UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

ÁUREA FONSECA LOPES GALINDO

Semi-parametric models for independent and longitudinal data with limited response

Modelos semi-paramétricos para dados independentes e longitudinais com resposta limitada

Campinas

2021

Áurea Fonseca Lopes Galindo

Semi-parametric models for independent and longitudinal data with limited response

Modelos semi-paramétricos para dados independentes e longitudinais com resposta limitada

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Estatística.

Dissertation presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Statistics.

Supervisor: Caio Lucidius Naberezny Azevedo

Co-supervisor: Juvêncio Santos Nobre

Este trabalho corresponde à versão final da Dissertação defendida pela aluna Áurea Fonseca Lopes Galindo e orientada pelo Prof. Dr. Caio Lucidius Naberezny Azevedo.

Campinas

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

	Galindo, Áurea Fonseca Lopes, 1997-
G133s	Semi-parametric models for independent and longitudinal data with limited response / Áurea Fonseca Lopes Galindo. – Campinas, SP : [s.n.], 2021.
	Orientador: Caio Lucidius Naberezny Azevedo.
	Coorientador: Juvêncio Santos Nobre.
	Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.
	1. Dados de proporção. 2. Modelos aditivos generalizados. 3. Análise de regressão. 4. Inferência estatística. I. Azevedo, Caio Lucidius Naberezny, 1979-. II. Nobre, Juvêncio Santos. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Modelos semi-paramétricos para dados independentes e longitudinais com resposta limitada
Palavras-chave em inglês:
Proportional data
Generalized additive models
Regression analysis
Statistical inference
Área de concentração: Estatística
Titulação: Mestra em Estatística
Banca examinadora:
Caio Lucidius Naberezny Azevedo [Orientador]
Fernanda de Bastiani
Guilherme Vieira Nunes Ludwig
Data de defesa: 01-03-2021
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-8874-600
- Currículo Lattes do autor: http://lattes.cnpq.br/7667588501852147

Dissertação de Mestrado defendida em 01 de março de 2021 e aprovada pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO

Prof(a). Dr(a). FERNANDA DE BASTIANI

Prof(a). Dr(a). GUILHERME VIEIRA NUNES LUDWIG

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

A Deus por todo amor e todas as bênçãos que tem me concedido.
Ao meu amado marido por ser minha base e minha inspiração.

Acknowledgements

Quero começar agradecendo a Deus, por sua imensa bondade comigo por todas as oportunidades e bênçãos derramadas em minha vida.

Agradeço à minha família, em especial meus pais, Inaldo e Neusa, meus avós, Albanita, Inaldo, Ludnardo, a minha saudosa vovó Vilma (in memorian), minha tia-avó, Aleide, e minha irmã, Aline, por todo o amor, cuidado e apoio ao longo dessa jornada acadêmica. Vocês estiveram presentes em um dos passos mais importantes que dei e, mesmo tendo sido difícil, me deram todo o suporte necessário para seguir em frente e concluir mais essa etapa. Agradeço cada ligação, cada visita, cada carinho, cada conforto. Devo a vocês todas minhas conquistas. Amo vocês.

Aos meus orientadores professor Caio Azevedo e professor Juvêncio Nobre. Sou imensamente grata pelas oportunidades, incentivos, conselhos e apoio dados a mim. Os senhores são inspiração para mim e levarei os ensinamentos que recebi de vocês para toda a vida.

Ao meu companheiro de vida, João Victor, por dividir todas as dúvidas, angústias, alegrias e conquistas. Você foi fundamental para este trabalho tanto quanto é na minha vida pessoal. Obrigada por toda ajuda e amor, sua paciência comigo foi essencial.

Aos meus amados amigos Andrezza, Yuri, Thales, André, Ernandes, Diego, Carolina, Juliana, Duda, Gaby e John por todo o carinho. Mesmo de longe, vocês estiveram comigo em cada parte me dando suporte e amor. Cada um sabe da importância que tem em meu coração, amo vocês.

A todos que fazem parte Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas. Agradeço aos professores por todos os ensinamentos compartilhados, em especial: Hildete Pinheiro, Diego Bernardini, Aluísio Pinheiro, Jesus Garcia, Hélio Migon, Veroníca González-Lopez e Caio Azevedo. Também aos funcionários, em especial, Luciana e Magali.

Aos professores do Departamento de Estatística e Matemática Aplicada da Universidade Federal do Ceará.

À professora Fernanda Bastiani, agradeço pela atenção e disponibilidade em esclarecer as minhas dúvidas.

À Roberta dos Santos por ter sido tão solicita em me ajudar com as rotinas computacionais. Agradeço sua atenção e gentileza.

Aos membros participantes da banca examinadora pelo aceite, pelas sugestões e contribuições positivas.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Por fim, agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro a este projeto, por meio de uma bolsa de Mestrado, processo n° 2018/26784-1.

Resumo

O principal foco deste trabalho consiste na modelagem de dados (in)dependentes no intervalo [0,1], e respectivos casos particulares. Nesse sentido desenvolvemos duas abordagens, sob o enfoque frequentista. A primeira corresponde a uma nova classe de modelos baseados nas distribuições beta, simplex e beta retangular e, em suas respectivas versões aumentadas em zeros e/ou uns. Consideramos preditores, com uma parte paramétrica e outra não paramétrica, para a média, e preditores paramétricos para a dispersão/precisão e probabilidades de ocorrência de zeros e uns (no caso das distribuições aumentadas). Uma das vantagens dessa estrutura é poder modelar de forma satisfatória fenômenos lineares e não lineares. Na segunda abordagem, o foco são dados dependentes. De modo análogo ao caso anterior, foram desenvolvidos modelos de regressão baseados nas supracitadas distribuições, também sob o enfoque semi-paramétrico para a média (uma parte paramétrica e outra não paramétrica para o preditor) e paramétrica para as demais quantidades, considerando estruturas de correlação via Equações de Estimação Generalizadas (EEG). As duas abordagens foram desenvolvidas utilizando funções de ligação, para a média, simétricas e assimétricas, o que corresponde a generalizações de metodologias existentes. Além das novas classes de modelos, técnicas de seleção estrutural e diagnóstico, medidas de influência global e local, foram desenvolvidas para todos os modelos, em ambas as abordagens. Além disso, foram conduzidos estudos de simulação apropriados e análise de dados reais para ilustrar o potencial das metodologias desenvolvidas.

Palavras-chave: Dados de proporção; Modelos aditivos generalizados; Análise de regressão; Inferência estatística.

Abstract

The main focus of this work consists on the modeling of (in)dependent data in the interval [0,1], and the respective particular cases. In this sense, we developed two approaches, under the frequentist paradigm. The first one corresponds to a new class of models based on the beta, simplex and beta rectangular distributions and, in their respective augmented zero and/or one versions. We consider predictors, with a parametric and a nonparametric part, for the mean, and parametric predictors for the dispersion / precision and probabilities of occurrences of zeros and ones (for the augmented distributions). One of the advantages of this structure is to model linear and non-linear phenomena, properly. In the second approach, we focus on dependent data. In a similar way to the previous case, regression models were developed based on the aforementioned distributions, also under the semi-parametric approach for the mean (a parametric part and a nonparametric part for the predictor) and parametric for the other quantities, considering correlation structures via Generalized Estimation Equations (GEE). The two approaches were developed considering symmetric and asymmetric link functions, which correspond to generalizations of existing methodologies. In addition to the new model classes, structural selection, model fit assessment tools, measures of global and local influence, were developed for all models, for both approaches. In addition, appropriate simulation studies and real data analysis were conducted to illustrate the potential of the developed methodologies.

Keywords: Proportion data; Generalized additive models; Regression analysis; Statistical inference.

