and we have $E(u_i|\epsilon_i)=\frac{\nu+1}{\nu+(\epsilon_i^2)/(\sigma^2)}$. Hence

$$E(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})=\Big(\frac{\nu^{(k)}+1}{\nu^{(k)}+\delta_1^{(k)}},\frac{\nu^{(k)}+1}{\nu^{(k)}+\delta_2^{(k)}},\ldots,\frac{\nu^{(k)}+1}{\nu^{(k)}+\delta_n^{(k)}}\Big)',\qquad \delta_i^{(k)}=\frac{(\epsilon_i^2)^{(k)}}{(\sigma^2)^{(k)}}.$$

We also have

$$E\big(\sum_{i=1}^{n}\ln u_i|\boldsymbol{y},\boldsymbol{\theta}^{(k)}\big)=\sum_{i=1}^{n}\Big\{\psi\big(\frac{\nu^{(k)}+1}{2}\big)-\ln\big(\frac{\nu^{(k)}+\delta_i^{(k)}}{2}\big)\Big\},$$

where $\psi(s)=\frac{\partial\ln\Gamma(s)}{\partial s}$ is the Digamma function.

And in M-step, the $(k+1)$th estimate of $\boldsymbol{\theta}$ maximizes

$$\begin{aligned}E\big(\ln L(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})\big)\ \propto\ \Big\{&\frac{\nu}{2}E\big(\sum_{i=1}^{n}\ln u_i|\boldsymbol{y},\boldsymbol{\theta}^{(k)}\big)-\frac{\nu}{2}\boldsymbol{\tau}_nE(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})-n\ln\Gamma(\frac{\nu}{2})+\\&n\frac{\nu}{2}\ln\frac{\nu}{2}-\frac{n}{2}\ln(\sigma^2)+\ln|\boldsymbol{I}_n-\rho\boldsymbol{W}|-\frac{\boldsymbol{\epsilon}'E(\boldsymbol{U}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})\boldsymbol{\epsilon}}{2\sigma^2}\Big\},\end{aligned}$$

where $\boldsymbol{\tau}_n$ is an n-dimensional vector of ones, $E(\boldsymbol{U}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})=diag\big(E(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})\big)$. Thus we can get $\nu^{(k+1)}$ by maximizing

$$\frac{\nu}{2}E\big(\sum_{i=1}^{n}\ln u_i|\boldsymbol{y},\boldsymbol{\theta}^{(k)}\big)-\frac{\nu}{2}\boldsymbol{\tau}_nE(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(k)})-n\ln\Gamma(\frac{\nu}{2})+n\frac{\nu}{2}\ln\frac{\nu}{2}.$$

And obtain $\{\rho^{(k+1)}, \boldsymbol{b}^{(k+1)}, (\sigma^2)^{(k+1)}\}$ by maximizing

$$\frac{n}{2}\ln(\sigma^2) + \ln|\boldsymbol{I}_n - \rho\boldsymbol{W}| - \frac{\boldsymbol{\epsilon}' E(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)})\boldsymbol{\epsilon}}{2\sigma^2}. \tag{3.7}$$

Observing the above if ρ is known, $\boldsymbol{b}^{(k+1)}$ and $(\sigma^2)^{(k+1)}$ are respectively,

$$\begin{aligned}
\boldsymbol{b}^{(k+1)}(\rho) &= \big(\boldsymbol{A}' E(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)})\boldsymbol{A}\big)^{-1}\big(\boldsymbol{A}' E(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)})\boldsymbol{y} - \rho\boldsymbol{A}' E(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)})\boldsymbol{W}\boldsymbol{y}\big) \\
(\sigma^2)^{(k+1)}(\rho) &= \frac{1}{n}\big(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{A}\boldsymbol{b}^{(k+1)}(\rho)\big)' \cdot E(\boldsymbol{U}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)}) \cdot \big(\boldsymbol{y} - \rho\boldsymbol{W}\boldsymbol{y} - \boldsymbol{A}\boldsymbol{b}^{(k+1)}(\rho)\big).
\end{aligned} \tag{3.8}$$

Thus similar to the quasi-maximum likelihood estimation method (Lee (2004)), we substitute $\boldsymbol{b}^{(k+1)}(\rho)$ and $(\sigma^2)^{(k+1)}(\rho)$ onto (3.7) and drop the constant term. Then $\rho^{(k+1)}$ is the solution of

$$\arg\max_{\rho} \quad \big\{ -\frac{n}{2}\ln\big((\sigma^2)^{(k+1)}(\rho)\big) + \ln|\boldsymbol{I}_n - \rho\boldsymbol{W}|\big\}.$$

The optimization of the above refers to Ord (1975). Once $\rho^{(k+1)}$ is obtained, replacing ρ with $\rho^{(k+1)}$ in (3.8) yields $\boldsymbol{b}^{(k+1)}$ and $(\sigma^2)^{(k+1)}$, respectively.

The estimator of $\boldsymbol{\theta}$ is finally obtained by repeating E-step and M-step until $\max\{|\rho^{(k+1)} - \rho^{(k)}|, |\boldsymbol{b}^{(k+1)} - \boldsymbol{b}^{(k)}|, |(\sigma^2)^{(k+1)} - (\sigma^2)^{(k)}|\}$ is within a given threshold. And $\hat{\beta}(t)$ can be reconstructed by the estimator of b, i.e. $\hat{\boldsymbol{b}} = (\hat{b}_1, \cdots, \hat{b}_m)$. That is

$$\hat{\beta}(t) = \sum_{j=1}^{m} \hat{b}_j \hat{\phi}_j(t). \tag{3.9}$$

For clarity, we summarize the main steps of the estimation procedure in Algorithm 1.

3.3 Choosing the truncation parameter for the RSSoFR

We introduce two ways to determine the truncation parameter m. The first is the percentage of variance explained (PVE) for predictors, which is often adopted for the FPC basis. The second is based on cross validation, which is a more general approach.

Algorithm 1 Main steps of the estimation procedure

1: Represent the functional predictor and the slope function using the FPLS basis or the FPC basis. In this step, after an appropriate truncation parameter is given, the RSSoFR approximates an SAR model with t-distributions whose covariates are the component scores of $x_i(t)$s, as shown in (3.4).

2: Determine the estimators for $\rho, \boldsymbol{b}, \sigma^2, \nu$ in (3.4) by the EM algorithm.

