Intrinsic Functional Partially Linear Poisson Regression Model for Count Data

Jiaqi Xu [1,†] , Yu Lu [1,†], Yuanshen Su [1,†], Tao Liu [2,*] , Yunfei Qi [3] and Wu Xie [3]

[1] Sydney Smart Technology College, Northeastern University, Shenyang 110004, China; 202219268@stu.neu.edu.cn (J.X.); 202219298@stu.neu.edu.cn (Y.L.); 202219294@stu.neu.edu.cn (Y.S.)

[2] School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

[3] Eighth Geological Brigade of Hebei Bureau of Geology and Mineral Resources Exploration (Hebei Center of Marine Geological Resources Survey), Qinhuangdao 066000, China; geo_yfqi@163.com (Y.Q.); 15232374023@163.com (W.X.)

* Correspondence: liutao@neuq.edu.cn or math.taoliu@gmail.com

† These authors contributed equally to this work.

Abstract: Poisson regression is a statistical method specifically designed for analyzing count data. Considering the case where the functional and vector-valued covariates exhibit a linear relationship with the log-transformed Poisson mean, while the covariates in complex domains act as nonlinear random effects, an intrinsic functional partially linear Poisson regression model is proposed. This model flexibly integrates predictors from different spaces, including functional covariates, vector-valued covariates, and other non-Euclidean covariates taking values in complex domains. A truncation scheme is applied to approximate the functional covariates, and the random effects related to non-Euclidean covariates are modeled based on the reproducing kernel method. A quasi-Newton iterative algorithm is employed to optimize the parameters of the proposed model. Furthermore, to capture the intrinsic geometric structure of the covariates in complex domains, the heat kernel is employed as the kernel function, estimated via Brownian motion simulations. Both simulation studies and real data analysis demonstrate that the proposed method offers significant advantages over the classical Poisson regression model.

1. Introduction

Count data are widely encountered across various fields, including biomedicine, industry, agriculture, and economics. Poisson regression is a common and effective statistical model for handling such data, particularly when modeling the frequency or count of events. It is used to explore the underlying relationships between count data, as the response variable, and a set of covariates. The application of Poisson regression has shown significant results across multiple domains. For instance, Getaneh et al. [1] utilized Poisson regression to investigate the significant determinants of neonatal mortality rates in Ethiopia. Loukas et al. [2] employed the model to analyze the number of goals scored in football matches. Nzuma and Mzera [3] applied Poisson regression to assess the control measures chosen by smallholder farmers in Kilifi County, Kenya, in response to aflatoxin contamination. Additionally, Sakane et al. [4] found that Poisson regression analysis revealed the relationship between the ability of type 1 diabetes patients to manage hypoglycemia and the occurrence of severe hypoglycemic episodes. These practical applications demonstrate that Poisson regression effectively captures the discreteness of data by modeling the probability distribution of event occurrences, leading to more accurate predictions and interpretations. This makes Poisson regression an extremely effective tool for analyzing count data.

With the rapid advancement of data collection and storage technologies, traditional Poisson regression models may no longer adequately address emerging data scenarios [5]. For instance, Weaver et al. [6] uncovered potential limitations of Poisson regression in their analysis of hospitalization data, highlighting that real-world datasets frequently fail to meet the underlying assumptions of the model. In the presence of multicollinearity among covariates, maximum likelihood estimation (MLE) can become unreliable. To mitigate this issue, Amin et al. [7] proposed the Poisson James–Stein estimator, which aims to reduce the inflated variance and standard errors associated with MLE under conditions of multicollinearity. Abdelwahab et al. [8] proposed the two-parameter Liu estimator that can effectively address multicollinearity issues. In contexts where the dimensionality of covariates is high, Fei Jiang and Yanyuan Ma [9] developed an estimator for Poisson regression that accommodates high-dimensional characteristics and additive measurement errors. Under the assumption of robust variance, Hagiwara et al. introduced a goodness-of-fit test specifically designed for modified Poisson regression. Additionally, Qiang Fu et al. [10] proposed an enhanced Poisson regression estimator tailored for grouped and right-censored count data, deriving their asymptotic properties. Furthermore, Wang and Yu [11] presented a nonparametric Poisson regression model informed by Euclidean covariates, along with corresponding estimation tools. Minggen Lu and Dana Loomis [12] introduced a spline-based partially linear Poisson single-index model to investigate the relationship between multidimensional air pollution exposure and mortality, utilizing B-splines to approximate the unknown regression function.

The methods mentioned above were proposed in the context of vector-valued covariates. However, with the diversification of data types, covariates may originate from different sample spaces, such as function spaces [13], Riemannian manifolds [14], Euclidean spaces with complex spatial constraints [15], and certain metric spaces [16]. For instance, the population density of a particular species in different regions is influenced not only by local factors such as forest coverage and lake area but also by time-varying environmental factors. Due to their rapid fluctuations, these environmental factors can be regarded as functional data. Moreover, population density is also related to the geographical coordinates of the regions, and these spherical coordinates are clearly not Euclidean data. Based on this, it is of significant importance and value to develop Poisson regression models that can integrate covariate information from diverse spatial sources [17].

For functional data, Wang et al. [13] first proposed the concept of functional data analysis (FDA), making a pioneering contribution to research in this field. Subsequently, research on FDA has mainly focused on the technique of functional principal component analysis (FPCA). For data over complex domains, Wood et al. [18] pointed out that traditional smoothing or modeling methods that do not take into account the intrinsic geometry of the space, especially boundary constraints, can yield poor results when used on data in complex domains. It is therefore important to take into account the intrinsic geometry of the region and its complex boundaries. In recent years, Gaussian processes (GPs) have been widely used for modeling unknown functions or surfaces, and it has been proposed to regress unknown manifolds with Intrinsic Gaussian Processes (IGPs) [19,20]. This clearly facilitates the construction of high-dimensional data in low-dimensional manifolds and fully respects the non-Euclidean properties of the data structure.

When the canonical parameter exhibits linearity, utilizing a generalized linear model with a Poisson likelihood is advantageous for analyzing count data. However, in many scenarios, this linear assumption may not hold, necessitating the appropriate handling of nonlinear dependencies. Nonparametric Poisson regression emerges as a preferred method for characterizing nonlinear relationships, offering increased flexibility in modeling count data. A critical consideration arises when predictor variables originate from disparate spaces, which can complicate the application of nonparametric Poisson regression. For instance, in spatial data analysis [21,22], it is often necessary to model predictive variables belonging to Euclidean space while also parameterizing non-specific random spatial effects. Spatial data, such as longitude and latitude, are distributed across irregular spatial domains

characterized by complex boundaries. Notably, pairs of locations that are close in terms of Euclidean distance may be intrinsically far apart if separated by geographical barriers. Some studies treat spatial effects as identical and independent Gaussian distributions. While this assumption is acceptable for unstructured spatial effects that can be approximated by independent random effects, it often fails to hold in practice, making it essential to address spatial dependence appropriately. Several nonparametric estimators, such as spline techniques [23–27], provide commendable methods for approximating functions on complex constrained domains. However, spline methodologies come with certain limitations. Many spline approaches rely on predetermined basis functions, such as polynomials or cubic splines. These methods can also face challenges related to dimensional complexities and increased computational demands, particularly in high-dimensional scenarios. Additionally, splines may struggle to effectively accommodate predictive variables from distinct spaces simultaneously. While nonparametric Poisson regression offers greater flexibility, it often suffers from poor convergence rates and a lack of interpretability. In classical regression analysis, semi-parametric regression models have garnered significant attention from the scientific community due to their advantages over linear and nonparametric models. These models combine the interpretability of linear slope functions with the flexibility of nonparametric link functions, making them widely applicable across various scientific fields [28–34].

The objective of this paper is to develop a Poisson regression model capable of handling predictor variables that include functional covariates, vector-valued covariates, and covariates defined in non-Euclidean spaces, with the response variable being count data. To balance the flexibility and interpretability of the regression model, we assume a linear relationship between the functional covariates and vector-valued covariates and the log-transformed Poisson mean response, and we posit a nonlinear relationship between the non-Euclidean covariates and the log-transformed Poisson mean response, represented through individual random effects. For the functional covariates, we employ the Karhunen–Loève expansion for approximation. In the case of the non-Euclidean covariates, we assume that the corresponding random effects function resides in a reproducing kernel Hilbert space, which can be effectively modeled using the reproducing kernel representation theorem. For the selection of non-Euclidean kernel functions, we utilize heat kernels based on the heat equation, which can be estimated through the transition probabilities of Brownian motion. The Karhunen–Loève expansion and heat kernel enable our model to handle complex functional and spatial data. The KL expansion reduces functional data into main components, similar to principal component analysis, making it easier to manage time-dependent covariates. The heat kernel captures the spatial dependencies in non-Euclidean domains by modeling local geometric relationships, which is essential for spatially structured data. The parameter estimation for the proposed method will be optimized using a Newton iteration algorithm. Furthermore, we extend the theoretical framework of our method to encompass the two-component mixture Poisson distribution. Analysis of both simulated and real data indicates that the proposed method demonstrates significant advantages over classical Poisson regression models.

The remainder of this article is structured as follows. Section 2 introduces the intrinsic functional partially linear Poisson regression model, providing details on the model formulation and parameter estimation procedures. Section 3 presents simulation studies and empirical data analyses to evaluate the efficacy of the proposed methodology. Section 4 offers a comprehensive summary of the key findings.

2. Methodology

2.1. Model Setting

In classic Poisson regression, it is assumed that the response variable $Y = y \in \{0, 1, \cdots\}$, and the number of occurrences of an event, given the Euclidean predictor variable $X = \boldsymbol{x} \in \mathbb{R}^D$, follows a Poisson distribution

$$p(Y = y \mid X = \boldsymbol{x}) = \frac{\lambda(\boldsymbol{x})^y}{y!} e^{-\lambda(\boldsymbol{x})}, \tag{1}$$

where $\ln(\lambda(\boldsymbol{x})) = \boldsymbol{x}^\top \boldsymbol{\beta}$ and $\boldsymbol{\beta}$ is a D-dimensional vector of canonical parameters. It is the most natural idea that links the linear predictor $\boldsymbol{x}^\top \boldsymbol{\beta}$ and the parameter $\lambda(\boldsymbol{x})$ with the logarithm. Then, the MLEs of $\boldsymbol{\beta}$ can be found numerically using the quasi-Newton method or iterative weighted least squares method.