Contents

1 Introduction

In several fields of knowledge, as Biology, Chemistry, Physics, Psychometric among others, it is very common to deal with limited (bounded) response variable. That is, variables assuming values on the interval [a,b], as well as the respective particular cases: (a,b], [a,b), (a,b), $-\infty < a < b < \infty$. These intervals can be translated to the respective [0,1] type intervals (including the respective particular cases). In the past, it was common to transform the response to the real interval (see Atkinson (1985)), allowing the use of the usual normal homoscedastic regression model. However, this approach tends to present some problems. One of them is the fact that the parameters can not be easily interpreted in terms of the original response. Another drawback is that even after transforming, asymmetry and/or heavy tails may still remain in the response distribution. A third problem is that the parameter estimates tend to be biased. Therefore, such approach is not recommended. Other models, commonly used, as the Generalized Linear Model (GLM), proposed by Nelder and Wedderburn (1972), establish some assumptions that limits their application in such situations, as the assumption of the response distribution belongs to the linear exponential family.

To overcome these limitations, several approaches, considering the original response, have been proposed. Among others, we can cite: the Simplex regression model (Barndorff-Nielsen and Jørgensen, 1991), which belongs to the class of Dispersion Models defined by Jørgensen (1987), the Beta regression model (Ferrari and Cribari-Neto, 2004), which is based on the class of GLM and the Beta Rectangular regression model (Bayes et al., 2012), proposed as a robust alternative, that is, to properly accommodate outliers (extreme observations).

Other interesting approaches are: the regression model based on the class of Johnson SB distributions, proposed by Lemonte and Bazán (2015), the Unit Gamma regression model proposed by Mousa et al. (2016), the log-weighted Exponential regression model introduced by Altun (2019) and the class of regression models based on the unit-Lindley distribution proposed by Mazucheli et al. (2019). Furthermore, Altun and Cordeiro (2020) proposed a regression model based on the unit-improved second-degree Lindley distribution, Barreto-Souza et al. (2020) developed the Bessel regression model and Altun et al. (2021) introduced the log-Bilal regression model.

On the other hand, within the context of bounded response modeling, several model diagnostic tools and information criteria for model selection, have been developed over the years. We can cite: Espinheira et al. (2008) who proposed residuals for the beta regression model with non-varying dispersion, Ferrari et al. (2011) who proposed diagnostic tools for beta regression models with varying dispersion, Zhao et al. (2014) and Bayer and Cribari-Neto (2017), who proposed variable selection for varying dispersion beta regression models, Rocha and Simas (2011) and Espinheira and Silva (2019) that developed residual and influence analysis to a general class of beta and simplex regression models, respectively.

All the above works are suitable for data on the (0,1) intervals. That is, they are not suitable for modeling data in the interval $[0, 1]$ (since they do not deal with the values 0 and/or 1). In this work, we will use the term “augmented” to refer to data/models that consider 0 and/or 1 values. Therefore, we will name as “augmented” the works that use the term “inflated”, under the above meaning. To handle this situation, some works have been developed as: the Zero-or-One Inflated Beta regression models (Ospina and Ferrari, 2012), the Zero-and-One Augmented Simplex regression model (Liu et al., 2020) and Zero-and-One Augmented Beta Rectangular regression model (Silva et al., 2020; Nogarotto et al., 2020), which allow zero and/or one observations.

Besides the regression for response mean, quantile regression models for bounded data has been developed. Some works are: the L-Logistic regression models (da Paz et al., 2019), median regression model using the generalized Johnson-SB distributions (Rodrigues et al., 2019), Johnson-t quantile regression model (Lemonte and Moreno-Arenas, 2019), the Kumaraswamy regression model (Mitnik and Baek, 2013), the logit skew Student-t and generalized Tobit regression models (Hossain et al., 2016).

All the above works deal with bounded independent data. However, the interest on properly modeling bounded longitudinal data has been growing (see Galvis et al. (2014) and Wang and Luo (2014) for more details). In this case, besides the peculiar nature of the response, it is also expected to observe within-units dependence. Therefore it is necessary to consider both features, in order to obtain reliable inference.

For longitudinal (dependent) data restrict to interval $(0, 1)$, different approaches have been proposed, mainly based on random effects regression models and Generalized Estimating Equations (GEE). For example, Song and Tan (2000) and X.-K. Song et al. (2004) proposed Simplex regression models with non-varying and varying dispersion, via GEE, respectively. Venezuela (2008) proposed a Beta Regression model via GEE, whereas, Qiu et al. (2008) introduced a random effects Simplex Model. On the other hand, Verkuilen and Smithson (2012) proposed a random effects Beta Regression model. Petterle et al. (2019) proposed a Quasi-Beta Longitudinal Regression Model which does not require distributional assumptions for the response vector, within an approach similar to the GEE one. Freitas et al. (2021) proposed a class of regression models based on the Simplex, Beta and Unit Gamma, with varying dispersion/precision parameter, via GEE.

Also, for longitudinal (dependent) bounded augmented ([0,1]) data some works have been developed. We can cite the Zero-or-One Augmented Beta Rectangular Regression with random effects (Wang and Luo, 2014) and Galvis et al. (2014), who developed a random effects Zero-and-One Augmented Beta and Simplex regression models.

For dependent (longitudinal) data the GEE based models have some advantages over the random effect models. Indeed, the former allows us: to properly model population average behavior over covariates of interest; to explicitly model the marginal distributions and to consider different types of within-subject correlation matrices.

Another issue is that, very often, the response and some covariates present an unclear or a non-linear relation. In this case, under independence assumption, the use of the Generalized Additive Partial Linear Models (GAPLM) or Semi-parametric Additive Models (SPAM) (Hastie and Tibshirani, 1990) can be quite useful. For dependent (longitudinal) data, the use of GAPLM based on GEE (Generalized Estimation Equations) is a useful approach (see, for example, Lian et al. (2014), Wang et al. (2014), Manghi et al.(2017,2019)). These models assume that the response distribution belongs to Linear Exponential Family, considering both linear and non-linear relations, between the response and the covariates, through parametric and non-parametric components, respectively. These models are extensions of the Generalized Additive Models (GAM), which consider a full non-parametric linear predictor. In this area of GAM and GAPLM models, we can cite some works as: Green and Yandell (1985), Hastie and Tibshirani (1990), Rigby and Stasinopoulos (1996), Rigby and Stasinopoulos (2005),Yu and Ruppert (2002), Wang et al. (2011), Hernando Vanegas and Paula (2016) and Yang et al. (2019).

Although the growning literature in this field, there are few works, under the frequentist approach, with semi-parametric predictors, for independent-bounded data. Within this context, we can cite: Rigby and Stasinopoulos (2005), who proposed a family of models which includes a semi-parametric beta regression model and Ibacache-Pulgar and Figueroa-Zuniga (2019), who proposed a semi-parametric additive Beta regression model considering non-varying precision. Under the Bayesian approach, Duan et al. (2019) proposed a semi-parametric additive Simplex regression model with heterogeneous dispersion. For correlated-bounded data under the frequentist approach with semi-parametric predictors, to the best of our knowledge, no works are available in the literature. Also, there are few works under the frequentist approach that handle with link function misspecification, see, for example, Pereira and Cribari-Neto (2014) and Canterle and Bayer (2019).

The main goal of this work is to develop two flexible class of semi-parametric models for bounded (augmented) data (and the respective particular cases). One for independent data and another for longitudinal (dependent) data. We considered the frequentist paradigm for developing all inferential tools. We consider the beta, the simplex and the beta rectangular distributions, as well as their zeros and/or ones augmented versions.