E-step : Calculate $u_i^{(k+1)}$ and $(\ln u_i)^{(k+1)}$.

$$u_i^{(k+1)} = E(u_i|\boldsymbol{y}, \boldsymbol{\theta}^{(k)}) = \frac{\nu^{(k)}+1}{\nu^{(k)}+\delta_i^{(k)}}$$

$$(\ln u_i)^{(k+1)} = E(\ln u_i|\boldsymbol{y}, \boldsymbol{\theta}^{(k)}) = \psi\big(\frac{\nu^{(k)}+1}{2}\big) - \ln\big(\frac{\nu^{(k)}+\delta_i^{(k)}}{2}\big)$$

where $\psi(s)$ is the Digamma function, $\delta_i^{(k)} = \frac{(\epsilon_i^2)^{(k)}}{(\sigma^2)^{(k)}}$ and $\epsilon_i^{(k)}$ is the ith element of $\boldsymbol{y} - \rho^{(k)}\boldsymbol{W}\boldsymbol{y} - \boldsymbol{A}\boldsymbol{b}^{(k)}$.

M-step : Update the parameters $\{\nu^{(k)}, \rho^{(k)}, \boldsymbol{b}^{(k)}, (\sigma^2)^{(k)}\}$.

$$\rho^{(k+1)} = \arg\max_{\rho} \quad \big\{ -\frac{n}{2}\ln\big((\sigma^2)^{(k+1)}(\rho)\big) + \ln|\boldsymbol{I}_n - \rho\boldsymbol{W}|\big\}$$

$$\boldsymbol{b}^{(k+1)} = \big(\boldsymbol{A}'\boldsymbol{U}^{(k+1)}\boldsymbol{A}\big)^{-1}\big(\boldsymbol{A}'\boldsymbol{U}^{(k+1)}\boldsymbol{y} - \rho^{(k+1)}\boldsymbol{A}'\boldsymbol{U}^{(k+1)}\boldsymbol{W}\boldsymbol{y}\big)$$

$$(\sigma^2)^{(k+1)} = \frac{1}{n}\big(\boldsymbol{y} - \rho^{(k+1)}\boldsymbol{W}\boldsymbol{y} - \boldsymbol{A}\boldsymbol{b}^{(k+1)}\big)' \cdot \boldsymbol{U}^{(k+1)} \cdot \big(\boldsymbol{y} - \rho^{(k+1)}\boldsymbol{W}\boldsymbol{y} - \boldsymbol{A}\boldsymbol{b}^{(k+1)}\big)$$

$$\nu^{(k+1)} = \arg\max_{\nu} \sum_{i=1}^{n}\Big\{\frac{\nu}{2}(\ln u_i)^{(k+1)} - \frac{\nu}{2}u_i^{(k+1)} - \ln\Gamma(\frac{\nu}{2}) + \frac{\nu}{2}\ln\frac{\nu}{2}\Big\}$$

where $\boldsymbol{U}^{(k+1)} = \mathrm{diag}\big((u_1^{(k+1)}, \cdots, u_n^{(k+1)})'\big)$.

3: Reconstruct the estimator for $\beta(t)$ in the RSSoFR. The slope function is constructed using the FPLS basis or the FPC basis mentioned in Step 1 and the coefficient $\hat{\boldsymbol{b}}$ estimated in Step 2, as shown in (3.9). The other estimators are obtained directly from Step 2.

To compare the new RSSoFR with the SSoFR and the SoFR in the numerical experiments, we use the first method to determine m. If the PVE is set to be 80%, the truncation parameter m is subject to $\min_l \{(\sum_{j=1}^{l}\hat{\lambda}_j)/(\sum_{j=1}^{n}\hat{\lambda}_j) \geq 80\%\}$.

In a non-spatial scenario, cross-validation method is commonly conducted by each time extracting an object from data as test set, and using the remaining as training set to predict the extracted one. In a spatial context, cross validation can be also applied but notice that the modification of spatial structure resulting from removing

one unit should be considered. Denote the objects in training set as in-sample units, whose dependent variable Y_s and independent variables X_s are both observed, and the objects in test set as out-sample units, whose observed variable is explanatory variable X_o and the response Y_o is unknown, which need to be predicted. We call the problem of predicting Y_o using X_s, X_o and Y_s the out-of-sample prediction.

Goulard et al. (2017) discussed the out-of-sample prediction for the SAR model and concluded the "BP" predictor, which is based on Goldberger formula, behaves well among other existing predictors. The calculation formula of the BP predictor is

$$\hat{Y}_o^{BP} = \hat{Y}_o^{TC} - Q_o^{-1}Q_{os}(Y_s - \hat{Y}_s^{TC}), \qquad \hat{Y}^{TC} = (I_n - \hat{\rho}W)^{-1}X\hat{\beta} = \begin{pmatrix} \hat{Y}_s^{TC} \\ \hat{Y}_o^{TC} \end{pmatrix},$$

where Q is the precision matrix and

$$Q = \frac{1}{\hat{\sigma}^2}(I_n - \hat{\rho}(W' + W) + \hat{\rho}^2 W'W) = \begin{pmatrix} Q_s & Q_{so} \\ Q_{os} & Q_o \end{pmatrix}.$$

In all, we use a spatial cross-validation method, which employs the BP predictor, to select the number of components. Specifically, we extract one observation at a time from the dataset and use the remaining to predict it. When all the observations in the dataset have been extracted once, we define the following cross-validated prediction error (PE) for truncation parameter m,

$$PE(m) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i^m)^2, \tag{3.10}$$

where $\hat{y}_i^m$ is the predicted y_i using m components and $n-1$ observations. For all the alternatives of m, we choose the one with the smallest PE.

4 SIMULATION STUDY

Several simulation studies are conducted to evaluate the finite-sample performances of the proposed estimators for $\rho, \nu, \beta(t)$. All of the computations were carried out in the R environment, and we used the R packages 'spdep', 'fda' and 'fda.usc' to implement the proposed procedure.

Because the spatial dependence and the t-distribution are of interest in this study, we compare the proposed RSSoFR with the following two models

(a) The SoFR with i.i.d. t-distributions

$$\boldsymbol{y} = \int_0^1 \boldsymbol{x}(t)\beta(t)dt + \boldsymbol{\epsilon}, \qquad \epsilon_i \sim t(\nu). \tag{4.1}$$

(b) The SSoFR with i.i.d. normal distributions

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \int_0^1 \boldsymbol{x}(t)\beta(t)dt + \boldsymbol{\epsilon}, \qquad \epsilon_i \sim N(0, \sigma^2). \tag{4.2}$$

These three models are estimated all by firstly expanding the functional regressor using the FPC basis. Here we set the PVE to be identical to 80%. Then we have

(1) The SSoFR (4.2) approximates an SAR model with i.i.d. normal distributions, which can be estimated by the quasi-maximum likelihood estimation method (Lee (2004)).

(2) The SoFR (4.1) approximates an ordinary multivariate linear model with i.i.d. t-distributions, which can be estimated by the EM algorithm (Peel and McLachlan (2000)).

Regarding $\hat{\beta}(t)$ for the three models, they can be reconstructed by the basis functions as that showed by (3.9). Different degrees of spatial effects are considered, $\rho = \{0, 0.5, 0.8\}$. Note that when $\rho = 0$, the SSoFR (2.3) reduces to the SoFR (4.1).