With the increasing diversity of data types, covariates are considered from various feature spaces, including square-integrable function spaces, Euclidean spaces, and non-Euclidean spaces. Functional and vector-valued covariates are assumed to exhibit a linear relationship with $\ln(\lambda)$, while covariates from non-Euclidean spaces are treated as random effects with a nonlinear relationship. Under these assumptions, the intrinsic functional partially linear Poisson regression model is defined as follows:

$$\ln(\lambda) = \sum_{j=1}^{J} \int_{t \in \mathcal{T}} X_j(t) \beta_j(t) dt + \boldsymbol{Z}^\top \boldsymbol{\eta} + \delta(\boldsymbol{s}) \tag{2}$$

where $\boldsymbol{X} := (X_1, X_2, \cdots, X_J)$ are multivariate functional covariates, $\mathcal{T}$ is a compact set, and $\boldsymbol{\beta} := \{\beta_1, \beta_2, \cdots, \beta_J\}$ are slope functions. $\boldsymbol{Z} \in \mathbb{R}^D$ are vector-valued covariates, and $\boldsymbol{\eta}$ is their corresponding coefficient vector. $\delta(\boldsymbol{s})$ is a nonlinear random effect related to non-Euclidean covariates $\boldsymbol{s} \in \mathcal{M}$, and $\mathcal{M}$ is a complex domain (such as a sphere, an ellipse, or hyperbolic space). In this paper, we assume the complex domain is a non-Euclidean manifold.

In the above assumption, each covariate type distinctly impacts the parameter λ, leveraging either linear or nonlinear interactions based on its characteristics. The functional covariates $X_j(t)$, which capture time-dependent data over a domain $\mathcal{T}$, influence λ linearly through the term $\int_{\mathcal{T}} X_j(t) \beta_j(t) \, dt$. This structure effectively captures gradual, continuous variations. The vector-valued covariates Z, residing in Euclidean space $\mathbb{R}^D$, impact λ via a simpler linear association $Z^\top \eta$, allowing the model to include fixed or categorical effects efficiently. The non-Euclidean covariates $s \in \mathcal{M}$, from non-Euclidean spaces (e.g., a manifold), contribute nonlinearly through a random effect $\delta(s)$, modeled by the heat kernel. This approach adapts to the intrinsic geometry of $\mathcal{M}$ and captures local dependencies, enabling the model to represent complex spatial relationships that indirectly but crucially affect λ.

For the functional covariates, if $J = 1$, a common method in functional data analysis is to use the Karhunen–Loève expansion to approximate the random function [35], which approximates each functional covariate as a sum of key principal components. This enables the model to handle continuous predictors in a compact, computationally feasible manner. Specifically, the random function $X(t)$ can be written as follows:

$$X(t) = \mu(t) + \sum_{l=1}^{\infty} a_l \psi_l(t), \quad l \geq 1, \tag{3}$$

where $\mu(t)$ is the mean function of $X(t)$, $a_l = \int_{t \in \mathcal{T}} (X(t) - \mu(t)) \psi_l(t) dt, l \geq 1$ are principal components, and $\psi_l(t)$s form an orthonormal system of eigenfunctions of the covariance operator of $X(t)$:

$$\int_{t \in \mathcal{T}} \text{Cov}(X(t), X(s)) \psi_j(s) ds = \lambda_l \psi_l(t), \forall t \in [0, \mathcal{T}] \tag{4}$$

Without loss of generality, we assume that $X(t)$ is a zero-mean stochastic process, i.e., $\mu(t) = 0$. Arranging the principal components in descending order according to their corresponding eigenvalues ($\lambda_1 \geq \lambda_2 \geq \cdots$), we define $X^{(Q)}(t)$ as the approximation of $X(t)$ obtained by truncating the expansion (3) after the first Q terms, where $Q \geq 1$:

$$X^{(Q)}(t) = \sum_{l=1}^{Q} Z_l \psi_l(t), \quad Q \geq 1. \tag{5}$$

For the multivariate functional covariates (i.e., $J > 1$), the multivariate Karhunen–Loève expansion can be applied, as outlined in [36].

For the random effects associated with the non-Euclidean covariates $\boldsymbol{s} \in \mathcal{M}$, we employ a reproducing kernel-based method [37] to capture the nonlinear structure, i.e.,

$$\delta(\boldsymbol{s}) = \sum_{j=1}^{\infty} \mu_j k(\boldsymbol{s}, \boldsymbol{s}_j), \tag{6}$$

where k is a kernel function and $\{\mu_j\}_{j=1}^{\infty}$ are expansion coefficients. Non-Euclidean covariates often represent spatial information on complex domains. By using the heat kernel, we can capture spatial dependencies on these domains, integrating local geometric structures and supporting accurate predictions on spatially structured data. Given the flexibility of the heat kernel in adapting to irregular and complex data, we choose it as the kernel function for our modeling approach. To be more specific, consider $\mathcal{M}$ to be an H-dimensional complete and compact Riemannian manifold with a Riemannian metric $\mathbf{g}$. Its boundary, $\partial \mathcal{M}$, is continuous and differentiable almost everywhere. Based on this metric tensor $\mathbf{g}$, the heat kernel k on $\mathcal{M}$ is a smooth function defined on $\mathcal{M} \times \mathcal{M} \times \mathbb{R}^+$ that satisfies the heat equation [38]:

$$\frac{\partial}{\partial \theta} k(\boldsymbol{s}_1, \boldsymbol{s}_2, \theta) = \frac{1}{2} \Delta_s k(\boldsymbol{s}_1, \boldsymbol{s}_2, \theta), \quad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{M}, \quad \theta \geq 0. \tag{7}$$

where Δ_s is associated with a Laplace–Beltrami operator. The heat kernel is a symmetric and positive semidefinite kernel on $\mathcal{M}$. We use the Neumann boundary condition as the boundary condition of the heat equation, which ensures the uniqueness of the heat kernel. In the special case where $\mathcal{M}$ in (7) is a Euclidean space, the heat kernel has a closed form corresponding to a Gaussian function. When $\mathcal{M}$ in (7) is a non-Euclidean manifold, constructing the heat kernel linked to the Laplace–Beltrami operator becomes a complex task, falling under the domain of partial differential equations and differential geometry. Rather than directly solving the heat equation, we can take advantage of the fact that heat kernels correspond to the transition density of Brownian motion (BM) on $\mathcal{M}$. Our approach involves simulating Brownian motion on $\mathcal{M}$, calculating the transition density numerically, and then using these results to approximate the heat kernel.

To make the connection between the heat kernel and the transition density of BM clear, consider $\mathcal{F}(\theta)$ as a BM on $\mathcal{M}$ starting from $\boldsymbol{s}_1$ at time $\theta = 0$. The probability that $\mathcal{F}(\theta)$ lies in any Borel set $\mathcal{A} \subset \mathcal{M}$ is given by

$$\mathbb{P}[\mathcal{F}(\theta) \in \mathcal{A} \mid \mathcal{F}(0) = \boldsymbol{s}_1] = \int_{\mathcal{A}} k(\boldsymbol{s}_1, \boldsymbol{s}, \theta) d\boldsymbol{s} \tag{8}$$

where the integral is defined in terms of the volume form of a Riemannian manifold $\mathcal{M}$. Under this framework, the Neumann boundary conditions imposed on the heat kernel will approximate the heat kernel by stopping the Brownian motion process at the boundary and reflecting it, resampling the next step until the stopping time is reached.

In summary, we term the above method intrinsic functional partially linear Poisson regression (iFPLPR), which accounts for the potentially complex boundary or interior conditions, as well as the intrinsic geometry of the spaces. The model can be simplified depending on the data characteristics. When there are no random effects or vector-valued covariates, it reduces to a functional Poisson regression, focusing solely on the functional covariates and their linear relationship with the log-transformed Poisson mean. If there are no functional covariates, the model becomes a Poisson regression with random effects in complex domains, using kernel methods to capture the nonlinear relationship between the non-Euclidean covariates and the response variable. In the simplest case, when there are neither functional covariates nor random effects, the model simplifies to the classical Pois-

son regression, where only the vector-valued covariates are linearly linked to the Poisson mean. This flexibility allows the iFPLPR model to accommodate various data types, ranging from purely functional data to complex spatial data or basic Poisson regression scenarios.

2.2. Parameter Estimation

In this part, we use the information from the sample for parameter estimation. First, for the functional covariates $\boldsymbol{X}$, we use the same orthogonal basis that was used to expand both $X_j(t)$ and β_j, namely the functional principal components basis, which is widely used in functional data analysis [13]. Then, assume we have independent and identically distributed observations in our sample, denoted as $\{X_{ij}, i = 1, 2, \ldots, n, j = 1, 2, 3, \cdots, J\}$, where X_{ij} is assumed to be fully observed. The mean function is denoted by $\mu_j(t) = \mathbb{E}X_j(t)$, and the covariance function is $\Sigma_j(s, t) = \mathbb{E}[(X_j(s) - \mu_j(s)) \otimes (X_j(t) - \mu_j(t))]$, where the tensor product, denoted by $\otimes$, is a mathematical operation that takes two vectors or matrices and produces a new tensor. It expands the dimensions of the input objects by creating all possible combinations of their elements. For finite samples, the estimates of the mean and the covariance functions can be computed as follows:

$$\widehat{\mu}_j(t) = \overline{X}_j(t) = \frac{1}{n}\sum_{i=1}^{n} X_{ij}(t), \widehat{\Sigma}_j(s, t) = \frac{1}{n}\sum_{i=1}^{n}\big(X_{ij}(s) - \overline{X}_j(s)\big)\big(X_{ij}(t) - \overline{X}_j(t)\big), \quad (9)$$

and by Mercer's theorem [39–41], we have

$$\Sigma_j(s, t) = \sum_{q=1}^{\infty} \lambda_{qj}\zeta_{qj}(s)\zeta_{qj}(t), \quad \widehat{\Sigma}_j(s, t) = \sum_{q=1}^{\infty} \widehat{\lambda}_{qj}\widehat{\zeta}_{qj}(s)\widehat{\zeta}_{qj}(t), \quad (10)$$

where the eigenvalues and eigenvectors of the operator Σ_j are denoted by λ_{qj} and ζ_{qj}. The eigenvalues and eigenvectors of the operator $\widehat{\Sigma}_j$ are denoted by $\widehat{\lambda}_{qj}$ and $\widehat{\zeta}_{qj}$. Then, we use these to approximate $X_{qj}(t)$ and $\beta_j(t)$:

$$X_{ij}(t) = \mu_j(t) + \sum_{q=1}^{\infty}\big\langle X_{ij}(t) - \mu(t), \widehat{\zeta}_{qj}(t)\big\rangle\widehat{\zeta}_{qj}(t),$$

$$\beta_j(t) = \sum_{q=1}^{\infty}\big\langle\beta_j(t), \widehat{\zeta}_{qj}(t)\big\rangle\widehat{\zeta}_{qj}(t), \quad j = 1, \cdots, J. \quad (11)$$