In Table 1, are presented a summary of all models that are proposal of our

Table 1 – Proposal models

Regression model	Linear predictor	Link function for mean	Data type
beta, simplex and BR	semi-parametric (μ) parametric (ϕ)	logit probit cauchit cloglog loglog	independent correlated
ZOAB, ZOAS and ZOABR	semi-parametric (μ) parametric (ϕ) parametric (p_0 and p_1)		

*μ is the response mean, ϕ is the dispersion/precision parameter, p_0 and p_1 probabilities of occurrences of zeros and ones, respectively.

work. We will propose twelve new models and considering five link functions for mean parameter, developing all inferential tools for them. Specifically, the goals of this work are:

1. To developed a general class of models for independent bounded (augmented) observations with a semi-parametric predictor for the mean, considering symmetric and asymmetric link functions (including the logit and probit ones), with parametric predictors for the dispersion/precision and probabilities of occurrences of zeros and ones (for the models based on the augmented distributions);

2. To propose a similar approach of the point 1, for longitudinal (dependent) data);

3. For the two above classes, to develop model fit assessment tools, model comparison statistics (information criteria) and influence analysis tools;

4. To perform appropriate simulation studies in order to study the performance of all developed tools, under different scenarios of practical interest, also comparing them with alternative approaches.

5. To present the modeling of real problems, based on our developments, where it is shown the advantages of them over usual approaches.

1.1 Distributions

In this section, we present a short review of the distributions here considered for the developed regression models.

1.1.1 Beta distribution

A random variable Y follows a beta distribution if its density is given by:

$$f_Y(y;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}\mathbb{I}_{(0,1)}(y),$$

where $\alpha, \beta > 0$ and $\Gamma(.)$ is the gamma function $\Gamma(\theta) = \int_0^\infty y^{\theta-1}e^{-y}$. Ferrari and Cribari-Neto (2004) proposed a reparametrization of the beta distribution, that is $\mu = \alpha/(\alpha+\beta)$ and $\phi = \alpha + \beta$, i.e, $\alpha = \mu\phi$ and $\beta = (1-\mu)\phi$. Thus, the respective density is given by

$$b_Y(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}\mathbb{I}_{(0,1)}(y), \tag{1.1}$$

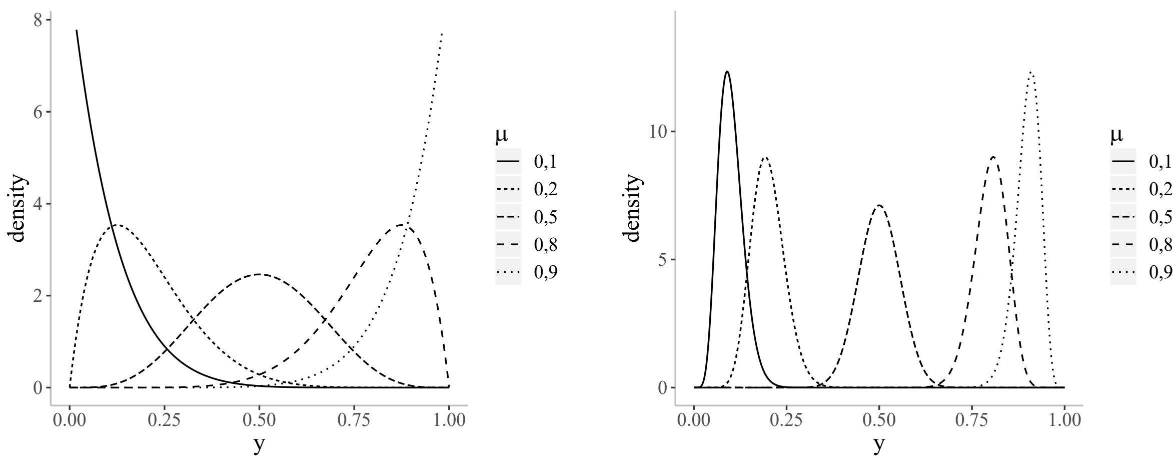
where $0 < \mu < 1$ and $\phi > 0$, with the notation $Y \sim \text{beta}(\mu, \phi)$. Also, the mean and the variance are given, respectively, by $\mathbb{E}(Y) = \mu$ and $\text{Var}(Y) = \frac{V(\mu)}{1+\phi}$, where $V(\mu) = \mu(1-\mu)$ and ϕ is a precision parameter. As shown in Figure 1, the beta distribution is very flexible. That is, the respective density can assumes different behaviors, as asymmetric, symmetric and U-shaped. Also, in Figure 2 are illustrated the behavior of the skewness and the kurtosis, as functions μ and ϕ, which are given, respectively, by

$$\text{Skewness} = \frac{\mathbb{E}[(Y-\mathbb{E}(Y))^3]}{\mathbb{E}[(Y-\mathbb{E}(Y))^2]^{3/2}} \quad \text{and} \quad \text{Kurtosis} = \frac{\mathbb{E}[(Y-\mathbb{E}(Y))^4]}{\mathbb{E}[(Y-\mathbb{E}(Y))^2]^2},$$

in particular, for beta distribution,

$$\text{Skewness} = \frac{2(\phi - 2\mu\phi)\sqrt{\phi+1}}{(\phi+2)\sqrt{\mu(1-\mu)\phi^2}} \quad \text{Kurtosis} = \frac{6(\phi+1)(2\mu\phi-\phi)^2}{\mu(1-\mu)\phi^2(\phi+2)(\phi+3)} - \frac{6}{(\phi+3)}.$$

It is possible to notice that as ϕ increases for $\mu > 0.5$, the skewness increases as well. On the other hand, when $\mu < 0.5$, as ϕ decreases, the skewness increases. Furthermore, for the kurtosis, as ϕ decreases and $\mu < 0.25$ or $\mu > 0.75$, the kurtosis increases, whereas the opposite behaviour is observed when $0.25 < \mu < 0.75$.



(a) $\phi = 10$ (b) $\phi = 80$

Figure 1 – Beta density for different values of μ, with $\phi = 10$ (a) and $\phi = 80$ (b).

1.1.2 Beta Rectangular distribution

Despite the beta distribution be very flexible, as pointed out by Hahn (2008), it presents light-tails. That is, it may not accommodate extreme values (outliers), properly.

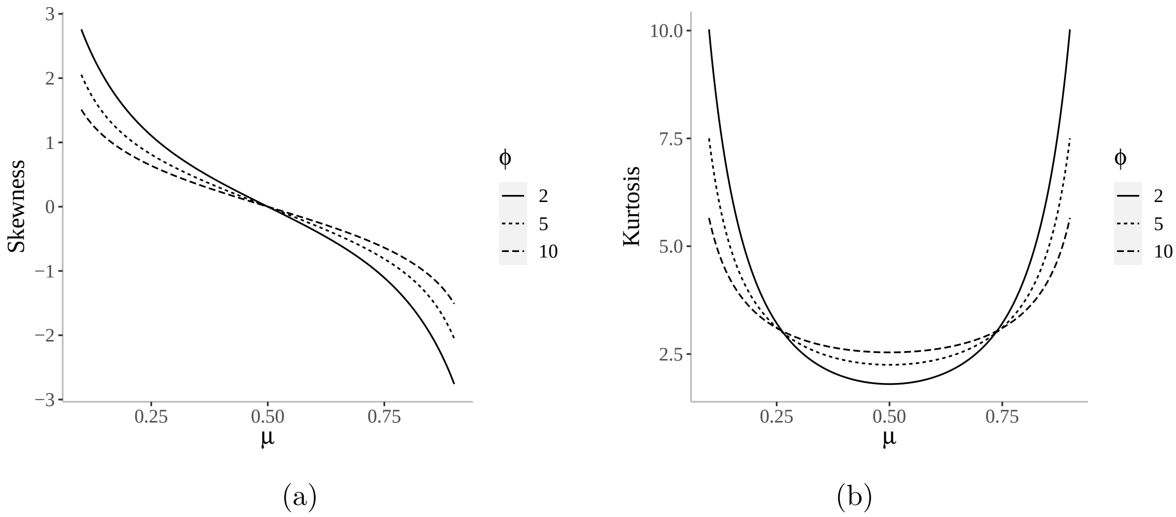


Figure 2 – Skewness (a) and Kurtosis (b) of the beta distribution *versus* the average (μ), for different values of ϕ.