As for the spatial scenario, we adopt the rook matrix by randomly apportioning n agents on a regular square grid of cells; each agent is located on a cell. In this context,

if the grid has R rows and T columns, then the sample size $n = R \times T$. Units that share an edge are neighbours. This definition ensures the units in the inner field of the grid have four neighbours, the units in the corners have two neighbours, and the units along the boarders have three neighbours. Therefore, the spatial matrix is an adjacent matrix with each entry $w_{ii'} = 1$ if units i and i' are neighbours and $w_{ii'} = 0$ otherwise. We set $n = \{10 \times 15, 15 \times 20, 20 \times 30\}$ in the simulation. The spatial weight matrix is row-normalized in all cases.

For the functional part of equation (2.3), we employ the same form as the functions in the SoFR by Hall and Horowitz (2007). Specifically, we generate the simulation data $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$ using

$$\boldsymbol{y} = (\boldsymbol{I}_n - \rho \boldsymbol{W})^{-1} \Big(\int_0^1 \boldsymbol{x}(t)\beta(t)dt + 0.5\boldsymbol{\epsilon} \Big)$$

(1) $\epsilon_i \sim N(0,1)$, $\epsilon_i \sim t(1)$, $\epsilon_i \sim t(3)$

(2) $\rho = 0$, $\rho = 0.5$, $\rho = 0.8$

The functional predictor $\boldsymbol{x}(t) = (x_1(t), x_2(t), \cdots, x_n(t))'$ is produced with values of $x_i(t)$ independently generated by $x_i(t) = \sum_{j=1}^{50} a_j Z_j \varphi_j(t)$, where $a_j = (-1)^{j+1} j^{-\gamma/2}$ with $\gamma = 1.1$ and 2, respectively; $Z_j \sim U[-\sqrt{3}, \sqrt{3}]$ and $\varphi_j(t) = \sqrt{2}\cos(j\pi t)$. Similarly, the coefficient function $\beta(t)$ is generated according to $\beta(t) = \sum_{j=1}^{50} b_j \varphi_j(t)$, where $b_1 = 0.3$ and $b_j = 4(-1)^{j+1} j^{-2}, j \geq 2$. Note that the eigenvalues of the covariance function $\hat{K}(u, v)$ play a vital role in determining the estimation accuracy of $\beta(t)$. We consider two cases. In case 1, $\gamma = 1.1$, where the eigenvalues are well spaced and the slope function can be accurately estimated. In case 2, $\gamma = 2$, where the closely spaced eigenvalues can cause the estimator $\hat{\beta}(t)$ to display poor performance.

The experiment is repeated 500 times in each setting. And we suppose $x_i(t)$s are observed on 101 equispaced points, i.e. $t_j = 0, \frac{1}{100}, \cdots, \frac{99}{100}, 1$. The selected number of components are $\{7, 8, 9\}$ for $\gamma = 1.1$ and 2 for $\gamma = 2$. We assess the behaviour of the estimator $\hat{\rho}$ in terms of the mean bias and its standard deviation. Concerning $\hat{\beta}(t)$, we evaluate its performances in terms of the integrated mean-square error $IMSE =$

$\int_0^1(\hat{\beta}(t)-\beta(t))^2dt$. We summarize the estimation results for ρ in Table 1, $\beta(t)$ in Table 2 and ν in Table 3. Examination of Table 1-3 leads to the following conclusions.

Table 1: The biases and its standard deviations (in brackets) of $\hat{\rho}$.

			$\gamma=1.1$			$\gamma=2$		
ρ	ϵ_i	n	RSSoFR	SSoFR	SoFR	RSSoFR	SSoFR	SoFR
0	normal	150	−0.0059 (0.0887)	−0.0035 (0.0769)	–	−0.0086 (0.0969)	−0.0055 (0.0862)	–
		300	−0.0025 (0.0608)	−0.0027 (0.0526)	–	−0.0033 (0.0646)	−0.0021 (0.0571)	–
		600	−0.0009 (0.0429)	−0.0004 (0.0362)	–	−0.0011 (0.0479)	−0.0006 (0.0435)	–
0	t(1)	150	0.0002 (0.0266)	−0.0145 (0.0828)	–	−0.0009 (0.0291)	−0.0165 (0.0884)	–
		300	0.0004 (0.0127)	−0.0092 (0.0600)	–	−0.0003 (0.0137)	−0.0090 (0.0552)	–
		600	−0.0006 (0.0067)	−0.0042 (0.0345)	–	−0.0003 (0.0071)	−0.0009 (0.0338)	–
0	t(3)	150	−0.0059 (0.0801)	−0.0055 (0.0865)	–	−0.0050 (0.0862)	−0.0101 (0.0952)	–
		300	−0.0061 (0.0550)	−0.0094 (0.0596)	–	−0.0020 (0.0640)	0.0000 (0.0708)	–
		600	−0.0021 (0.0398)	−0.0011 (0.0467)	–	−0.0026 (0.0450)	−0.0033 (0.0511)	–
0.5	normal	150	−0.0121 (0.0733)	−0.0103 (0.0622)	–	−0.0129 (0.0775)	−0.0114 (0.0682)	–
		300	−0.0078 (0.0500)	−0.0058 (0.0430)	–	−0.0039 (0.0554)	−0.0021 (0.0495)	–
		600	−0.0041 (0.0362)	−0.0030 (0.0315)	–	−0.0021 (0.0404)	−0.0014 (0.0346)	–
0.5	t(1)	150	−0.0016 (0.0220)	−0.0192 (0.0765)	–	0.0016 (0.0217)	−0.0146 (0.0674)	–
		300	−0.0012 (0.0111)	−0.0085 (0.0464)	–	−0.0005 (0.0096)	−0.0077 (0.0487)	–
		600	−0.0002 (0.0062)	−0.0053 (0.0364)	–	0.0000 (0.0052)	−0.0034 (0.0296)	–
0.5	t(3)	150	−0.0105 (0.0712)	−0.0177 (0.0754)	–	−0.0190 (0.0785)	−0.0218 (0.0815)	–
		300	−0.0036 (0.0467)	−0.0069 (0.0533)	–	−0.0082 (0.0504)	−0.0089 (0.0557)	–
		600	−0.0021 (0.0309)	−0.0016 (0.0346)	–	−0.0024 (0.0354)	−0.0059 (0.0393)	–
0.8	normal	150	−0.0132 (0.0482)	−0.0094 (0.0396)	–	−0.0204 (0.0523)	−0.0179 (0.0439)	–
		300	−0.0085 (0.0312)	−0.0076 (0.0272)	–	−0.0114 (0.0380)	−0.0103 (0.0333)	–
		600	−0.0038 (0.0217)	−0.0031 (0.0186)	–	−0.0040 (0.0253)	−0.0037 (0.0216)	–
0.8	t(1)	150	−0.0018 (0.0243)	−0.0230 (0.0502)	–	−0.0025 (0.0150)	−0.0192 (0.0465)	–
		300	−0.0002 (0.0061)	−0.0113 (0.0375)	–	−0.0006 (0.0059)	−0.0113 (0.0308)	–
		600	0.0001 (0.0028)	−0.0077 (0.0263)	–	−0.0001 (0.0029)	−0.0075 (0.0219)	–
0.8	t(3)	150	−0.0138 (0.0453)	−0.0151 (0.0455)	–	−0.0140 (0.0475)	−0.0181 (0.0521)	–
		300	−0.0054 (0.0293)	−0.0073 (0.0338)	–	−0.0067 (0.0321)	−0.0099 (0.0362)	–
		600	−0.0038 (0.0196)	−0.0052 (0.0227)	–	−0.0041 (0.0220)	−0.0052 (0.0251)	–