For a suitably chosen truncation number Q_j, we set $a_{iqj} = \big\langle X_{ij}(t) - \mu_j(t), \widehat{\zeta}_{qj}(t)\big\rangle$ and $b_{qj} = \big\langle\beta_j(t), \widehat{\zeta}_{qj}(t)\big\rangle, q = 1, 2, \cdots, Q_j$. Then, we use these to approximate $X_{ij}(t)$ and $\beta_j(t)$:

$$X_{ij}^{Q_j}(t) = \mu_j(t) + \sum_{q=1}^{Q_j} a_{iqj}\widehat{\zeta}_{qj}(t) = \widehat{\mu}_j(t) + \boldsymbol{a}_{ij}^{\top}\widehat{\boldsymbol{\zeta}}_{Q_jj}(t),$$

$$\beta_j^{Q_j}(t) = \sum_{q=1}^{Q_j} b_{qj}\widehat{\zeta}_{qj}(t) = \boldsymbol{b}_j^{\top}\widehat{\boldsymbol{\zeta}}_{Q_jj}(t), \quad j = 1, \cdots, J. \quad (12)$$

where $\boldsymbol{a}_{ij} = \big(a_{i1j}, \ldots, a_{iQ_jj}\big)^{\top}$, $\boldsymbol{b}_j = \big(b_{1j}, \ldots, b_{Q_jj}\big)^{\top}$, and $\widehat{\boldsymbol{\zeta}}_{Q_jj}(t) = \big(\widehat{\zeta}_{1j}(t), \ldots, \widehat{\zeta}_{Q_jj}(t)\big)^{\top}$. In practice, the number of terms Q_j can be determined using criteria such as the fraction of variance explained or through cross-validation.

Second, for the random effects associated with the non-Euclidean covariates $\boldsymbol{s} \in \mathcal{M}$, we employ a kernel-based method to capture the nonlinear structure, with the heat kernel being an appropriate choice. However, in most cases, the heat kernel lacks an analytical expression, requiring us to estimate it. This estimation is performed by simulating BM sample paths and numerically evaluating the transition probability to approximate the integral in Equation (8). We simulate N paths, $\{\mathcal{F}(\theta) : \theta > 0\}$ on $\mathcal{M}$, starting from $\mathcal{F}(0) = \boldsymbol{s}_1$. For any $\theta > 0$ and $\boldsymbol{s} \in \mathcal{M}$, we estimate the probability of $\mathcal{F}(\theta)$ falling within a small

neighborhood $\mathcal{A}$ of s by counting how many BM sample paths reach $\mathcal{A}$, with the assumption that the BM has a higher probability of reaching the neighborhood of the target point. The transition probability is approximated as

$$\mathbb{P}[\mathcal{F}(\theta) \in \mathcal{A} \mid \mathcal{F}(0) = s_1] \approx \frac{C}{N}, \tag{13}$$

where N is the total number of simulated BM sample paths and C is the number of paths $\mathcal{A}$ at time θ. The transition density of $S(\theta)$ at s is then approximated by

$$k(s_1, s_2, \theta) \approx \frac{1}{V(\mathcal{A})} \cdot \frac{C}{N}, \tag{14}$$

where $V(\mathcal{A})$ is the Riemannian volume of the neighborhood $\mathcal{A}$, which is parameterized by the radius of $\mathcal{A}$. The Riemannian volume refers to the volume of a set in Euclidean space computed using Riemann integration, which partitions the space into small sub-regions (often rectangles) and sums their volumes.

Then, we simulate the BM paths on $\mathcal{M}$ to estimate the transition density of BM on $\mathcal{M}$. Let $\mathbf{a}(l_0) \in \mathbb{R}^{d_e}$ be such that $\phi(\mathbf{a}(l_0)) = s_0$, and let $\phi : \mathbb{R}^{d_e} \to \mathcal{M}$ be a smooth local parameterization of $\mathcal{M}$ around $s_1 \in \mathcal{M}$, where d_e is the dimension of a potential Euclidean space. In this paper, we consider the local parameterization ϕ to be known. If ϕ is unknown, it can be learned through latent variable models using nonlinear dimension reduction techniques. Under this assumption, simulating a sample path of BM on $\mathcal{M}$ with starting point s_0 is equivalent to simulating a stochastic process in $\mathbb{R}^{d_e}$ with starting point $\mathbf{a}(l_0)$. The BM on a Riemannian manifold in the local coordinate system follows a system of stochastic differential equations (SDEs) in Ito form. The Riemannian metric $\mathbf{g}$ is used to construct the SDEs:

$$da^{q_1}(l) = \frac{1}{2}\mathbf{G}^{-1/2} \sum_{q_2=1}^{d_e} \frac{\partial}{\partial x^j}\left(\mathbf{g}^{q_1 q_2}\mathbf{G}^{1/2}\right)dl + \left(\mathbf{g}^{-1/2}dB(l)\right)_{q_1}; \quad q_1, q_2 = 1, 2, \cdots, d_e, \tag{15}$$

where a^{q_1} represents the q_1-th dimension of the latent space. $\mathbf{G}$ is the determinant of $\mathbf{g}$, and $B(l)$ represents an independent BM in Euclidean space. The discrete form of the SDE is defined using Euler's formula:

$$q(\mathbf{a}(l) \mid \mathbf{a}(l-1)) = \mathbb{N}\left(\mathbf{a}(l) \mid \mu(\mathbf{a}(l-1), \Delta l), \Delta l \mathbf{G}^{-1}\right). \tag{16}$$

where

$$\mu(\mathbf{a}^{q_1}(l-1), \Delta l)_{q_1} = \mathbf{a}^{q_1}(l-1) + \frac{1}{2}\sum_{q_2=1}^{d_e}\left(-\mathbf{g}^{-1}\frac{\partial \mathbf{g}}{\partial \mathbf{a}^{q_2}}\mathbf{g}^{-1}\right)_{q_1 q_2}\Delta l + \frac{1}{4}\sum_{q_2=1}^{d_e}\left(\mathbf{g}^{-1}\right)_{q_1 q_2}\mathrm{tr}\left(\mathbf{g}^{-1}\frac{\partial \mathbf{g}}{\partial \mathbf{a}^{q_2}}\right)\Delta l.$$

The error of the heat kernel estimator, as defined in (14), is $\mathbf{O}(w^2) + \mathbf{O}(w^{-d_e}N_z^{-1})$, where w is a window size. When $N_z \to +\infty$, $w \to 0$, and $\frac{w^{-d_e}}{N_z} \to 0$, the error converges to 0. The estimation algorithm for the heat kernel is shown in Algorithm 1, where the parameter selection criteria are based on [38]. In [38], the authors also discussed the details of the convergence criteria and the optimal number of simulations. However, these are not the focus of this paper; instead, we concentrate on using the kernel functions. In Algorithm 1, the input part of the algorithm requires $X_{ij}(t)$ and $\beta_j(t)$, which represent the functional covariates and their corresponding coefficient functions, where t is in the domain $\mathcal{T}$, and Z is a vector of Euclidean covariates. s represents the non-Euclidean covariates in the domain $\mathcal{M}$. A is the radius of the neighborhood around the point of interest on the manifold $\mathcal{M}$. Finally, the algorithm outputs $k_{\mathrm{heat}}^\theta(s_1, s_2)$, which is the estimated transition probability density for the heat kernel between two points s_1 and s_2 on $\mathcal{M}$ calculated using simulated BM paths.

Algorithm 1 The estimation algorithm for the heat kernel

Input: $\{X_{ij}(t), \beta_j(t), Z \in \mathbb{R}^D, s \in \mathcal{M}, N$ (number of BM sample paths), radius of neighborhood $A\}$.
Output: Estimated heat kernel $k_{\text{heat}}^\theta(s_1, s_2)$.
  1: Initialize parameters: n, N (number of simulated BM paths)
  2: **for all** random effect $s \in \mathcal{M}$ **do**
  3:    Simulate N Brownian motion (BM) sample paths $\{S(\theta) : \theta > 0\}$ on $\mathcal{M}$ starting at s_1.
  4:    **for all** $\theta > 0$ **do**
  5:       Track the probability of each BM path reaching the neighborhood A of s;
  6:       Count C, the number of paths that reach A at time θ;
  7:       Estimate the transition probability using Formula (13).
  8:    **end for**
  9:    Approximate the transition density using Formula (14).
 10: **end for**
 11: **return** Estimated heat kernel $k_{\text{heat}}^\theta(s_1, s_2)$.

Third, by combining (12) and (14), the sample version of (2) is as follows:

$$\ln \lambda_i = \sum_{j=1}^{J} a_{ij}^\top b_j + Z_i^\top \eta + \mu^\top k_i, \quad i = 1, 2, \cdots, n \tag{17}$$

where $\mu = (\mu_1, \cdots, \mu_n)^\top$ and $k_i = (k_{\text{heat}}^\theta(s_i, s_1), \cdots, k_{\text{heat}}^\theta(s_i, s_n))^\top$. Next, we need to estimate the parameters b_j, η, μ. Let $S_i^\top \widetilde{b} = \sum_{j=1}^{J} a_{ij}^\top b_j + Z_i^\top \eta$. Then, we have

$$\ln \lambda_i = S_i^\top \widetilde{b} + \mu^\top k_i. \tag{18}$$

Let $\Psi = (\widetilde{b}^\top, \mu^\top)^\top$. The logarithmic likelihood function is as follows:

$$L(\Psi) = \ln\left[\prod_{i=1}^{n} P(Y_i = y_i | X_i, Z_i, s_i, \Psi) \right] \tag{19}$$

To make the model recognizable, we use the penalty logarithm maximum likelihood function as follows:

$$L_p(\Psi) = \sum_{i=1}^{n} \left(y_i (S_i^\top \widetilde{b} + \mu^\top k_i) - e^{S_i^\top \widetilde{b} + \mu^\top k_i} + \ln y_i! \right) - \frac{1}{2} c_1 \widetilde{b}^\top \widetilde{b} - \frac{1}{2} c_2 \mu^\top K \mu \tag{20}$$

where c_1 and c_2 are penalty parameters and $K = (k_{\text{heat}}^\theta(s_i, s_j))_{1 \le i, j \le n}$, with the penalty parameters regulating the trade-off between model fit and complexity, thus preventing overfitting. We note that a detailed analysis of the asymptotic properties of the estimators (including a preliminary discussion of the consistency and asymptotic normality of the estimators, as well as the role of penalization terms in convergence) would strengthen the article's theoretical foundation and address concerns regarding the robustness of the model. Many papers have conducted research in these areas, such as [42–45]. Although this is not the focus of this paper, we point out that a more systematic and comprehensive study of the asymptotic properties of the estimators we have presented is possible in the future.