This can be a limitation for data modeling, since it is not uncommon to observe extreme observation (outliers). To circumvent this limitation, Hahn (2008) proposed the beta rectangular distribution (BR), which is a mixture between a beta distribution and the standard rectangular (uniform) distribution. The respective density is given by

$$f_Y(y;\delta,\phi,\varepsilon) = \varepsilon\mathbb{I}_{(0,1)}(y) + (1-\varepsilon)b_Y(y;\delta,\phi),$$

where $0 \le \varepsilon \le 1$ is a mixture parameter and $b_Y(y;.,.)$ is the beta density, given in Equation (1.1). It is possible to notice that when $\varepsilon = 0$ the beta distribution is recovered, as well as when $\varepsilon = 1$ the standard rectangular distribution is obtained. Due to its mixture structure, the BR distribution can accommodate more properly, extreme observations.

Usually, for regression purposes, the response mean is modeled, which is the case of this work. However, according to Bayes et al. (2012) if $\mathbb{E}(Y) = \dfrac{\varepsilon}{2} + (1-\varepsilon)\delta = \mu$ is considered, we have that $0 < \varepsilon < 1 - |2\mu - 1|$, which makes more difficulty to define suitable regression models.

In order to obtain a more adequate regression structure, Bayes and Bazán (2014) define

$$\mu = \frac{\varepsilon}{2} + (1-\varepsilon)\delta \quad \text{and} \quad \alpha = \frac{\frac{\varepsilon}{2}\left(1-\frac{\varepsilon}{2}\right)}{\frac{\varepsilon}{2}\left(1-\frac{\varepsilon}{2}\right) + (1-\varepsilon)^2\delta(1-\delta)},$$

such that the parameter space of μ and α is now the rectangle given by $\{0 \le \mu \le 1, 0 \le \alpha \le 1\}$. Under this parametrization, we have that

$$\varepsilon = 1 - \sqrt{1 - 4\alpha\mu(1-\mu)} \quad \text{and} \quad \delta = \frac{\mu - \frac{1}{2} + \frac{1}{2}\sqrt{1-4\alpha\mu(1-\mu)}}{\sqrt{1-4\alpha\mu(1-\mu)}}, \tag{1.2}$$

and the implied density is given by

$$h_Y(y;\mu,\phi,\alpha) = \left(1-\sqrt{1-4\alpha\mu(1-\mu)}\right)\mathbb{I}_{(0,1)}(y) + \sqrt{1-4\alpha\mu(1-\mu)}\times \\ \times\, b_Y\left(\frac{\mu-\frac{1}{2}+\frac{1}{2}\sqrt{1-4\alpha\mu(1-\mu)}}{\sqrt{1-4\alpha\mu(1-\mu)}},\phi\right). \tag{1.3}$$

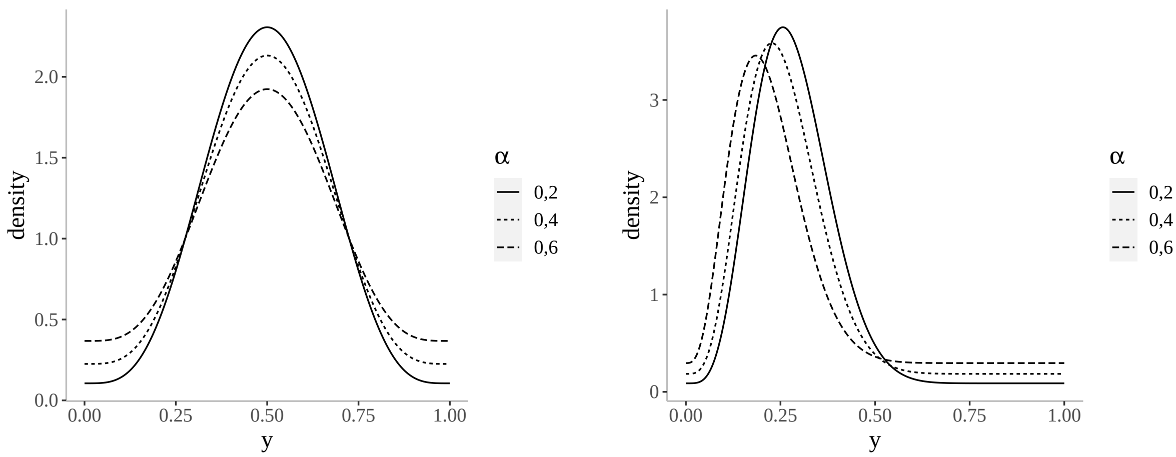
In this case, we define $Y \sim \mathrm{BR}(\mu,\phi,\alpha)$, where the quantity α is a shape parameter, which is associated with the behavior of the tails of the distribution and ϕ is a precision parameter. The associated mean and variance are given, respectively, by

$$\mathbb{E}(Y) = \mu \tag{1.4}$$

$$\mathrm{Var}(Y) = \frac{V(\delta)}{1+\phi}(1-\varepsilon)[1-\varepsilon(1+\phi)] + \frac{\varepsilon}{12}(4-3\varepsilon), \tag{1.5}$$

where $V(\delta) = \delta(1-\delta)$.

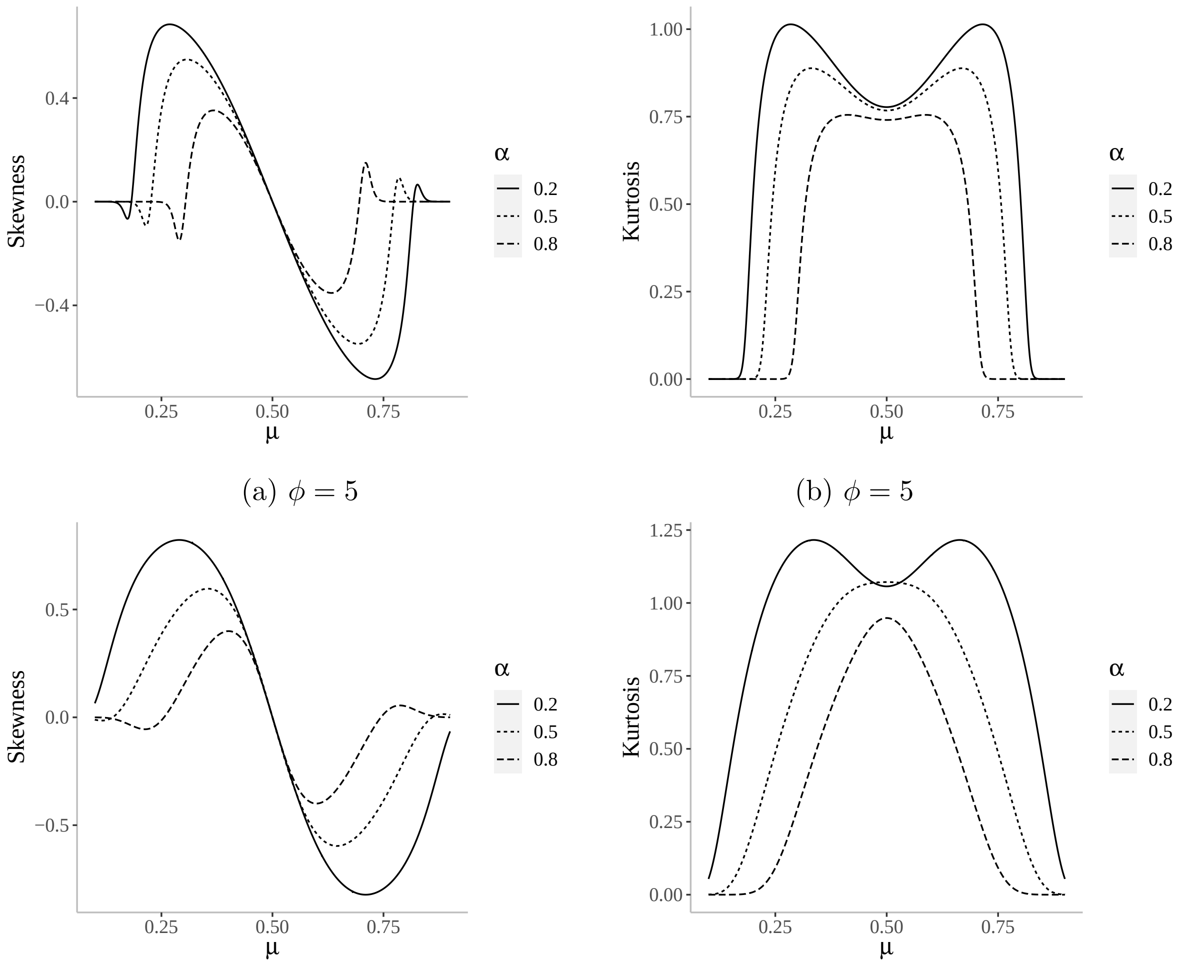
From Figure 3, it can be noticed that the BR distribution is as flexible as the beta model; However the behavior of the respective tails are more flexible, presenting an equal or smaller kurtosis, compared with the beta distribution. As α increases, the tails become heavier, that is, the outliers can be more properly accommodate. In Figure 4, it is possible to see the behavior of the skewness and kurtosis for different parameter values. The skewness, for $\mu < 0.5$, increases as α decreases, whereas, for $\mu > 0.5$, the skewness increases as α increases. On the other hand, the kurtosis increases, as α decreases. Compared with the beta distribution, despite the ϕ values, for most values of μ, the kurtosis of the BR is lower than the respective value for the beta distribution. In general, the BR distribution can present heavier tails as well higher (absolute) values, for the skewness than the beta distribution.