(1) In Table 1 for $\hat{\rho}$, the biases are small and the standard deviations display a decrease pattern as sample size increases. When $\epsilon_i \sim t(1)$, $\hat{\rho}$ of the RSSoFR always performs

Table 2: The empirical average IMSE and its standard deviations (in brackets) of $\hat{\beta}(t)$.

			$\gamma = 1.1$			$\gamma = 2$		
ρ	ϵ_i	n	RSSoFR	SSoFR	SoFR	RSSoFR	SSoFR	SoFR
0	normal	150	0.1386 (0.0567)	0.1267 (0.0531)	0.1377 (0.0549)	0.1727 (0.0776)	0.1695 (0.0755)	0.1723 (0.0773)
		300	0.0788 (0.0346)	0.0721 (0.0316)	0.0786 (0.0342)	0.1418 (0.0455)	0.1403 (0.0451)	0.1417 (0.0454)
		600	0.0419 (0.0168)	0.0383 (0.0153)	0.0418 (0.0168)	0.1279 (0.0246)	0.1273 (0.0243)	0.1279 (0.0247)
0	t(1)	150	0.2676 (0.1497)	30034 (565196)	0.2632 (0.1431)	0.2080 (0.0932)	431 (6064)	0.2084 (0.0934)
		300	0.1524 (0.0752)	12790 (192370)	0.1513 (0.0745)	0.1569 (0.0523)	1356 (17250)	0.1568 (0.0521)
		600	0.0755 (0.0336)	4633 (71726)	0.0754 (0.0337)	0.1350 (0.0267)	43476 (593702)	0.1349 (0.0266)
0	t(3)	150	0.1841 (0.0885)	0.3278 (0.2543)	0.1832 (0.0875)	0.1923 (0.0859)	0.2204 (0.1064)	0.1917 (0.0859)
		300	0.0998 (0.0456)	0.1845 (0.1220)	0.0996 (0.0452)	0.1499 (0.0500)	0.1648 (0.0604)	0.1500 (0.0501)
		600	0.0535 (0.0218)	0.0970 (0.0700)	0.0534 (0.0218)	0.1290 (0.0250)	0.1369 (0.0327)	0.1290 (0.0250)
0.5	normal	150	0.1366 (0.0590)	0.1231 (0.0520)	0.1984 (0.0904)	0.1742 (0.0787)	0.1719 (0.0786)	0.1946 (0.0882)
		300	0.0761 (0.0321)	0.0698 (0.0281)	0.1184 (0.0544)	0.1410 (0.0465)	0.1394 (0.0461)	0.1540 (0.0501)
		600	0.0430 (0.0171)	0.0395 (0.0158)	0.0687 (0.0268)	0.1271 (0.0255)	0.1265 (0.0255)	0.1385 (0.0288)
0.5	t(1)	150	0.2822 (0.1471)	2827 (25820)	1.1746 (0.9887)	0.2104 (0.0960)	3744 (49109)	0.4115 (0.3345)
		300	0.1543 (0.0820)	3862 (49921)	0.6101 (0.3746)	0.1596 (0.0525)	22641 (488421)	0.2519 (0.1293)
		600	0.0740 (0.0335)	9897 (192399)	0.3027 (0.1699)	0.1345 (0.0261)	134 (1001)	0.1836 (0.0591)
0.5	t(3)	150	0.1940 (0.0929)	0.3148 (0.1809)	0.3114 (0.1594)	0.1891 (0.0861)	0.2216 (0.1347)	0.2164 (0.0981)
		300	0.1042 (0.0470)	0.1808 (0.1095)	0.1722 (0.0826)	0.1483 (0.0472)	0.1630 (0.0576)	0.1676 (0.0560)
		600	0.0531 (0.0224)	0.0926 (0.0591)	0.0908 (0.0388)	0.1289 (0.0241)	0.1376 (0.0329)	0.1420 (0.0283)
0.8	normal	150	0.1432 (0.0636)	0.1298 (0.0581)	0.6378 (0.3425)	0.1759 (0.0803)	0.1734 (0.0799)	0.3524 (0.1792)
		300	0.0763 (0.0296)	0.0697 (0.0264)	0.4035 (0.1882)	0.1425 (0.0455)	0.1407 (0.0450)	0.2884 (0.1148)
		600	0.0416 (0.0168)	0.0377 (0.0149)	0.2581 (0.0978)	0.1288 (0.0248)	0.1282 (0.0246)	0.2494 (0.0652)
0.8	t(1)	150	0.2877 (0.2352)	1071 (7593)	10.7320 (24.6818)	0.2029 (0.0916)	737 (10172)	3.0490 (9.5959)
		300	0.1518 (0.0730)	2446 (19969)	4.4654 (4.5632)	0.1540 (0.0508)	932 (12959)	1.0579 (1.0765)
		600	0.0782 (0.0394)	7359 (101428)	1.9833 (1.3628)	0.1343 (0.0267)	628 (7479)	0.6228 (0.5507)
0.8	t(3)	150	0.1819 (0.0920)	0.3356 (0.5530)	0.8924 (0.5139)	0.1877 (0.0833)	0.2223 (0.1090)	0.4125 (0.2617)
		300	0.0989 (0.0472)	0.1779 (0.1357)	0.5534 (0.2888)	0.1470 (0.0505)	0.1623 (0.0561)	0.3185 (0.1527)
		600	0.0524 (0.0216)	0.0899 (0.0580)	0.3368 (0.1328)	0.1294 (0.0256)	0.1362 (0.0298)	0.2671 (0.0917)

better than that of the SSoFR. And when ϵ_i follows $N(0,1)$ or $t(3)$, $\hat{\rho}$ of the two models behave similarly.