Next, to ensure an optimal fit to the data, we use the quasi-Newton iterative algorithm [46] to solve for the parameters Ψ, which iteratively refines the parameter estimates while balancing model complexity and data fit. Specifically, we take the derivative of the log-likelihood function. The following are the first-order and second-order derivatives:

$$\frac{\partial L_p(\Psi)}{\partial \widetilde{b}} = \sum_{i=1}^{n} \left(y_i S_i - S_i e^{S_i^\top \widetilde{b} + \mu^\top k_i} \right) - \frac{1}{2} c_1 \widetilde{b} \tag{21}$$

$$\frac{\partial L_p(\Psi)}{\partial \boldsymbol{\mu}} = \sum_{i=1}^{n} \left(y_i \boldsymbol{k}_i - \boldsymbol{k}_i e^{S_i^\top \widetilde{b} + \boldsymbol{\mu}^\top \boldsymbol{k}_i} \right) - \frac{1}{2} c_2 K \boldsymbol{\mu} \tag{22}$$

$$\frac{\partial^2 L_p(\Psi)}{\partial \widetilde{b} \partial \widetilde{b}^\top} = \sum_{i=1}^{n} \left(-S_i e^{S_i^\top \widetilde{b} + \boldsymbol{\mu}^\top \boldsymbol{k}_i} S_i^\top \right) - \frac{1}{2} c_1 \boldsymbol{I} \tag{23}$$

$$\frac{\partial^2 L_p(\Psi)}{\partial \widetilde{b} \partial \boldsymbol{\mu}^\top} = \sum_{i=1}^{n} \left(-S_i e^{S_i^\top \widetilde{b} + \boldsymbol{\mu}^\top \boldsymbol{k}_i} \boldsymbol{k}_i^\top \right) \tag{24}$$

$$\frac{\partial^2 L_p(\Psi)}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^\top} = \sum_{i=1}^{n} \left(-\boldsymbol{k}_i e^{S_i^\top \widetilde{b} + \boldsymbol{k}_i^\top \mu} \boldsymbol{k}_i^\top \right) - \frac{1}{2} c_2 K \tag{25}$$

where $\boldsymbol{I}$ is an identity matrix. The following parameter estimation is performed using the quasi-Newton method:

$$\widetilde{b}^{(k+1)} = \widetilde{b}^{(k)} - \left(\frac{\partial^2 L_p(\Psi)}{\partial \widetilde{b} \partial \widetilde{b}^\top} \bigg|_{\widetilde{b}^{(k)}} \right)^{-1} \frac{\partial L_p(\Psi)}{\partial \widetilde{b}} \bigg|_{\widetilde{b}^{(k)}} \tag{26}$$

Based on the DFP method, let $\frac{\partial L_p(\Psi)}{\partial \widetilde{b}} \Big|_{\widetilde{b}^{(k)}} = g_k$ and $\left(\frac{\partial^2 L_p(\Psi)}{\partial \widetilde{b} \partial \widetilde{b}^\top} \Big|_{\widetilde{b}^{(k)}} \right)^{-1} = H_k^{-1}$. Then,

$$H_{k+1}^{-1} - H_k^{-1} = \frac{\Delta \widetilde{b}^{(k)} \Delta \left(\widetilde{b}^{(k)} \right)^\top}{\Delta g_k^\top \Delta \widetilde{b}^{(k)}} - \frac{H_k^{-1} \Delta g_k \Delta g_k^\top H_k^{-1}}{\Delta g_k^\top H_k^{-1} \Delta g_k}. \tag{27}$$

By setting a constant ε, we stop the iteration when $g_{k+1} < \varepsilon$. The same criterion is applied to the other two parameters. The iteration continues until convergence, and the final result is denoted as $\widehat{\Psi} = (\widehat{\widetilde{b}}^\top, \widehat{\boldsymbol{\mu}}^\top)^\top$. All regularization parameters and kernel parameters can be selected using the Generalized Cross-Validation (GCV) method. The estimation algorithm for the proposed method is shown in Algorithm 2. The significance of the parameters in the input section is the same as the explanation provided in Algorithm 1. The result $\widehat{\Psi}$ contains the estimated values of the functional coefficients, random effect terms, and other model parameters after the optimization process converges. The result is used to make further inferences based on the data.

Algorithm 2 The estimation algorithm for the proposed method

Input: $\left\{ X_{ij}(t), \beta_j(t), t \in \mathcal{T}, i = 1, 2, \ldots, n, j = 1, 2, 3, \cdots, J \right\}, \mathbf{Z} \in \mathbb{R}^D, \boldsymbol{s} \in \mathcal{M}$
Output: $\widehat{\Psi}$.
1: **for all** $X_{ij}(t), \beta_j(t), t \in \mathcal{T}, j = 1, 2, ..., J$ **do**
2:     Calculate the eigenvalues $\widehat{\lambda}_{qj}$ and eigenvectors $\widehat{\zeta}_{qj}$ of the sample covariance $\widehat{\Sigma}_j$. Expand $X_{qj}(t)$ and $\beta_j(t)$ using Formula (11);
3:     Choose the truncation number Q_j using the fraction of variance explained criterion or cross-validation method;
4:     Approximate $X_{ij}(t)$ and $\beta_j(t)$ using Formula (12). Then, we have $\boldsymbol{a}_{ij}, \boldsymbol{b}_j$.
5: **end for**
6: **for all** $\boldsymbol{s} \in \mathcal{M}$ **do**
7:     Simulate the BM sample paths and evaluate the transition probability using Formula (13);
8:     Approximate the transition density of $S(\theta)$ at $\boldsymbol{s}$ using Formula (14).
9: **end for**
10: Submit the data, obtain Formula (17), iterate Formulas (26) and (27) continuously until they converge, and obtain the final result $\widehat{\Psi} = (\widehat{\widetilde{b}}^\top, \widehat{\boldsymbol{\mu}}^\top)^\top$.
11: **return** $\widehat{\Psi}$.

2.3. Two-Component Poisson Mixture Regression with Random Effects

The intrinsic functional two-component partially linear Poisson mixture regression with bivariate random effects for the modeling of count data characterized by functional covariates, vector-valued covariates, and non-Euclidean covariates is defined as follows:

$$P(Y_i = y_i) = p\frac{e^{-\lambda_{1i}}\lambda_{1i}^{y_i}}{y_i!} + (1-p)\frac{e^{-\lambda_{2i}}\lambda_{2i}^{y_i}}{y_i!} \tag{28}$$

where $i = 1, 2, \cdots, n$, and

$$\ln(\lambda_{1i}) = \sum_{j_1=1}^{J_1} \int_{t\in\mathcal{T}} X_{1ij_1}(t)\beta_{1j_1}(t)\mathrm{d}t + \mathbf{Z}_{1i}^{\top}\boldsymbol{\eta}_1 + \delta_{1i}(s_{1i}) \tag{29}$$

and

$$\ln(\lambda_{2i}) = \sum_{j_2=1}^{J_2} \int_{t\in\mathcal{T}} X_{2ij_2}(t)\beta_{2j_2}(t)\mathrm{d}t + \mathbf{Z}_{2i}^{\top}\boldsymbol{\eta}_2 + \delta_{2i}(s_{2i}) \tag{30}$$

where p is a weight parameter and $i = 1, 2, \cdots, n$, $\ln(\frac{p}{1-p}) = \varepsilon$. λ_{1i} and λ_{2i} denote the mean of the two components. For any $i = 1, \cdots, n$, $j_1 = 1, \cdots, J_1$, and $j_2 = 1, \cdots, J_2$, $X_{1ij_1}(t)$, $X_{2ij_2}(t)$, $\beta_{1j_1}(t)$, $\beta_{2j_2}(t) \in L^2(\mathcal{T})$, $\mathbf{Z}_{1i} \in \mathbb{R}^{D_1}$ and $\mathbf{Z}_{2i} \in \mathbb{R}^{D_2}$, $s_{1i} \in \mathcal{M}_1$ and $s_{1i} \in \mathcal{M}_2$. We model the functional covariates and random effects using techniques similar to those in the previous section, resulting in the following version of the estimate

$$\ln(\lambda_{1i}) = S_{1i}^{\top}\widetilde{b}_1 + \boldsymbol{\mu}_1^{\top}\mathbf{k}_{1i}, \quad \ln(\lambda_{2i}) = S_{2i}^{\top}\widetilde{b}_2 + \boldsymbol{\mu}_2^{\top}\mathbf{k}_{2i} \tag{31}$$

where $S_{1i}^{\top}\widetilde{b}_1 = \sum_{j=1}^{J_1} a_{1ij}^{\top}b_{1j} + \mathbf{Z}_{1i}^{\top}\boldsymbol{\eta}_1$ and $S_{2i}^{\top}\widetilde{b}_2 = \sum_{j=1}^{J_2} a_{2ij}^{\top}b_{2j} + \mathbf{Z}_{2i}^{\top}\boldsymbol{\eta}_2$. $\boldsymbol{\mu}_1 = (\mu_{11}, \cdots, \mu_{1n})^{\top}$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \cdots, \mu_{2n})^{\top}$. $\mathbf{k}_{1i} = (k_{\text{heat}}^{\theta}(s_{1i}, s_{11}), \cdots, k_{\text{heat}}^{\theta}(s_{1i}, s_{1n}))^{\top}$ and $\mathbf{k}_{2i} = (k_{\text{heat}}^{\theta}(s_{2i}, s_{21}), \cdots, k_{\text{heat}}^{\theta}(s_{2i}, s_{2n}))^{\top}$. Thus, we obtain the parameters to be estimated: $\Psi = \{\widetilde{b}_1^{\top}, \widetilde{b}_2^{\top}, \boldsymbol{\mu}_1^{\top}, \boldsymbol{\mu}_2^{\top}, \varepsilon\}^{\top}$. The corresponding penalty logarithm maximum likelihood function is

$$L(\Psi) = \ln\left[\prod_{i=1}^{n} P(Y_i = y_i)\right] - \frac{1}{2}c_1\widetilde{b}_1^{\top}\widetilde{b}_1 - \frac{1}{2}c_2\boldsymbol{\mu}_1^{\top}K_1\boldsymbol{\mu}_1 - \frac{1}{2}c_3\widetilde{b}_2^{\top}\widetilde{b}_2 - \frac{1}{2}c_4\boldsymbol{\mu}_2^{\top}K_2\boldsymbol{\mu}_2$$

where c_1, c_2, c_3, and c_4 are penalty parameters. In the following, we use the EM algorithm for parameter estimation. First, we introduce the latent variables z_i, that is,

$$z_i = \begin{cases} 1, & y_i \text{ is from Poisson}(\lambda_{1i}), \\ 0, & y_i \text{ is from Poisson}(\lambda_{2i}). \end{cases}$$