(a) $\phi = 10$ and $\mu = 0,5$

(b) $\phi = 80$ and $\mu = 0,2$

Figure 3 – Beta rectangular density for different values of α (indicated in the graphs), with $\phi = 10$ and $\mu = 0,5$ (a) and $\phi = 80$ and $\mu = 0,2$ (b).



(a) $\phi = 5$ (b) $\phi = 5$

(c) $\phi = 50$ (d) $\phi = 50$

Figure 4 – Skewness (a) and (c) and Kurtosis (b) and (d) of beta rectangular distribution *versus* μ for different values of α (indicated in the graphs) considering $\phi = 5$ ((a) and (b)) and $\phi = 50$ ((c) and (d)).

1.1.3 Simplex distribution

The simplex distribution is a member of a broader class, proposed by Barndorff-Nielsen and Jørgensen (1991), defined on the interval $(0, 1)$. A random variable Y following a simplex distribution, with mean $\mu \in (0, 1)$ and dispersion parameter $\phi > 0$, denoted by $Y \sim \text{simplex}(\mu, \phi)$, has a density given by

$$p_y(y; \mu, \phi) = \frac{1}{\sqrt{2\pi\phi[y(1-y)]^3}} \exp\left\{-\frac{1}{2\phi} d(y; \mu)\right\} \mathbb{I}_{(0,1)}(y), \tag{1.6}$$

where

$$d(y; \mu) = \frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}. \tag{1.7}$$

Jorgensen (1997) showed that the mean and the variance of distribution are given, respectively, by

$$\mathbb{E}(Y) = \mu \quad \text{and}$$
$$\text{Var}(Y) = \mu(1-\mu) - \frac{1}{\sqrt{2\phi}} \exp\left(\frac{1}{2\phi\mu^2(1-\mu)^2}\right) \Gamma\left(\frac{1}{2}, \frac{1}{2\phi\mu^2(1-\mu)^2}\right), \tag{1.8}$$

where $\Gamma(.,.)$ represents an incomplete gamma function, that is, $\Gamma(a,b) = \int_b^\infty y^{a-1}e^{-y}dy$.

In Figure 5 we show different curves for the simplex density, under different values of the parameters. Notice that the smaller ϕ, the more symmetric is the density. Bonat et al. (2018) made a comparison study between the simplex and the beta distributions, concluding that, usually, the beta's tails are heavier than those of the simplex distribution.

Figure 6 focus on the skewness and kurtosis of the simplex distribution, in terms of μ and ϕ. We can see that as *phi* increases, the density becomes more skewed. In addition, as ϕ increases and μ moves away from 0.5, the kurtosis increases.

Comparing to the two previous distributions, the simplex and beta have similar shapes for kurtosis and skewness, whereas BR proved have heavier tails under certain conditions and higher (absolute) values, for the skewness, which can implies more robustness and flexibility in some cases.

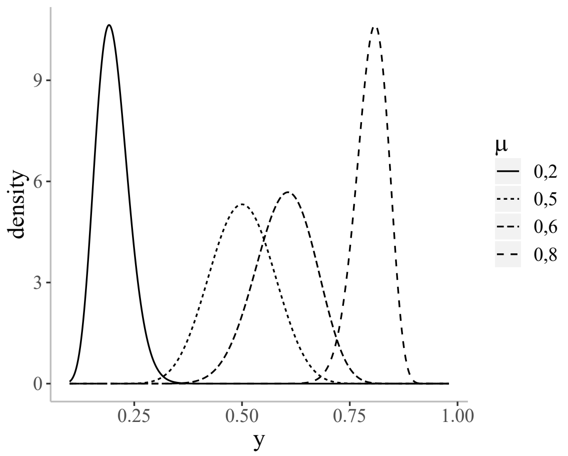
1.1.4 Zero-and/or-one Augmented distributions

The Zero-and/or-One Augmented distributions were developed for dealing with $[0,1]$ data. That is, besides the continuous values within the interval (0,1), the observations 0 and 1 are considered. It is obtained from the mixing between some continuous distribution with support in $(0,1)$ and the Bernoulli distributions. The continuous distribution is used to model the continuous component, whereas the Bernoulli distribution accounts for the discrete part, that is, the probabilities related to zeros and ones.

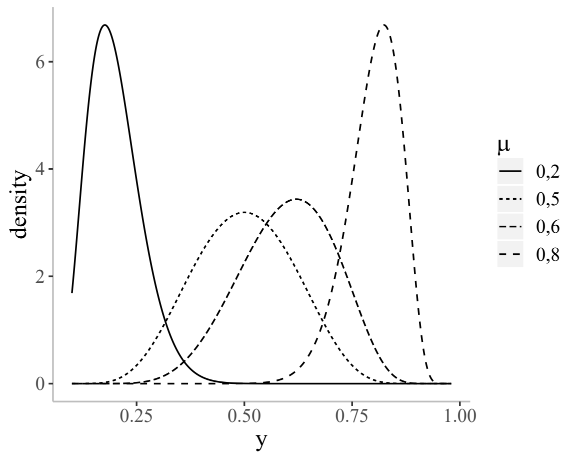
It is said that Y has a Zero-and-One Augmented distribution, if its density is given by

$$r_Y(y;\upsilon,\varrho,\varpi) = [\upsilon(1-\varrho)^{1-y}\varrho^y]\mathbb{I}_{\{0,1\}}(y) + (1-\upsilon)b_Y^*(y;\varpi)\mathbb{I}_{(0,1)}(y), \tag{1.9}$$