(2) In Table 2 compare $\hat{\beta}(t)$ of the RSSoFR and the SoFR. In case $\rho = 0$, RSSoFR's $\hat{\beta}(t)$ approximates the SoFR's. But in case $\rho \neq 0$, $\hat{\beta}(t)$ of the RSSoFR has much smaller IMSE than that of the SoFR has. Moreover, as ρ increases, the differences

Table 3: The estimates and its standard deviations (in brackets) of ν.

			$\gamma = 1.1$			$\gamma = 2$		
ρ	ϵ_i	n	RSSoFR	SSoFR	SoFR	RSSoFR	SSoFR	SoFR
0	normal	150	4.0080 (0.0892)	–	4.0080 (0.1549)	4.0260 (0.1593)	–	4.0420 (0.2105)
		300	4.0040 (0.0632)	–	4.0040 (0.0632)	4.0000 (0.0000)	–	4.0000 (0.0000)
		600	4.0000 (0.0000)	–	4.0000 (0.0000)	4.0000 (0.0000)	–	4.0000 (0.0000)
0	t(1)	150	1.7960 (0.4635)	–	1.8260 (0.4474)	1.8560 (0.4557)	–	1.8820 (0.4477)
		300	1.8720 (0.3462)	–	1.8880 (0.3281)	1.8880 (0.3401)	–	1.8920 (0.3355)
		600	1.9560 (0.2053)	–	1.9620 (0.1914)	1.9620 (0.1914)	–	1.9660 (0.1814)
0	t(3)	150	3.2300 (0.4213)	–	3.2360 (0.4250)	3.1660 (0.3725)	–	3.2040 (0.4034)
		300	3.1380 (0.3452)	–	3.1440 (0.3514)	3.1260 (0.3322)	–	3.1580 (0.3651)
		600	3.0880 (0.2836)	–	3.1080 (0.3107)	3.0640 (0.2450)	–	3.0960 (0.2949)
0.5	normal	150	4.0060 (0.1342)	–	4.0320 (0.1872)	4.0180 (0.1331)	–	4.0500 (0.2601)
		300	4.0020 (0.0447)	–	4.0000 (0.0000)	4.0000 (0.0000)	–	4.0000 (0.0000)
		600	4.0000 (0.0000)	–	4.0000 (0.0000)	4.0000 (0.0000)	–	4.0000 (0.0000)
0.5	t(1)	150	1.7800 (0.4817)	–	1.9580 (0.6520)	1.8460 (0.4457)	–	1.9640 (0.6321)
		300	1.9100 (0.2934)	–	1.9340 (0.5120)	1.8840 (0.3388)	–	1.8900 (0.5200)
		600	1.9540 (0.2097)	–	1.8900 (0.3978)	1.9700 (0.1708)	–	1.8600 (0.3959)
0.5	t(3)	150	3.2520 (0.4346)	–	3.4940 (0.5005)	3.1960 (0.3974)	–	3.4460 (0.4976)
		300	3.1360 (0.3431)	–	3.5140 (0.5003)	3.0900 (0.2865)	–	3.3760 (0.4849)
		600	3.0680 (0.2520)	–	3.4600 (0.4989)	3.0700 (0.2554)	–	3.3920 (0.4887)
0.8	normal	150	4.0020 (0.0775)	–	4.0540 (0.2349)	4.0120 (0.1260)	–	4.0520 (0.2478)
		300	4.0000 (0.0000)	–	4.0060 (0.0773)	4.0020 (0.0447)	–	4.0120 (0.1090)
		600	4.0000 (0.0000)	–	4.0000 (0.0000)	4.0000 (0.0000)	–	4.0000 (0.0000)
0.8	t(1)	150	1.7960 (0.4458)	–	2.2380 (0.7891)	1.8460 (0.4457)	–	2.2240 (0.7791)
		300	1.8880 (0.3157)	–	2.2120 (0.6928)	1.9060 (0.2989)	–	2.1300 (0.7114)
		600	1.9540 (0.2097)	–	2.0380 (0.6398)	1.9660 (0.1814)	–	2.0280 (0.6261)
0.8	t(3)	150	3.2240 (0.4173)	–	3.8120 (0.4012)	3.1760 (0.3812)	–	3.7780 (0.4160)
		300	3.1140 (0.3181)	–	3.8760 (0.3299)	3.0800 (0.2716)	–	3.8520 (0.3555)
		600	3.0820 (0.2746)	–	3.9580 (0.2008)	3.0580 (0.2340)	–	3.8720 (0.3344)

in the IMSE between the RSSoFR and the SoFR also raise. Thus we conclude that overall our RSSoFR behaves better than the SoFR.

(3) In Table 2 observe $\hat{\beta}(t)$ of the RSSoFR and the SSoFR. We can find under normal distributions, our proposed method performs basically consistent with the SSoFR. Whereas under t-distributions with thick tails, the RSSoFR produces better results

for $\beta(t)$ than the SSoFR. Thus the RSSoFR is more robust than the SSoFR.

(4) From Table 2, we can also find the IMSE of $\hat{\beta}(t)$ of the RSSoFR decreases with the increasing sample size. Besides, $\beta(t)$ is more accurately estimated given $\gamma = 1.1$ than $\gamma = 2$ when the other simulation parameters are held equal, as the case in Hall and Horowitz (2007). The estimation results for ν are presented in Table 3.