We treat $\{z_i, i = 1, 2, ..., n\}$ as missing data, denoted as Y_m, and then record the observable data $\{(X_{1ij_1}(t), X_{2ij_2}(t), \mathbf{Z}_{1i}, \mathbf{Z}_{2i}, \delta_{1i}, \delta_{2i}, y_i) \mid i = 1, 2, ..., n; j_1 = 1, 2, \cdots, J_1; j_2 = 1, 2, \cdots, J_2\}$ as Y_0. Record the complete data as $Y_c = (Y_0, Y_m)$. Then, the penalized log-likelihood function based on the complete data Y_c can be expressed as

$$\begin{aligned} L_p^c &= \ln\left[\prod_{i=1}^{n} P(Y_i = y_i)\right] - \frac{1}{2}c_1\widetilde{b}_1^{\top}\widetilde{b}_1 - \frac{1}{2}c_2\boldsymbol{\mu}_1^{\top}K_1\boldsymbol{\mu}_1 - \frac{1}{2}c_3\widetilde{b}_2^{\top}\widetilde{b}_2 - \frac{1}{2}c_4\boldsymbol{\mu}_2^{\top}K_2\boldsymbol{\mu}_2 \\ &= -n\ln(1 + e^{\varepsilon}) + \sum_{i=1}^{n}\left[z_i\varepsilon + z_i\ln f_{1i} - \ln(y_i!) + (1 - z_i)\ln f_{2i}\right] \\ &\quad - \frac{1}{2}c_1\widetilde{b}_1^{\top}\widetilde{b}_1 - \frac{1}{2}c_2\boldsymbol{\mu}_1^{\top}K_1\boldsymbol{\mu}_1 - \frac{1}{2}c_3\widetilde{b}_2^{\top}\widetilde{b}_2 - \frac{1}{2}c_4\boldsymbol{\mu}_2^{\top}K_2\boldsymbol{\mu}_2 \end{aligned}$$

where $f_{1i} = e^{-\lambda_{1i}}\lambda_{1i}^{y_i}$ and $f_{2i} = e^{-\lambda_{2i}}\lambda_{2i}^{y_i}$. Then, the model parameters can be estimated using the Expectation Maximization (EM) algorithm [47]. Concretely, the $r + 1$ step of the EM algorithm that maximizes the above equation consists of the two steps described below.

For the E-step, assuming that the parameter estimates at step r are known to be $\Psi^{(r)}$, we have the following mathematical expectation:

$$
\begin{aligned}
Q(\Psi|\Psi^{(r)}) &= E[L_p^c(\Psi|Y_c)|Y_0, \Psi^{(r)}] \\
&= -n\ln(1 + e^{\varepsilon}) + \sum_{i=1}^{n}\left[z_i^{(r)}\varepsilon + z_i^{(r)}\ln f_{1i} - \ln(y_i!) + (1 - z_i^{(r)})\ln f_{2i} \right] \\
&\quad - \frac{1}{2}c_1\widetilde{b}_1^{(r)\top}\widetilde{b}_1^{(r)} - \frac{1}{2}c_2\boldsymbol{\mu}_1^{(r)\top}K_1\boldsymbol{\mu}_1^{(r)} - \frac{1}{2}c_3\widetilde{b}_2^{(r)\top}\widetilde{b}_2^{(r)} - \frac{1}{2}c_4\boldsymbol{\mu}_2^{(r)\top}K_2\boldsymbol{\mu}_2^{(r)}
\end{aligned}
\tag{32}
$$

where

$$
z_i^{(r)} = E(z_i|Y_0, \Psi^{(r)}) = \left.\frac{pf_{1i}}{pf_{1i} + (1-p)f_{2i}}\right|_{\Psi^{(r)}}
\tag{33}
$$

For the M-step, the maximum value of the objective function is solved as follows:

$$
\Psi^{(r+1)} = \underset{\Psi}{\arg\max}\ Q(\Psi|\Psi^{(r)})
\tag{34}
$$

It is sufficient to continue iterating until convergence using the quasi-Newton method, and then the final estimate is $\widehat{\Psi} = \{\widehat{\widetilde{b}}_1^{\top}, \widehat{\widetilde{b}}_2^{\top}, \widehat{\boldsymbol{\mu}}_1^{\top}, \widehat{\boldsymbol{\mu}}_2^{\top}, \widehat{\varepsilon}\}^{\top}$. The estimation algorithm for the proposed method is shown in Algorithm 3.

Algorithm 3 Two-component Poisson mixture regression with random effects

---

Input: Functional covariates $\{X_{1ij}(t), X_{2ij}(t)\}$, vector-valued covariates $\{Z_{1i}, Z_{2i}\}$, non-Euclidean covariates $\{s_{1i}, s_{2i}\}$, count data $\{y_i\}$, initial parameters $\{\widetilde{b}_1^{\top}, \widetilde{b}_2^{\top}, \boldsymbol{\mu}_1^{\top}, \boldsymbol{\mu}_2^{\top}, \varepsilon\}$.

Output: Estimated parameters $\widehat{\widetilde{b}}_1^{\top}, \widehat{\widetilde{b}}_2^{\top}, \widehat{\boldsymbol{\mu}}_1^{\top}, \widehat{\boldsymbol{\mu}}_2^{\top}, \widehat{\varepsilon}$.

1: Initialize parameters $\widetilde{b}_1^{\top}, \widetilde{b}_2^{\top}, \boldsymbol{\mu}_1^{\top}, \boldsymbol{\mu}_2^{\top}, \varepsilon$.
2: Set convergence tolerance ξ.
3: **while** $\left|\Psi^{(r+1)} - \Psi^{(r)}\right| \geq \xi$ **do**
4:     **E-step:**
5:     **for all** $i = 1, 2, \ldots, n$ **do**
6:         Calculate the expected value of the latent variable using Formula (32).
7:     **end for**
8:     **M-step:**
9:     Maximize the penalized log-likelihood Formula (34);
10:     Update parameters using the quasi-Newton method.
11: **end while**
12: **return** Estimated parameters $\widehat{\widetilde{b}}_1^{\top}, \widehat{\widetilde{b}}_2^{\top}, \widehat{\boldsymbol{\mu}}_1^{\top}, \widehat{\boldsymbol{\mu}}_2^{\top}, \widehat{\varepsilon}$.

---

3. Data Analysis

3.1. Simulation Studies

In this section, a Monte Carlo simulation study is conducted to assess the performance of the proposed model. Let y_i be the i-th response variable of the i-th subject corresponding to the covariates, including the functional covariates, vector-valued covariates, and non-Euclidean covariates.

Assuming a Poisson process for y_i, we have a Poisson distribution with mean λ_i given the predictors $X_i(t)$, $\boldsymbol{Z}_i$, and $\boldsymbol{s}_i$. We fit a Poisson regression model of the form

$$
\ln(\lambda_i) = \int_{t \in [0,1]} X_i(t)\beta(t)dt + \boldsymbol{Z}_i^{\top}\boldsymbol{\eta} + \delta(\boldsymbol{s}_i)
\tag{35}
$$

where $X_i(t) = \sum_{k=1}^{100} a_k k^{-3/4}\phi_k(t)$ and $\{\phi_k(t)\}_{k=1}^{100}$ are Fourier basis functions, the a_k values are generated from a normal distribution $N(0,1)$, and $\beta(t) = \sum_{k=1}^{100} k^{-2}\phi_k(t)$. Each curve $X_i(t)$ is observed at 100 regular design points. The $\boldsymbol{Z}_i$ values are generated from a five-dimensional normal distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \text{diag}(0.5)$. $\boldsymbol{\eta}$ is generated from a five-dimensional

normal distribution $N(\mathbf{0}, \mathbf{\Sigma})$, with $\mathbf{\Sigma}$ = diag(1). For the non-Euclidean covariates s_i, we consider the classical U-shaped region data, i.e., $s_i = (s_{1i}.s_{2i})$. The U-shaped structured domain is defined as a subset of $\mathbb{R}^2$, as shown in Figure 1(left). The values of the finite-area test function δ (i.e., the colors of the mappings) vary smoothly from the bottom-right to the top-right corner of the domain, ranging from −6 to 6. The black crosses represent 100 observations. The goal is to estimate the test function and make predictions at 450 equally spaced grid points in the domain.
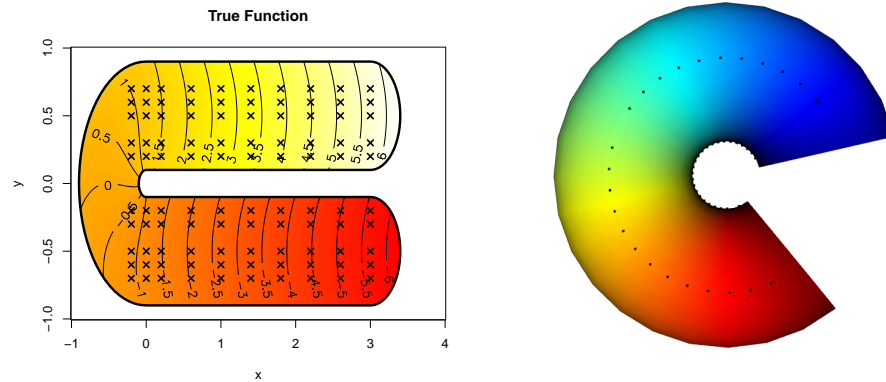
Figure 1. Sample points on U-shaped (**left**) and Bitten Torus domains (**right**).

The heat kernel is utilized to estimate the random effect of the regression function in this study. We choose the penalty parameters using the GCV method. To evaluate the performance improvement of the proposed method compared to the classical Poisson regression model, we use the ratio of the difference in absolute and relative prediction errors between the two models to the prediction errors of the Poisson regression as the evaluation criterion. We are also concerned with the accuracy of model parameter estimation. Therefore, we also present the estimation errors for the parameters Ψ. Furthermore, in order to comprehensively test the applicability and robustness of the model, we set different model parameters as follows:

(1) n = 50, 100, 150, 200, and the other parameters remain unchanged.
(2) a_k values are generated from a normal distribution $N(0, \sigma^2)$, where σ = 0.01, 0.1, 0.5, 1, and the other parameters remain unchanged.
(3) $\delta_1(s_i) = \sin(s_{1i} + s_{2i})$, $\delta_2(s_i) = \sin(s_{1i}) + s_{2i}^2$, $\delta_3(s_i) = \sin(s_{1i}) + e^{1/2s_{2i}}$, $\delta_4(s_i) = e^{1/2s_{1i}} + \cos(s_{2i})$, and the other parameters remain unchanged.
(4) Let $X(t)$ be a standard Brownian motion, $\beta(t) = \sin(3\pi/4t)$ or $\beta(t) = \cos(\pi/2t)$ or $\beta(t) = t^2$ or $\beta(t) = \pi/2 \arcsin(t)$, and the other parameters remain unchanged.
(5) Let $\mathcal{M}$ be a U-shaped region or Bitten Torus, and the other parameters remain unchanged.