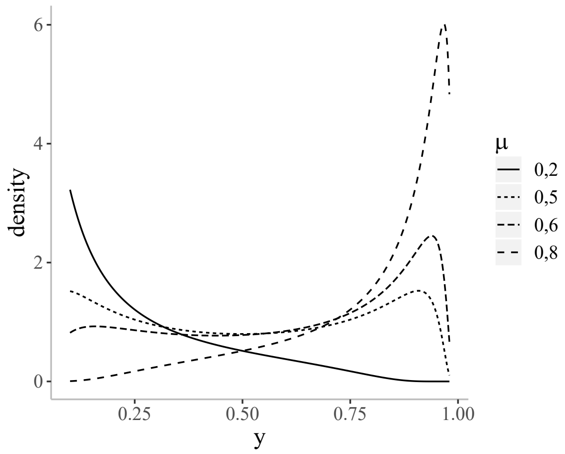
where, $(\upsilon,\varrho) \in (0,1)^2$, $\upsilon = \mathbb{P}(Y=0) + \mathbb{P}(Y=1)$, i.e, υ is the probability of an observation be zero or/and one and $\varrho = (\mathbb{P}(Y) = 1|\mathbb{I}_{\{0,1\}}(y))$, that is ϱ is the probability of the observation be equals to one given that it belongs to the discrete part. We can notice that $\mathbb{P}(Y=1) = \upsilon\varrho$, $\mathbb{P}(Y=0) = \upsilon(1-\varrho)$ and, for $0 < a < b < 1$, $\mathbb{P}(Y \in (a,b)) = (1-\upsilon)\int_a^b b_Y^*(y;\varpi)dy$. Furthermore, $b_Y^*(y;\varpi)$ can be, for example, the density of beta, simplex and BR distributions, which are the distributions adopted in this work. Below is described the generated distribution for each choice for $b_Y^*(y_i;\varpi)$:



(a) $\phi = 0,6$

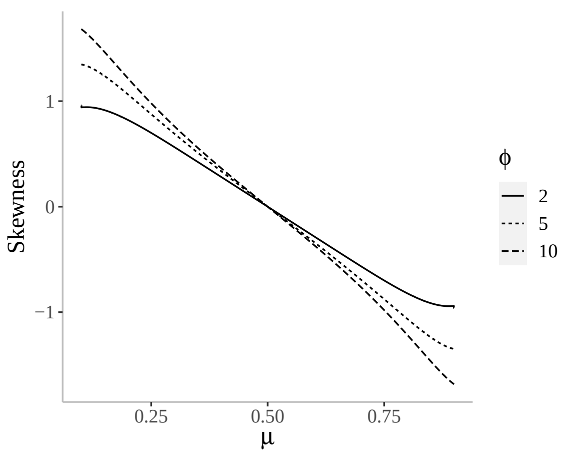


(b) $\phi = 1$

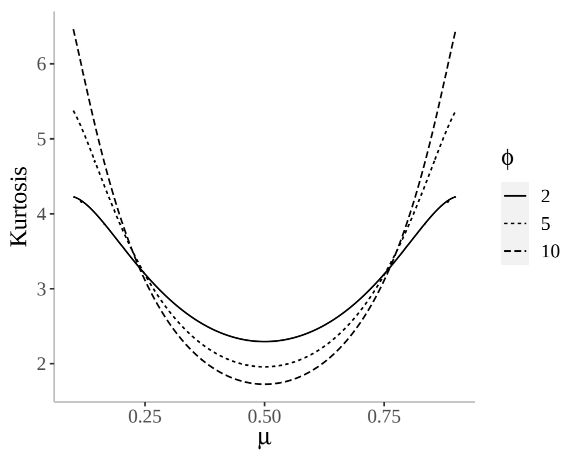


(c) $\phi = 4$

Figure 5 – Simplex densities for different values of μ, with $\phi = 0,6$ (a), $\phi = 1$ (b), $\phi = 4$ (c).



(a)



(b)

Figure 6 – Skewness (a) and Kurtosis (b) of the simplex distribution *versus* the average (μ), for different values of ϕ.

- When $b_Y^*(y_i;\varpi)$ is the beta density given in Equation (1.1), it implies that Y follows a Zero-and-One Augmented Beta distribution ($Y \sim \text{ZOAB}(\upsilon, \varrho, \mu, \phi)$) and $\varpi = (\mu, \phi)$;
- When $b_Y^*(y_i;\varpi)$ is the BR density given in Equation (1.3), it implies that Y follows a Zero-and-One Augmented Beta Rectangular distribution ($Y \sim \text{ZOABR}(\upsilon, \varrho, \mu, \phi, \alpha)$) and $\varpi = (\mu, \phi)$;
- When $b_Y^*(y_i;\varpi)$ is the simplex density given in Equation (1.6), it implies that Y follows a Zero-and-One Augmented Simplex distribution ($Y \sim \text{ZOAS}(\upsilon, \varrho, \mu, \phi)$) and $\varpi = (\mu, \phi, \alpha)$.

The respective mean for all distribution is $\mathbb{E}(Y) = \upsilon\varrho + (1-\upsilon)\mu$, whereas the variance, for each case, is given by:

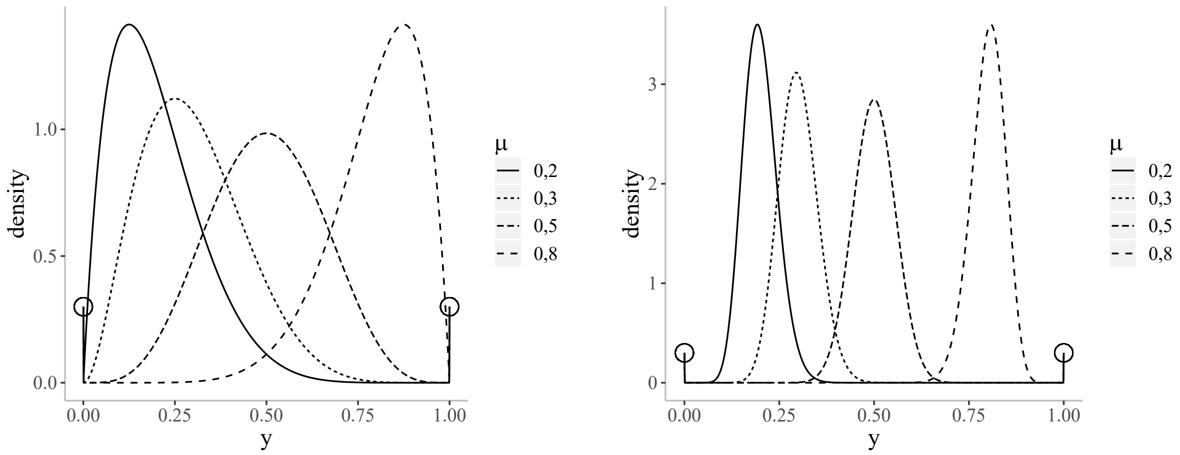
- For ZOAB:
$$\text{Var}(Y) = \upsilon\varrho(1-\varrho) + (1-\upsilon)\left[\frac{\mu(1-\mu)}{1+\phi}\right] + \upsilon(1-\upsilon)(\varrho-\mu)^2.$$
- For ZOABR:
$$\text{Var}(Y) = \upsilon V_1 + (1-\upsilon)V_2 + \upsilon(1-\upsilon)(\varrho-\mu)^2,$$
where $V_1 = \varrho(1-\varrho)$ and $V_2 = \dfrac{\varepsilon}{3} + (1-\varepsilon)\left[\dfrac{\delta(1-\delta)}{1+\phi} + \delta^2\right] - \mu^2$.
- For ZOAS:
$$\text{Var}(Y) = \upsilon\varrho(1-\varrho) + (1-\upsilon)V_3 + \upsilon(1-\upsilon)(\varrho-\mu)^2,$$
where V_3 is the variance of the simplex distribution as shown in Equation (1.8).