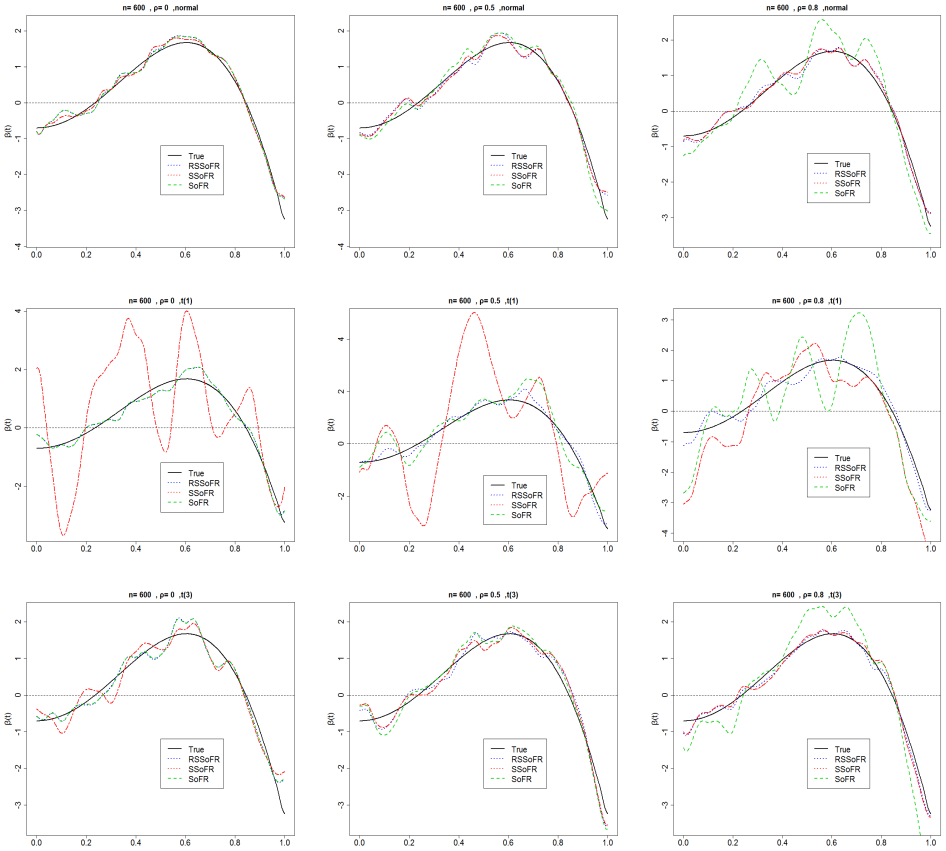

Figure 2: The estimated $\beta(t)$ vs. the true $\beta(t)$ when $n = 600$, $\gamma = 1.1$, $\rho = 0, 0.5, 0.8$ and $\epsilon_i \sim N(0,1), t(1), t(3)$ respectively.

To show the performances of $\hat{\beta}(t)$ intuitively, for each case we select one result from 500 repetitions when $n = 600$, $\gamma = 1.1$, $\rho = 0, 0.5, 0.8$ and $\epsilon_i \sim N(0,1), t(1), t(3)$. Figure 2 displays the estimated $\hat{\beta}(t)$ vs. the true value of $\beta(t)$ for the RSSoFR, the SSoFR and the SoFR. From the figure, we can draw the similar conclusions for $\beta(t)$ as those derived from Table 2. In addition, we have also implemented numerical experiments with the FPLS basis used for the functional regressor expansion, which leads to the

similar conclusions as those from results with the FPC basis in Table 1-3. For sake of space, we do not report the corresponding results and will illustrate the FPLS basis in the real data analysis. To sum up, our proposed model performs better than the SoFR and is more robust than the SSoFR, which provides a competitive alternative for the existing methods in FDA.

5 REAL DATA ANALYSIS

In this section, we revisit the weather dataset presented in Section 1 to assess the application of the RSSoFR. Specifically, we add records that correspond to the weather data for 2008 derived from the China Meteorological Yearbook. In this dataset, records from 2005 to 2007 serve as training set while those from 2008 serve as test set. Note that each of the involved 34 Chinese cities covers a large area with the average land size around 16147 km^2, and each observation represents the global picture of the monthly mean temperature and monthly total precipitation across the whole city instead of the particular point, which integrates records from tens or even hundreds of meteorological stations in that city area rather than only one weather station (that is commonly treated as point-referenced data). From this perspective, here the dataset can be viewed as areal data instead of point-reference data. We compare the RSSoFR with the SoFR by their fitting and prediction results.

During preprocessing, we smooth the discrete temperature records $x_i(t_j), i = 1, 2, \ldots, 34,\ t_j = \frac{1}{12}, \frac{1}{6}, \ldots, \frac{11}{12}, 1$ (each j represents a month) using the Epanechnikov Kernel to get $x_i(t)$. Figure 3 (left) presents the functional predictors.

Let the response y_i be the logarithm of the mean annual total precipitation for the ith city. We build the RSSoFR model as

$$y_i = \rho \sum_{i \neq i'} w_{ii'} y_{i'} + \int_0^1 x_i(t)\beta(t)dt + \epsilon_i, \tag{5.1}$$

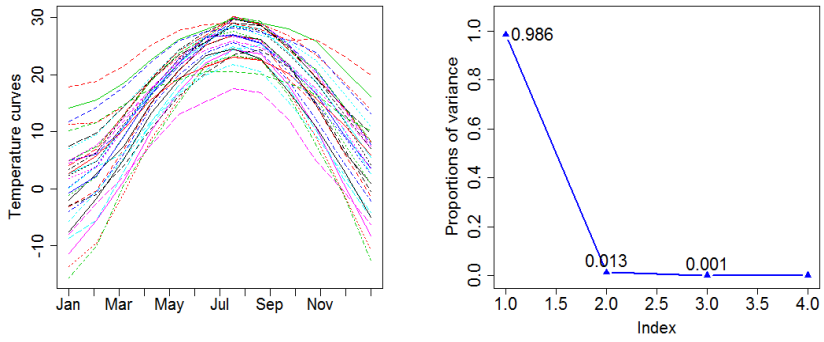

Figure 3: The temperature curves of 34 cities in training set (left), the eigenvalues of the sample covariance function (right).

where $w_{ii'}$ is the weight between city i and i'. We also build the SoFR as

$$y_i = \int_0^1 x_i(t)\beta(t)dt + \epsilon_i \tag{5.2}$$

to enable comparison with the RSSoFR.

The spatial weight matrix $\boldsymbol{W} = (w_{ii'})_{n\times n}$ is formed by the reciprocal of the distance $d_{ii'}$ between centers of two cities, i.e.,

$$w_{ii'} = \frac{1}{d_{ii'}},$$

where $d_{ii'}$ is computed using the haversine formula based on longitudes and latitudes of cities' centers. Besides, we considered two influence factors that may effect model (5.1)'s performances during the process of constructing $\boldsymbol{W}$. The first is the threshold distance d_0 (in kilometers). We know if city i is very far from city i', the spatial dependence between them will be very small. Thus we set $w_{ii'} = 0$ if $d_{ii'} > d_0$. The second is the number of the nearest neighbors k. For any city i, there exists competitions among weights $w_{ii'}, i' = 1, \cdots, 34$. Suppose city i has many close neighbors $i_1, \cdots, i_{n_0}$, $\max\{d_{ii_1}, \cdots, d_{ii_{n_0}}\} < d_0$ and n_0 is relatively great, then it is very important to determine the value of k. As the smaller k is, the greater each $w_{ii'}$ becomes, i.e., the stronger influence each neighbor i' has on city i.

Table 4: The values of the Moran's I statistic under different values of k and d_0.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
d_0 (km)	613.5	626.0	626.0	626.0	613.5	625.4	625.4	613.5
Moran's I	0.746	0.801	0.732	0.753	0.754	0.747	0.741	0.743

We choose the $\boldsymbol{W}$ under which the Moran's I statistic takes the greatest value. Set $k = \{1, 2, 3, 4, 5, 6, 7, 8\}$. And for each k, we search a d_0 where the value of the Moran's I statistic is the greatest. Table 4 displays the results for different values of k. It can be seen $\boldsymbol{W}$ with $k = 2$ and $d_0 = 626.0$ is the best option. We also found this $\boldsymbol{W}$ performs better than the spatial weight matrix formed by the negative exponential of distance (2.2). Figure 4 presents the locations of 34 major cities on map of China based on the longitudes and the latitudes of the cities' centers. Because Urumchi and Lhasa are located far from the other cities (all the distances are greater than 1250 km), we remove their records from the weather dataset. Note that $\boldsymbol{W}$ is row-normalized after construction.