As n increases, the relative error first decreases and then gradually converges. The four U-shaped structured domains, where $\delta(s_i)$ takes different forms while the other parameters remain unchanged, are shown in Figure 2. The percentage reduction in the absolute and relative errors, compared to the traditional Poisson regression model, for different forms of $n, \sigma, \delta(s_i)$, and $\beta(t)$, is shown in Table 1.
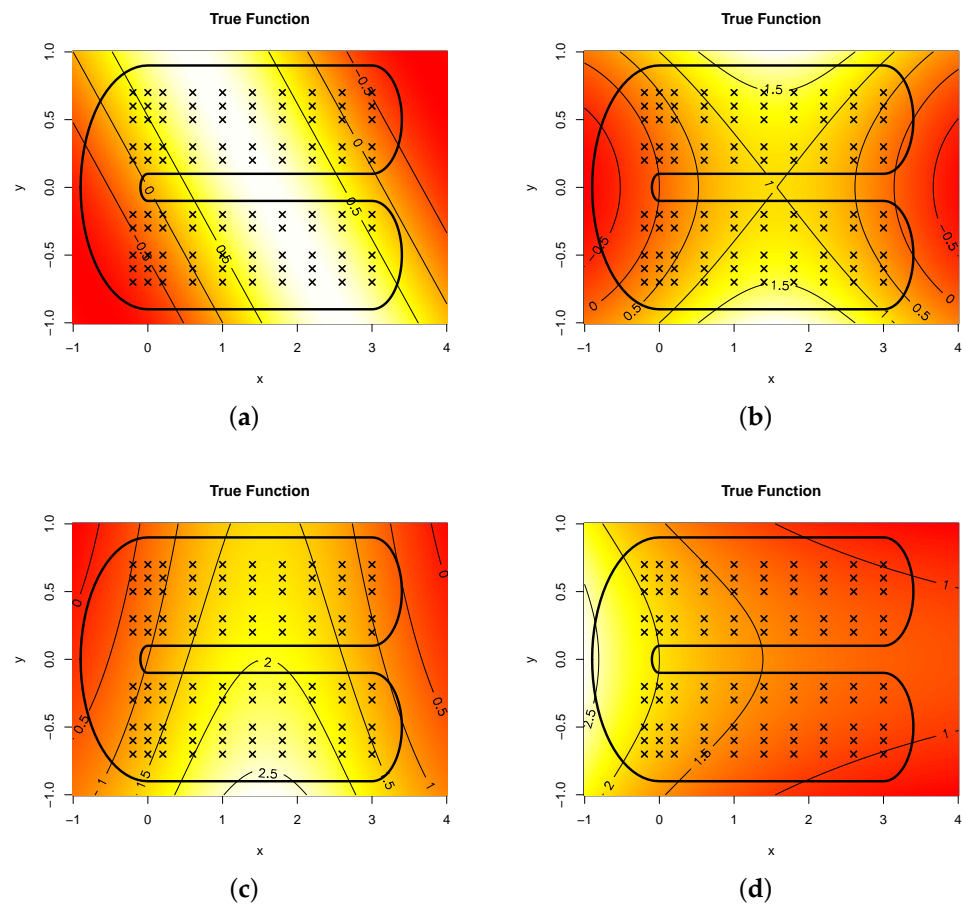
Figure 2. U-shaped domains with varying $\delta(s)$ functions. (**a**) $\delta_1(s_i) = \sin(s_{1i} + s_{2i})$; (**b**) $\delta_2(s_i) = \sin(s_{1i}) + s_{2i}^2$; (**c**) $\delta_3(s_i) = \sin(s_{1i}) + e^{-\frac{1}{2}s_{2i}}$; (**d**) $\delta_4(s_i) = e^{1/2s_{1i}} + \cos(s_{2i})$.

Table 1. The percentage reduction in absolute and relative errors.

	Definite Form	**Absolute Error (std)**	**Relative Error (std)**
n	$n = 50$	67.93% (8.69%)	80.30% (10.13%)
	$n = 100$	61.92% (9.46%)	81.61% (7.49%)
	$n = 150$	55.02% (12.10%)	82.71% (8.46%)
	$n = 200$	56.33% (8.30%)	85.38% (4.98%)
σ	$\sigma = 0.01$	58.61% (8.99%)	71.77% (9.11%)
	$\sigma = 0.1$	58.15% (9.52%)	71.77% (8.12%)
	$\sigma = 0.5$	61.56% (8.23%)	80.32% (7.80%)
	$\sigma = 1$	60.28% (8.87%)	81.42% (6.27%)
$\delta(s_i)$	$\delta(s_i) = \sin(s_{1i} + s_{2i})$	57.09% (8.62%)	79.97% (5.93%)
	$\delta(s_i) = \sin(s_{1i}) + s_{2i}^2$	55.98% (8.47%)	73.74% (11.03%)
	$\delta(s_i) = \sin(s_{1i}) + e^{1/2s_{2i}}$	54.65% (6.85%)	70.58% (8.93%)
	$\delta(s_i) = e^{1/2s_{1i}} + \cos(s_{2i})$	52.64% (6.35%)	69.82% (9.15%)
$\beta(t)$	$\beta(t) = \sin(3\pi/4t)$	58.96% (9.23%)	74.50% (10.21%)
	$\beta(t) = \cos(\pi/2t)$	58.96% (9.39%)	74.11% (7.49%)
	$\beta(t) = t^2$	57.67% (7.71%)	70.20% (8.47%)
	$\beta(t) = \pi/2\arcsin(t)$	57.15% (15.21%)	73.80% (7.04%)

When $\mathcal{M}$ is a U-shaped region and the other parameters remain unchanged, the absolute error and the relative error are reduced by 55.88% and 81.82%, respectively,

compared to the traditional Poisson regression model. The structured domain when $\mathcal{M}$ is a Bitten Torus and the other parameters remain unchanged is shown in Figure 1(right), with the absolute error and the relative error are reduced by 58.39% and 71.91%, respectively, compared to the traditional Poisson regression model.

Compared to classical Poisson regression, the distributions of the absolute error reduction percentage and the relative error reduction percentage are shown in Figures 3 and 4. The horizontal axis represents the variations in the four experimental settings (n, α_k, $\delta(s)$, and $\beta(t)$), while the vertical axis indicates the corresponding error reduction percentages. Figure 3a shows that as the sample size increases, the distribution range of the absolute error reduction percentage slightly widens but remains relatively stable overall, with a median around 0.6 and an interquartile range between 0.4 and 0.8. However, larger sample sizes may lead to the appearance of outliers. Figure 4a demonstrates that the median of the relative error reduction percentage shows an upward trend, while the distribution range changes minimally, indicating the stability of our estimates. As σ increases, the percentage of error reduction generally rises, suggesting that the value of α_k significantly impacts error reduction. The results show that setting the distribution characteristic σ of α_k around 0.5 yields the best error reduction effect and more stable data. Figures 3 and 4 also show that as $\delta(s)$ changes, the median gradually decreases, and outliers appear with certain $\delta(s)$ values, indicating that $\delta(s)$ has a considerable influence on error reduction. The figures demonstrate that with variations in $\beta(t)$, the absolute error reduction percentage concentrates around 0.6, while the relative error reduction percentage centers around 0.7. The shorter box length and similar width of absolute error indicate that changes in $\beta(t)$ have little impact on absolute error reduction.
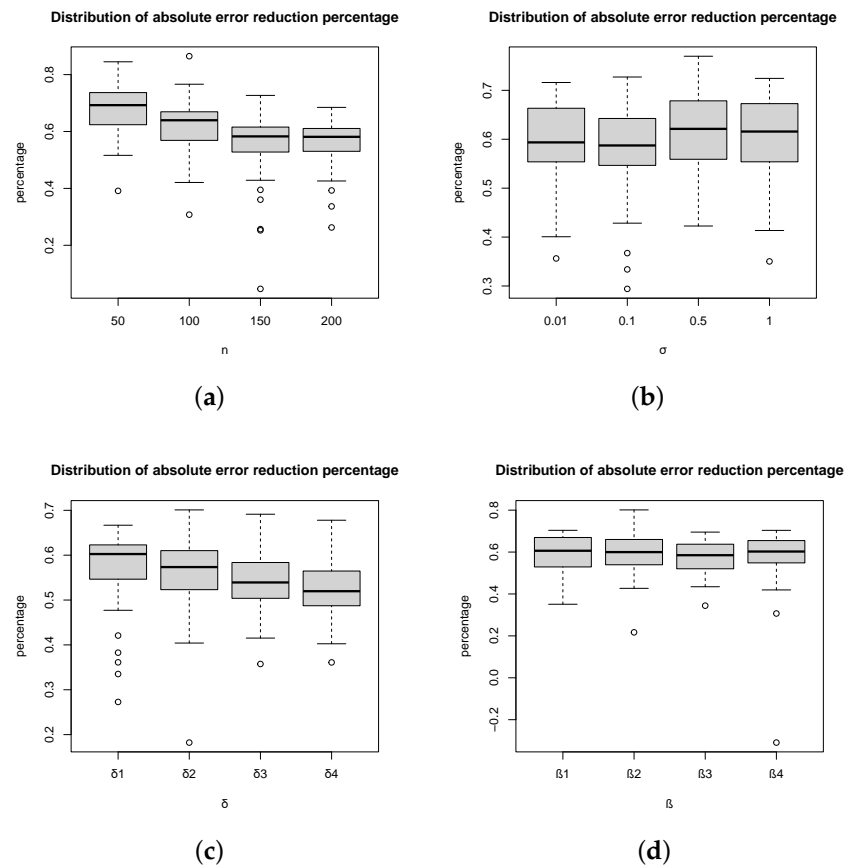