Following the proposal of Ospina (2008); Galvis (2014), it is possible to consider a reparameterization where the probabilities of occurrence of the discrete values are explicitly modeled. This is useful for building regression models. It consists on to define $\varrho = p_1/\upsilon$ and $\upsilon = p_0 + p_1$. Thus, we have that

$$r_Y(y; p_0, p_1, \mu, \phi) = p_0^{1-y} p_1^{y} \mathbb{I}_{\{0,1\}}(y) + (1 - p_0 - p_1) b_y(y; \mu, \phi) \mathbb{I}_{(0,1)}(y), \qquad (1.10)$$

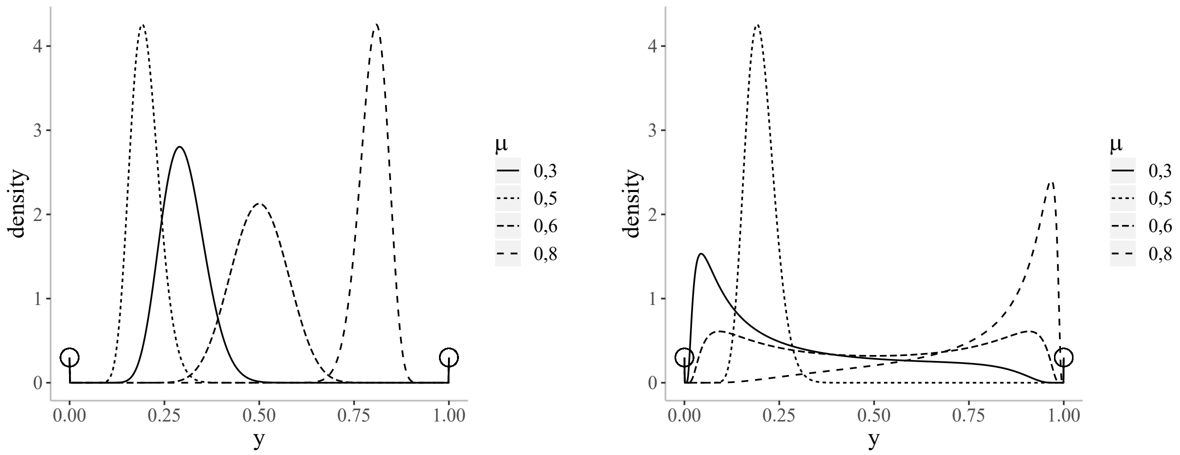
where $0 < p_0 + p_1 \leq 1$ and $b_y^*(y;\varpi)$ can be either the density of beta, simplex and BR distributions.

It is possible to see, from Figures 7, 8 and 9 depend on the values that ϕ takes, the dispersion may become smaller or higher, while μ are responsible for symmetry or asymmetry of the distributions. In addition, for ZOABR distribution (Figure 9), the higher is α, the heavier are the tails. Still observing Figure 9, we notice that in (0,1) when $\mu < 1/2$ and $\phi \leq 2$ the density presents a "J"-inverted shape, whereas, when $\mu > 1/2$ and $\phi \leq 2$ the density presents a "J" shape.



(a) $\phi = 10$ (b) $\phi = 80$

Figure 7 – Zero-and/or-One Augmented Beta density for different values of μ, for $p_0 = p_1 = 0,3$ and $\phi = 10$ (a), $\phi = 80$ (b).

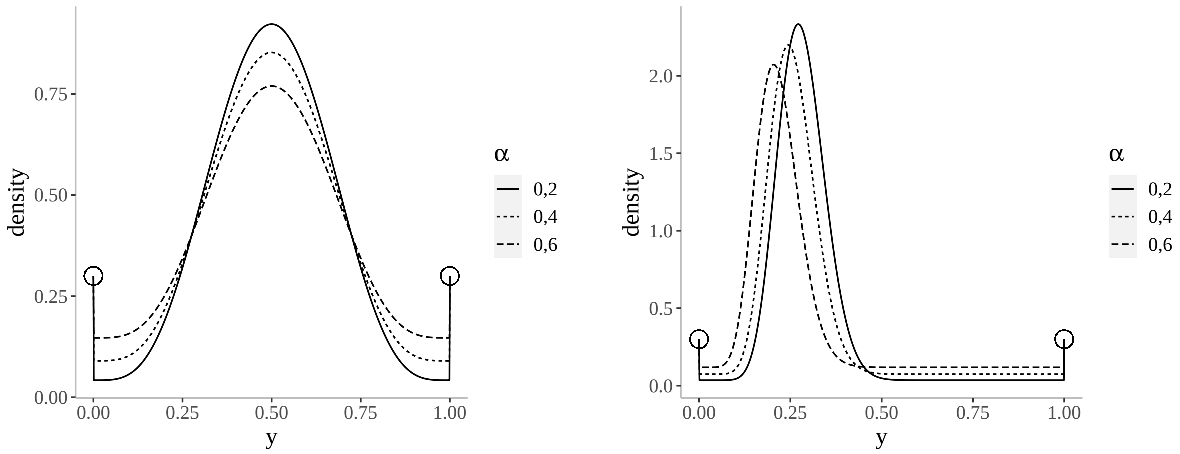


(a) $\phi = 0,6$ (b) $\phi = 4$

Figure 8 – Zero-and/or-One Augmented Simplex density for different values of μ (indicated in the graphs), with $p_0 = p_1 = 0,3$ for both graphs, besides $\phi = 0,6$ (a), $\phi = 4$ (b).

1.2 Basic concepts of Semi-parametric Additive Models

In this section we present some concepts of Semi-parametric Additive Models (SPAM), which will be necessary for the next sections. The SPAM (also called Generalized Additive Partial Linear Models, GAPLM) are an extension of the Generalized Additive Models (GAM), both proposed by Hastie and Tibshirani (1990). While the GAM has only a non-parametric regression structure, the SPAM is a mixing between the parametric and non-parametric approaches. These models and their extensions have been explored over the years by several authors as, among others, Buja et al. (1989), Hastie and Tibshirani (1987), Hastie and Tibshirani (1990), Linton and Härdle (1996), Guisan et al. (2002), Rigby and Stasinopoulos (2005), Wood (2008), Wang et al. (2011), Wood (2017).



(a) $p_0 = p_1 = 0,3$, $\phi = 10$ and $\mu = 0,5$ (b) $p_0 = p_1 = 0,3$, $\phi = 50$ and $\mu = 0,3$

Figure 9 – Augmented Beta Rectangular in zero and one density for different values of α (indicated in the graphs), with $p_0 = p_1 = 0,3$ for all graphs, besides $\phi = 10$ and $\mu = 0,5$ (a) and $\phi = 50$ and $\mu = 0,3$ (b).

The SPAM has some advantages compared to models with fully parametric predictors, since they allow to model a more diversity of relations between the response and each covariate. Furthermore, they have two particular cases, namely, parametric and non-parametric models.

Within the methodology of additive models, it is usual to approximate the non-linear component(s) by a linear combination of basis functions, such as: Fourier expansion, wavelets, splines as B-splines and natural cubic splines, see, for example, Hastie and Tibshirani (1990), Kooperberg and Stone (1991), Vidakovic (2009), Silverman (2018). In this work we focus on the use of splines due to its simple structure and good approximation properties, which definition is presented in next section.

1.2.1 Splines

Wegman and Wright (1983) published a review article about the use of splines in Statistics. Specifically, they present them as a non-parametric estimation technique for regression analysis. Under this methodology, some functions can be approximated by a sum of basis functions, which present great flexibility and are able to capture local behavior changes. These basis functions will be chosen, here, as the splines, where the main underlying idea is an approximation by polynomial functions.