Figure 4: The locations of 34 major cities on map of China based on the longitudes and the latitudes of the cities' centers.

Figure 3 (right) shows the eigenvalues of the sample covariance function. The eigenvalues clearly decay quickly, and the first eigenvalue accounts for 98.6% of the total variance. Therefore, we consider using one component in the RSSoFR. As the FPLS basis behaves better than the FPC basis in this dataset, we employ the FPLS basis

for both the RSSoFR and the SoFR to make a comparison. The fitted and predicted results of the RSSoFR and the SoFR are summarized in Table 5. Figure 5 shows the related consequences.

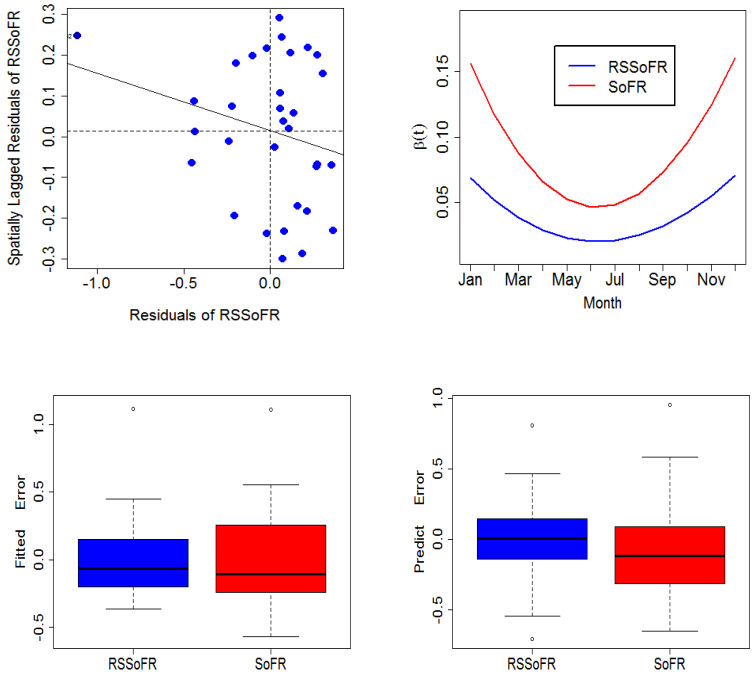

Figure 5: The Moran Scatter plot of the residuals of the RSSoFR (top-left), the estimated $\beta(t)$ of the RSSoFR and the SoFR (top-right), the fitted error of the RSSoFR and the SoFR (bottom-left), and the predicted error of the RSSoFR and the SoFR (bottom-right), respectively.

Table 5: The fitting and prediction results of the RSSoFR and the SoFR.

models	$\hat{\rho}$	Moran's I statistic(residuals)	MSE(fitted error)	MSE(prediction error)	$\hat{\nu}$
SoFR	—	0.487	0.128	0.114	—
RSSoFR	0.675	-0.14	0.088	0.076	15

Firstly, we discuss the fitting results of the RSSoFR and the SoFR. Recall from Section 1 that significant spatial correlation presents among the residuals of the SoFR. We can see from Table 5 that the value of the Moran's I statistic of the residuals of the RSSoFR

is small (−0.14), which suggests a majority of spatial correlation in the residuals has been removed. Figure 5 (top-left) shows the Moran Scatterplot of the residuals of the RSSoFR. As for the fitted values, the MSE of the RSSoFR is nearly half of that of the MSE of the SoFR, where $MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$. Thus, the RSSoFR fits the mean total annual precipitation better than the SoFR. Figure 5 (bottom-left) presents boxplots of the residuals of the RSSoFR and the SoFR. Regarding the estimated parameters, $\hat{\rho}$ is 0.705, which is also very significant with its p-value smaller than 0.001. The slope functions of the RSSoFR and the SoFR are provided in Figure 5 (top-right). We can find the two estimated curves have similar shapes, and $\hat{\beta}(t)$ of the RSSoFR is smoother than that of the SoFR. We conclude that the precipitation is much more strongly influenced by the temperatures during spring and winter than in the other seasons. Then we focus on the prediction results. The predicted error of the RSSoFR is 0.076 while that of the SoFR is 0.114, which indicates the RSSoFR has better prediction performance than the SoFR. Figure 5 (bottom-right) presents boxplots of the prediction error under the RSSoFR and the SoFR.

6 CONCLUSION AND DISCUSSION

The SoFR is popular in studies of links between a scalar response and functional predictors. However, the existing SoFR cannot address the dependencies in a cross-sectional spatial scenario. We propose a robust spatial autoregressive scalar-on-function regression that incorporates a spatial lagged term into the SoFR to accommodate the spatial dependence and allows for thick tailed noise term. An estimation method based on basis expansion and EM algorithm is developed to obtain the estimators of the spatial autoregressive parameter and the slope function. Specifically, the FPC basis and the FPLS basis can be applied and the spatial cross-validation method is introduced to choose the truncation parameter. Our simulation study demonstrates the consistency of the proposed estimators. In particular, the new model performs better than the SoFR when the spatial correlation is present, and the SSoFR when the error

term has thick tails. An examination of a real dataset illustrates the superiority of the RSSoFR over the SoFR. In short, the robust spatial autoregressive scalar-on-function regression model with t-distribution presented in this paper constitutes a practical statistical tool for modelling the spatial dependent data with functional covariates and scalar response that complements the widely popular spatial autoregressive scalar-on-function regression with normality assumption and may help to further understanding in many fields of applied research.

We discussed the RSSoFR under the assumption that only one functional predictor is involved. The RSSoFR with multiple functional predictors also deserve attention. Based on the estimation method introduced in Section 3, the methods presented here can be easily generalized to the RSSoFR with multiple functional predictors.

One referee raised concerns about the model uncertainty. Honestly, apart from the estimation method, the problem of statistical inference is also of great interest as it provides an overall assessment of the association of the functional covariates with the responses ($\beta(t)$, ρ), as well as the whole model; however, it remains challenging due to the infinite dimensionality of the functional covariates. To overcome this issue, a natural strategy is to reduce the dimension. With representing the functional covariates and the coefficient function by linear combinations of a set of basis functions, the testing problem for $\beta(t)$ reduces to the hypothesis testing under a classical linear model. Along with this line, there is a plethora of literature that develops statistical methods. For example, Su et al. (2017) proposed a Wald-type test with varying thresholds in selecting the number of PCs for the functional covariates; Garca-Portugus et al. (2014) introduced the projected Cramér-von Mises (PcVM) test-a testing method which is derived by using random projection, and whose null distribution is approximated by bootstrap. We refer the reader to Tekbudak et al. (2019) for an extensive comparison of testing methods in scalar-on-function regression.