Figure 3. Absolute error reduction by sample size and settings. (**a**) Variation in n; (**b**) Variation in σ; (**c**) Variation in δ; (**d**) Variation in $\beta(t)$.
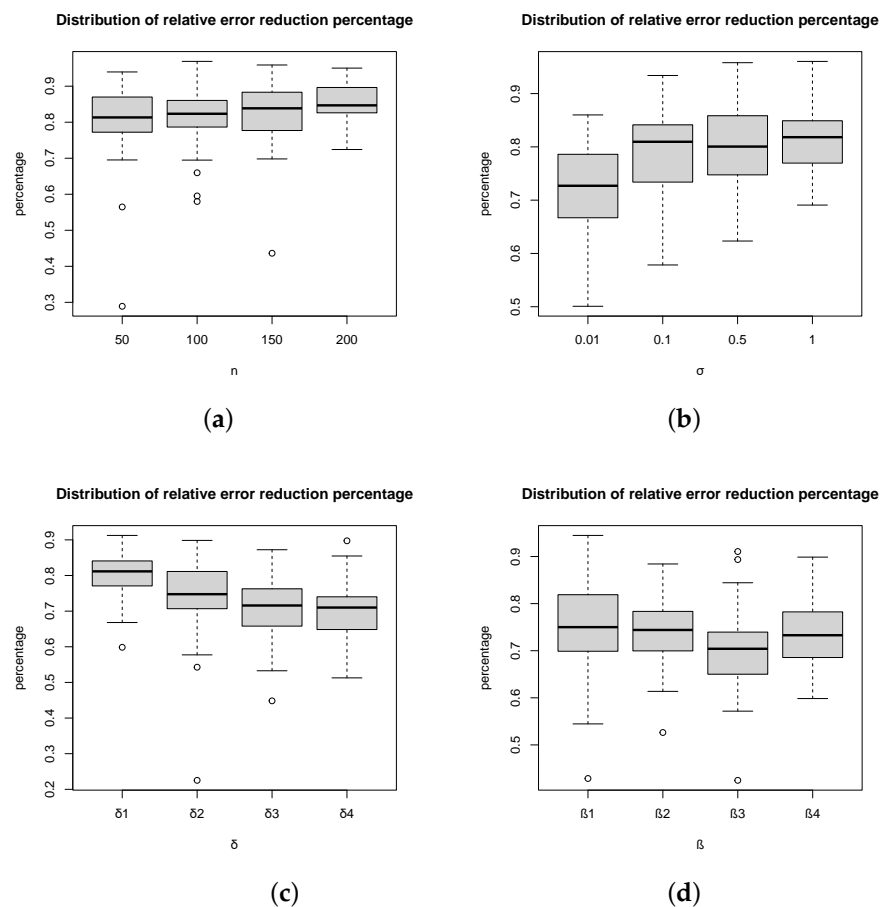
Figure 4. Relative error reduction across experimental settings. (**a**) Variation in n; (**b**) Variation in σ; (**c**) Variation in δ; (**d**) Variation in $\beta(t)$.

We should note that our simulation study can also be applied to more complex situations, such as multicollinearity of covariates or varying heterogeneity. In this paper, we consider sensitivity analysis, which is similar to the approach in [37]. This approach is more amenable to analytical treatment and allows for direct comparisons with previous literature in the field. Expanding the simulation to include more challenging scenarios, such as those involving highly correlated covariates or different types of non-Euclidean data, would provide a better understanding of the model's robustness and adaptability, offering a more comprehensive assessment, which we will explore in subsequent work. Extension to case studies involving more complex data, such as geospatial data or sensor data with temporal variables, is possible and will better illustrate the effectiveness of the proposed model in dealing with challenging real-world datasets. Although collecting real-world data is time-consuming, this will also be considered in future work.

3.2. Real Data Analysis

Taking the prediction of forest coverage for the 50 U.S. states as an example, we evaluate the performance of the proposed method in real-world data applications. A visualization of the original data is shown in Figure 5. In this paper, the forest cover percentage for the 50 U.S. states in 2023 is used as the response variable. For the predictor variables, we use the GDP time series (from 1997 to 2023, in millions of dollars, sourced from the BEA Interactive Data Application) for the 50 U.S. states as a functional covariate. We include total population and population density (people per square mile, sourced from worldpopulationreview.com), precipitation (in inches), and average temperature (in degrees Fahrenheit, sourced from the National Centers for Environmental Information) for

the 50 U.S. states in 2023 as the Euclidean covariates. For the non-Euclidean covariates, we select the latitude and longitude coordinates of the 50 U.S. states, which are then transformed into 3D coordinates on the unit sphere.
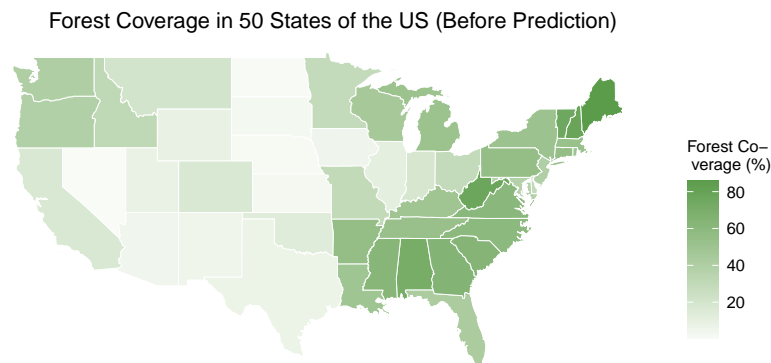
Forest Coverage in 50 States of the US (Before Prediction)

Figure 5. Observed forest coverage in U.S. states.

Next, we randomly split the dataset into a training set and a prediction set in a 16:9 ratio. For the functional data, we use B-spline basis functions for smoothing and apply functional principal component analysis (FPCA) based on cumulative contribution rate criteria for approximation. This approximation utilizes the Karhunen—Loève expansion, where each functional covariate $X_i(t)$ is expressed as

$$X_i(t) = \mu(t) + \sum_{l=1}^{Q} a_{il} \psi_l(t),$$

where $\mu(t)$ is the mean function, a_{il} are the principal component scores, and $\psi_l(t)$ are the eigenfunctions obtained from the covariance operator of $X_i(t)$. For the random effects, we employ a heat kernel function and optimize the regularization parameter using Generalized Cross-Validation (GCV). Parameter estimation in the iFPLPR model is performed by maximizing the penalized log-likelihood function:

$$L_p(\Psi) = \sum_{i=1}^{n} \left(y_i(S_i^\top b + \mu^\top k_i) - e^{S_i^\top b + \mu^\top k_i} + \ln(y_i!) \right) - \frac{1}{2} c_1 b^\top b - \frac{1}{2} c_2 \mu^\top K \mu,$$

where $S_i^\top b$ represents the linear effects and $\mu^\top k_i$ captures the nonlinear random effects. Regularization parameters c_1 and c_2 are selected through Generalized Cross-Validation (GCV), and parameter optimization is achieved using the quasi-Newton method.

The performance metrics—mean squared error, absolute error, and relative error reduction percentages—indicate the proposed model's accuracy in predicting forest coverage across U.S. states compared to classical Poisson regression models. Specifically, achieving a mean squared error reduction of 53.26%, along with reductions in the absolute and relative errors of 22.74% and 10.02%, respectively, reflects substantial improvements in predictive reliability. These error reductions are significant in practical terms, as they indicate that the iFPLPR model can more accurately capture the multifaceted relationships between forest coverage and its predictors, even when those predictors involve nonlinear, spatial, and functional data interactions. For environmental and regional planning applications, this improvement translates to better predictions of ecological outcomes, allowing for more informed decision making regarding resource allocation, conservation efforts, and urban planning in forested regions. By minimizing prediction errors, the iFPLPR model aligns closely with the study's objective to provide a flexible yet precise tool for analyzing count data within complex domains. This adaptability is particularly valuable in contexts where traditional regression models may struggle with multi-type covariates or non-Euclidean data, as often encountered in ecological, geographical, and public health studies. Additionally, Figure 6 presents a visualization of the predicted average forest cover values.

Compared to Figure 5, the results indicate that the model effectively fits the real forest cover data for the 50 U.S. states in 2023, demonstrating the method's strong performance and accuracy.
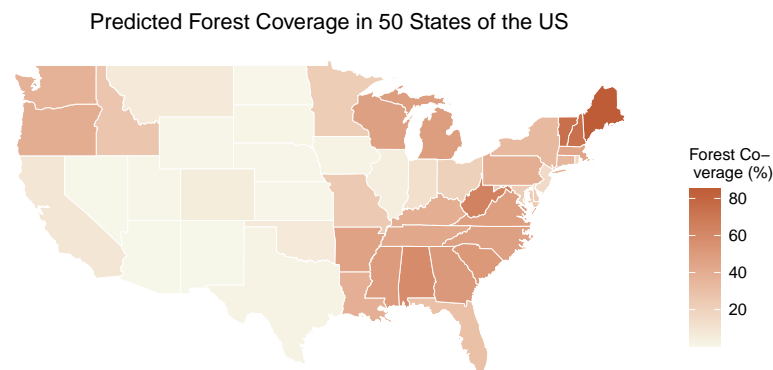
Predicted Forest Coverage in 50 States of the US

Figure 6. Predicted forest coverage using iFPLPR model.

The prediction of forest coverage across U.S. states exemplifies the versatility and robustness of the iFPLPR model, especially when dealing with count data influenced by diverse covariate types. Forest coverage is not only a critical metric for assessing ecological health but also for evaluating the effectiveness of conservation policies, urban planning, and climate resilience efforts. Accurate modeling of forest coverage can inform policymakers about sustainable development goals, help identify at-risk regions for deforestation, and support climate change mitigation strategies.

4. Conclusions

Poisson regression is a statistical method tailored for analyzing count data. In this article, we propose an intrinsic functional partially linear Poisson regression model, which applies when the functional covariates and vector-valued covariates have a linear relationship with the log-transformed Poisson mean, while nonlinear random effects exist between the non-Euclidean covariates and the Poisson mean. Compared to classical Poisson regression models, this model incorporates the intrinsic functional structure of time-series data as well as the inherent geometric information of non-Euclidean data. In the estimation process, we use a truncation method to approximate the functional covariates and apply kernel methods to model the random effects associated with the non-Euclidean covariates. The model parameters are estimated using the quasi-Newton iterative algorithm. Additionally, for the selection of the kernel function, we employ the heat kernel and estimate it using the transition probabilities of Brownian motion. Simulation studies and real data analysis show that this method demonstrates significant advantages over classical Poisson regression models.

Specifically, the iFPLPR model achieves significant reductions in absolute and relative errors, with absolute error reductions ranging from 55.88% to 67.93% and relative error improvements up to 85.38%. For instance, in scenarios with a U-shaped structured domain, the iFPLPR model reduced the absolute and relative errors by 55.88% and 81.82%, respectively, compared to the classical Poisson regression model. These enhancements illustrate the robustness of the iFPLPR model in minimizing prediction errors across diverse experimental settings, including variations in sample size (n), variance (σ), and the functional forms of $\delta(s)$ and $\beta(t)$. Such improvements underscore the iFPLPR model's utility in applications where count data exhibit complex spatial or functional structures. By effectively capturing intrinsic geometric dependencies through the use of heat kernels and smoothing techniques for functional covariates, the iFPLPR model offers more precise and dependable parameter estimates, providing substantial advantages for data analyses involving non-Euclidean or high-dimensional covariates.

For future research, on the one hand, we could explore cases where the relationship between the functional covariates, Euclidean covariates, and log-transformed Poisson mean is nonlinear. In such cases, modeling can be performed using reproducing kernel methods, particularly through multi-kernel learning to integrate data from different spaces. On the other hand, since many longitudinal datasets involve sparse functional covariates, sparse functional principal component analysis (sparse FPCA) is needed to handle such cases. As for the multi-component mixed Poisson regression model, we have only conducted theoretical derivations, and its practical performance still requires extensive simulation experiments and real data analysis for validation. Finally, the asymptotic statistical theory for model parameter estimation needs to be developed, and this work will be addressed in the future.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available because the research data are confidential.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Getaneh, F.B.; Belete, A.G.; Ayres, A.; Ayalew, T.; Muche, A.; Derseh, L. A generalized Poisson regression analysis of determinants of early neonatal mortality in Ethiopia using 2019 Ethiopian mini demographic health survey. *Sci. Rep.* **2024**, *14*, 2784. [CrossRef] [PubMed]
2. Loukas, K.; Karapiperis, D.; Feretzakis, G.; Verykios, V.S. Predicting Football Match Results Using a Poisson Regression Model. *Appl. Sci.* **2024**, *14*, 7230. [CrossRef]
3. Nzuma, J.M.; Mzera, U.I. Evaluating aflatoxin contamination control practices among smallholder maize farmers in Kilifi County, Kenya: A Poisson regression analysis. *Environ. Dev. Sustain.* **2024**, *26*, 10029–10041. [CrossRef]
4. Sakane, S.; Kato, K.; Hata, S.; Nishimura, E.; Araki, R.; Kouyama, K.; Hatao, M.; Matoba, Y.; Matsushita, Y.; Domichi, M.; et al. Association of hypoglycemia problem-solving abilities with severe hypoglycemia in adults with type 1 diabetes: A Poisson regression analysis. *Diabetol. Int.* **2024**, *15*, 1–9. [CrossRef]
5. Irshad, M.R.; Chesneau, C.; Shibu, D.S.; Monisha, M.; Maya, R. Lagrangian Zero Truncated Poisson Distribution: Properties Regression Model and Applications. *Symmetry* **2022**, *14*, 1775. [CrossRef]
6. Weaver, C.G.; Ravani, P.; Oliver, M.J.; Austin, P.C.; Quinn, R.R. Analyzing hospitalization data: Potential limitations of Poisson regression. *Nephrol. Dial. Transplant.* **2015**, *30*, 1244–1249. [CrossRef]
7. Amin, M.; Akram, M.N.; Amanullah, M. On the James-Stein estimator for the Poisson regression model. *Commun. Stat.-Simul. Comput.* **2022**, *51*, 5596–5608. [CrossRef]
8. Abdelwahab, M.M.; Abonazel, M.R.; Hammad, A.T.; El-Masry, A.M. Modified Two-Parameter Liu Estimator for Addressing Multicollinearity in the Poisson Regression Model. *Axioms* **2024**, *13*, 46. [CrossRef]
9. Jiang, F.; Ma, Y. Poisson regression with error corrupted high dimensional features. *Stat. Sin.* **2022**, *32*, 2023–2046. [CrossRef]
10. Fu, Q.; Zhou, T.Y.; Guo, X. Modified Poisson regression analysis of grouped and right-censored counts. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2021**, *184*, 1347–1367. [CrossRef]
11. Lu, M.; Loomis, D. Spline-based semiparametric estimation of partially linear Poisson regression with single-index models. *J. Nonparametric Stat.* **2013**, *25*, 905–922. [CrossRef]
12. Wang, Y.; Yu, Z. A kernel regression model for panel count data with nonparametric covariate functions. *Biometrics* **2022**, *78*, 586–597. [CrossRef] [PubMed]
13. Wang, J.L.; Chiou, J.M.; Müller, H.G. Functional data analysis. *Annu. Rev. Stat. Its Appl.* **2016**, *3*, 257–295. [CrossRef]
14. Dai, X.; Lin, Z.; Müller, H.G. Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics* **2021**, *77*, 1328–1341. [CrossRef]
15. Singh, P.K. Data with non-Euclidean geometry and its characterization. *J. Artif. Intell. Technol.* **2022**, *2*, 3–8. [CrossRef]

16. Suárez, J.L.; García, S.; Herrera, F. A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges. *Neurocomputing* **2021**, *425*, 300–322. [CrossRef]
17. Abdelwahab, M.M.; Shalaby, O.A.; Semary, H.E.; Abonazel, M.R. Driving Factors of NOx Emissions in China: Insights from Spatial Regression Analysis. *Atmosphere* **2024**, *15*, 793. [CrossRef]
18. Wood, S.N.; Bravington, M.V.; Hedley, S.L. Soap Film Smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2008**, *70*, 931–955. [CrossRef]
19. Lin, L.; Mu, N.; Cheung, P.; Dunson, D. Extrinsic Gaussian Processes for Regression and Classification on Manifolds. *Bayesian Anal.* **2019**, *14*, 887–906. [CrossRef]
20. Niu, M.; Dai, Z.; Cheung, P.; Wang, Y. Intrinsic Gaussian process on unknown manifolds with probabilistic metrics. *J. Mach. Learn. Res.* **2023**, *24*, 1–42.
21. Demšar, U.; Harris, P.; Brunsdon, C.; Fotheringham, A.S.; McLoone, S. Principal component analysis on spatial data: An overview. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 106–128. [CrossRef]
22. Erdélyi, J.; Kopáčik, A.; Kyrinovič, P. Spatial data analysis for deformation monitoring of bridge structures. *Appl. Sci.* **2020**, *10*, 8731. [CrossRef]
23. Yang, G.; Zhang, B.; Zhang, M. Estimation of knots in linear spline models. *J. Am. Stat. Assoc.* **2023**, *118*, 639–650. [CrossRef]
24. Kim, K.R.; Dryden, I.L.; Le, H.; Severn, K.E. Smoothing splines on Riemannian manifolds, with applications to 3D shape space. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2021**, *83*, 108–132. [CrossRef]
25. Mancinelli, C.; Puppo, E. Splines on Manifolds: A Survey. *Comput. Aided Geom. Des.* **2024**, *112*, 102349. [CrossRef]
26. Irshad, M.R.; Aswathy, S.; Maya, R.; Nadarajah, S. New One-Parameter Over-Dispersed Discrete Distribution and Its Application to the Nonnegative Integer-Valued Autoregressive Model of Order One. *Mathematics* **2024**, *12*, 81. [CrossRef]
27. Irshad, M.R.; Archana, K.; Al-Omari, A.I.; Maya, R.; Alomani, G. Extropy Based on Concomitants of Order Statistics in Farlie-Gumbel-Morgenstern Family for Random Variables Representing Past Life. *Axioms* **2023**, *12*, 792. [CrossRef]
28. Amewou-Atisso, M.; Ghosal, S.; Ghosh, J.K.; Ramamoorthi, R. Posterior consistency for semi-parametric regression problems. *Bernoulli* **2003**, *9*, 291–312. [CrossRef]
29. Athey, S.; Bickel, P.J.; Chen, A.; Imbens, G.W.; Pollmann, M. Semi-parametric estimation of treatment effects in randomised experiments. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2023**, *85*, 1615–1638. [CrossRef]
30. Zhao, Y.; Gijbels, I.; Van Keilegom, I. Parametric copula adjusted for non-and semiparametric regression. *Ann. Stat.* **2022**, *50*, 754–780. [CrossRef]
31. Taupin, M.L. Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Stat.* **2001**, *29*, 66–93. [CrossRef]
32. Karapiperis, D.; Tzafilkou, K.; Tsoni, R.; Feretzakis, G.; Verykios, V.S. A Probabilistic Approach to Modeling Students' Interactions in a Learning Management System for Facilitating Distance Learning. *Information* **2023**, *14*, 440. [CrossRef]
33. Karapiperis, D.; Tjortjis, C.; Verykios, V.S. A Suite of Efficient Randomized Algorithms for Streaming Record Linkage. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 2803–2813. [CrossRef]
34. Abdelwahab, M.M.; Al-Karawi, K.A.; Semary, H.E. Integrating gene selection and deep learning for enhanced Autisms' disease prediction: A comparative study using microarray data. *AIMS Math.* **2024**, *9*, 17827–17846. [CrossRef]
35. Sang, P.; Wang, L.; Cao, J. Parametric functional principal component analysis. *Biometrics* **2017**, *73*, 802–810. [CrossRef]
36. Happ, C.; Greven, S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Stat. Assoc.* **2018**, *113*, 649–659. [CrossRef]
37. Zhang, H.; Gan, J. A Reproducing Kernel-Based Spatial Model in Poisson Regressions. *Int. J. Biostat.* **2012**, *8*, 28. [CrossRef]
38. Niu, M.; Cheung, P.; Lin, L.; Dai, Z.; Lawrence, N.; Dunson, D. Intrinsic Gaussian processes on complex constrained domains. *J. R. Stat. Soc. Ser. B* **2019**, *81*, 603–627. [CrossRef]
39. Steinwart, I.; Scovel, C. Mercer's Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs. *Constr. Approx.* **2012**, *35*, 363–417. [CrossRef]
40. Mercer, J. Functions of positive and negative type, and their connection with the theory of integral equations. *Proc. R. Soc. Lond. Ser. A* **1909**, *83*, 69–70.
41. De Vito, E.; Umanità, V.; Villa, S. An extension of Mercer theorem to matrix-valued measurable kernels. *Appl. Comput. Harmon. Anal.* **2013**, *34*, 339–351. [CrossRef]
42. Ning, Y.; Liu, H. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Stat.* **2017**, *45*, 158–195. [CrossRef]
43. Belloni, A.; Chernozhukov, V.; Wang, L. Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **2011**, *98*, 791–806. [CrossRef]
44. Knight, K.; Fu, W. Asymptotics for Lasso-Type Estimators. *Ann. Stat.* **2000**, *28*, 1356–1378.
45. Liu, W. Gaussian graphical model estimation with false discovery rate control. *Ann. Stat.* **2013**, *41*, 2948–2978. [CrossRef]
46. Jin, S.; Madariaga, R.; Virieux, J.; Lambar, G. Two-dimensional asymptotic iterative elastic inversion. *Geophys. J. R. Astron. Soc.* **2010**, *108*, 575–588. [CrossRef]
47. Moon, T.K. The expectation-maximization algorithm. *IEEE Signal Process. Mag.* **1996**, *13*, 47–60. [CrossRef]