However, in the current paper we adopt the EM algorithm to implement the estimation procedure, which cannot obtain the corresponding standard errors of the proposed estimators. A common practice is to compute the standard errors of the proposed estima-

tors via the bootstrap. There are some existing methods for estimating the standard error of an estimator obtained from the EM algorithm. For example, Louis (1982) developed a formula for computing the observed information matrix in terms of the complete and missing information matrices, while the calculation of the missing information involves the conditional expectation of the outer product of the complete-data score vector; Jamshidian and Jennrich (2000) proposed to use numerical differentiation to yield the Hessian matrix. After obtaining the standard errors, one may follow the analogous methods to perform the statistical inference for the proposed model. Investigations along this direction may be of interest and merit further research but are beyond the scope of the paper.

References

Aguilera, A. M., Escabias, M., Preda, C., and Saporta, G. (2010). Using basis expansions for estimating functional PLS regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems*, **104**(2), 289 – 305.

Aguilera-Morillo, M. C., Durbán, M., and Aguilera, A. M. (2017). Prediction of functional data with spatial dependence: a penalized approach. *Stochastic Environmental Research and Risk Assessment*, **31**(1), 7–22.

Ait-Saïdi, A., Ferraty, F., Kassa, R., and Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics*, **42**(6), 475–494.

Anselin, L. (1998). *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, **27**(3), 247–267.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica*, **59**(4), 953–965.

Case, A. C., Rosen, H. S., and Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, **52**(3), 285 – 307.

Cliff, A. and Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, **4**(3), 267–284.

Crainiceanu, C. M., Staicu, A.-M., and Di, C.-Z. (2009). Generalized Multilevel Functional Regression. *Journal of the American Statistical Association*, **104**(488), 1550–1561.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, **37**(1), 35–72.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**(1), 136 – 154.

De Jong, S. (1993). PLS fits closer than PCR. *Journal of Chemometrics*, **7**(6), 551–557.

Decey, M. (1968). A review of measure of contuity for two and k-color maps. *In Spatial Analysis: a Reader in Statistical Geography*, pages 479–495.

Delaigle, A. and Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *Ann. Statist.*, **40**(1), 322–352.

Fang, Y., Yuejiao, F., and Lee, T. C. M. (2011). Functional mixture regression. *Biostatistics*, **12**(2), 341–353.

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

Garca-Portugus, E., Gonzlez-Manteiga, W., and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, **23**(3), 761–778.

Giraldo, R., Delicado, P., and Mateu, J. (2017). Spatial prediction of a scalar variable based on data of a functional random field. *Comunicaciones en Estadística*, **10**(2), 315–344.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society*, **61**(3), 453–469.

Goulard, M., Laurent, T., and Thomas-Agnan, C. (2017). About predictions in spatial autoregressive models: optimal and almost optimal strategies. *Spatial Economic Analysis*, **12**(2-3), 304–325.

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

Hastie, T. and Mallows, C. (1993). [a statistical view of some chemometrics regression tools]: Discussion. *Technometrics*, **35**(2), 140–143.

Isard, W. et al. (1970). *General theory: social, political, economic and regional.* Cambridge, Mass./London: Mass. Inst. Technol.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **64**(3), 411–432.

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *Journal of the American Statistical Association*, **100**(470), 565–576.

James, G., Hastie, T., and Sugar, C. (2000). Principal component models for sparse functional data. *Biometrika*, **87**(3), 587–602.

Jamshidian, M. and Jennrich, R. I. (2000). Standard errors for em estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(2), 257–270.

Kelejian, H. H. and Prucha, I. R. (2001). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, **40** (2), 509–533.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, **72**(6), 1899–1925.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, **137**(2), 489–514. ISSN 0304-4076.

Lesage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.*, **38** (6), 3321–3351.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44** (2), 226–233.

Marx, B. D. and Eilers, P. H. C. (1999). Generalized linear regression on sampled signals and curves: A P-spline approach. *Technometrics*, **41**(1), 1–13.

Menafoglio, A. and Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: a review of object oriented spatial statistics. *European journal of operational research*, **258**(2), 401–410.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, **2**(1), 321–359.

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

Nerini, D., Monestiez, P., and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, **101**(2), 409 – 418.

Olubusoye, O. E., Korter, G. O., and Salisu, A. A. (2016). Modelling road traffic crashes using spatial autoregressive model with additional endogenous variable. *Statistics in Transition. New Series*, **17**(4), 659–670.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, **70**(349), 120–126.

Peel, D. and McLachlan, G. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, **10**, 339–348.

Pineda-Ríos, W., Giraldo, R., and Porcu, E. (2019). Functional SAR models: With application to spatial econometrics. *Spatial statistics*, **29**, 145–159.

Preda, C. and Saporta, G. (2005). PLS regression on a stochastic process. *Computational Statistics & Data Analysis*, **48**(1), 149 – 158.

Preda, C. and Saporta, G. (2007). *PCR and PLS for Clusterwise Regression on Functional Data*, pages 589–598. Springer Berlin Heidelberg, Berlin, Heidelberg.

Preda, C., Saporta, G., and Lévéder, C. (2007). PLS classification of functional data. *Computational Statistics*, **22**(2), 223–235.

Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, **184**(2), 209–232.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society*, **53**(3), 539–572.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review*, **85**(2), 228–249.

Schabenberger, O. and Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. CRC press.

Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference*, **139**(10), 3405 – 3418.

Su, Y.-R., Di, C.-Z., and Hsu, L. (2017). Hypothesis testing in functional linear models. *Biometrics*, **73**(2), 551–561.

Tekbudak, M. Y., Alfaro-Córdoba, M., Maity, A., and Staicu, A.-M. (2019). A comparison of testing methods in scalar-on-function regression. *AStA Advances in Statistical Analysis*, **103**(3), 411–436.

Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment. *The Review of Economic Studies*, **68**(2), 261–295.

Wang, H., Gu, J., Wang, S., and Saporta, G. (2019). Spatial partial least squares autoregression: Algorithm and applications. *Chemometrics and Intelligent Laboratory Systems*, **184**, 123 – 131.

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

Zhang, J., Clayton, M. K., and Townsend, P. A. (2011). Functional concurrent linear regression model for spatial images. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**(1), 105–130.

Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A., and Morris, J. S. (2016). Functional CAR models for large spatially correlated functional datasets. *Journal of the American Statistical Association*, **111**(514), 772–786.