Then, let us suppose a function of interest, say f, limited by the interval $[a, b]$, i.e., f: $\mathbb{R} \to [a, b]$. This interval is, then partitioned in k^* subintervals, say, $[\epsilon_{t-1}, \epsilon_t], 1 \leq t \leq k^*$, where $a < \epsilon_0 < \cdots < \epsilon_{k^*} < b$. Then, a polynomial, say p_t, is used to approximate the function over each (sub)interval $[\epsilon_{t-1}, \epsilon_t], t = 1, \ldots, k^*$. This procedure leads to a piecewise polynomial approximation function $s(.)$, i.e., $s(x) = p_t(x)$ in $[\epsilon_{t-1}, \epsilon_t], t = 1, \ldots, k^*$. The values $\epsilon_0, \dots, \epsilon_{k^*}$ are called knots, being ϵ_0 and ϵ_{k^*} the external knots, whereas the others, $\epsilon_1, \dots, \epsilon_{k^*-1}$, are the so-called interval knots.

In the general case, the parts of polynomial $p_t(x)$ are chosen independently of each other and, then, they do not form a continuous function, say $s(x)$, in $[a, b]$. Therefore, to obtain a smooth approximation for a function, it is necessary that the parts of the polynomial be joined smoothly in the internal knots $\epsilon_1, \dots, \epsilon_{k^*-1}$ and to have all derivatives up to a certain order, coincide at knots. Thus, to obtain a smooth piecewise polynomial function, called spline, some restrictions should be imposed in the general definition of spline (Cunha, 2003).

Definition 1. *The function s is called an $d-$order (order=degree+1) spline, with knots $\{\epsilon_t\}_{t=1}^{k^*}$, associated with a partition of $[a, b]$, if s:*

- *is a polynomial with $d-1$ degree in each subinterval $[\epsilon_{t-1}, \epsilon_t]$;*
- *has $d-2$ continuous derivatives in each ϵ_t and, thus, in $[a, b]$;*
- *has the following form:*

$$s(x) = \sum_{l_1=0}^{d-1} c_{l_1} x^{l_1} + \sum_{l_2=1}^{k^*-1} a_{l_2}(x - \epsilon_{l_2})_+^{d-1},$$

where $c_0, \dots, c_{d-1} \in \mathbb{R}$, $a_1, \dots, a_{k-1} \in \mathbb{R}$ and, $(x - \epsilon_{l_2})_+^{d-1}$ is a $d-1$ degree truncated power function, which is defined as,

$$(x - \epsilon_{l_2})_+^{d-1} = \begin{cases} (x - \epsilon_{l_2})^{d-1}, \ if \ (x - \epsilon_{l_2}) \geq 0, \\ 0, \ if \ u < 0. \end{cases}$$

Therefore, it can be concluded that any spline function is a linear combination of $d + k$ basis functions. For more details about splines, see Wahba (1990) and Wood (2017).

In the next subsections, the B-splines and the P-splines, which will be used in our regression models, are presented.

B-splines

The B-splines proposed by De Boor (1978), are built using the so-called B-splines basis. To define q B-spline basis, following Wood (2017), first let us assume a total of $k^* = q + d + 1$ knots. Then the d-th order spline can be represented as $s(x) = \sum_{l=1}^{q} B_{l,d}\gamma_l$, where, γ_l is an (unknown) parameter and the B-spline basis functions are recursively defined as follows:

$$B_{l,d} = B_{l,d}(x) = \frac{x - \epsilon_l}{\epsilon_{l+d-1}(x) - \epsilon_t(x)} B_{l,d-1}(x) + \frac{\epsilon_{l+d} - x}{\epsilon_{l+d} - \epsilon_{l+1}} B_{l+1,d-1}(x),$$

where

$$B_{l,d}(x) = \begin{cases} 1, & if\ \epsilon_l \leq x \leq \epsilon_{l+1}, \\ 0, & \text{otherwise} \end{cases}$$

$d \leq 1$ is the order of the B-spline, $x \in [a, b]$, $a < \epsilon_1, \ldots, \epsilon_{k^*} < b$ and the knots are mutually equidistant.

P-splines

The P-splines were proposed by Eilers and Marx (1996) and they are a combination of B-splines and a difference penalty function of order d, namely Δ^d, applied to adjacent estimated coefficients $\boldsymbol{\gamma}_j$ of B-splines. The differences $\Delta^d \boldsymbol{\gamma}_j$ are defined as

$$\Delta\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_j - \boldsymbol{\gamma}_{j-1}; \Delta^2\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_j - 2\boldsymbol{\gamma}_{j-1} + \boldsymbol{\gamma}_{j-2}; \Delta^d\boldsymbol{\gamma}_j = \Delta\left(\Delta^{d-1}\boldsymbol{\gamma}_j\right), \text{ with } j = 1, \ldots, k,$$

where d is the difference orders. For details about the choice of the number and the position of knots, see for example Eilers and Marx (1996) and Ruppert (2002)

1.2.2 Penalization criterion

As seen in the previous section, each P-spline $s(x)$ depend on some parameters $\boldsymbol{\gamma}$. In order to obtain the fit of $s(x)$, these parameters have to be estimated through some estimation process with the addition of a penalty function and an associated parameter, say λ, called smoothness parameter, in order to enables a higher control over the spline smoothing, avoiding both an overfitting and a high smoothness of the fitted function (Green, 1987; Pulgar, 2009). For example, if $s(x)$ is an approximation for a regression function and the estimation process considering is the maximum likelihood method, the penalty function is adding in the log-likelihood to control the smoothness of the fitted regression function.

Then, the penalty function related to a given P-spline is given by

$$\boldsymbol{\gamma}\boldsymbol{\Lambda}^d\boldsymbol{\gamma}, \tag{1.11}$$

where $\boldsymbol{\gamma}$ are the coefficients of the related B-splines, $\boldsymbol{\Lambda}^d = (\Delta^d)^\top(\Delta^d)$ and Δ^d is the d order difference operator, for $d = 1, 2, \ldots$. For example, for $d = 1$, the penalized function is defined as

$$\boldsymbol{\gamma}^\top\boldsymbol{\Delta}^\top\boldsymbol{\Delta}\boldsymbol{\gamma} = (\gamma_2 - \gamma_1)^2 + (\gamma_3 - \gamma_2)^2 + \cdots + (\gamma_k - \gamma_{k-1})^2,$$

where the difference operator of dimension $k \times k$ is given by

$$\boldsymbol{\Delta} = \begin{bmatrix} -1 & 1 & 0 & \ldots \\ 0 & -1 & 1 & \ldots \\ 0 & 0 & -1 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad \text{and} \quad \boldsymbol{\gamma}^\top\boldsymbol{\Delta}^\top\boldsymbol{\Delta}\boldsymbol{\gamma} = \boldsymbol{\gamma}^\top \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots \\ -1 & 2 & -1 & 0 & \ldots \\ 0 & -1 & 2 & -1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \boldsymbol{\gamma}.$$

Adopting $d = 2$, the penalization is given by

$$\boldsymbol{\gamma}^\top(\boldsymbol{\Delta}^2)^\top(\boldsymbol{\Delta}^2)\boldsymbol{\gamma} = (\gamma_1 - 2\gamma_2 - \gamma_3)^2 + (\gamma_2 - 2\gamma_3 - \gamma_4)^2 + \cdots + (\gamma_{k-2} - 2\gamma_{k-1} - \gamma_k)^2,$$

and the difference operator, $\boldsymbol{\Delta}^2$, can be written as

$$\boldsymbol{\Delta}^2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots \\ 0 & 1 & -2 & 1 & \dots \\ 0 & 0 & 1 & -2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

For more details about penalties criterion, see for example Eilers and Marx (2010) and Eilers et al. (2015).

2 Semi-parametric Additive Models for independent and limited data

In this Chapter we develop the Semi-parametric Additive Models for independent and bounded data. The proposal models will be the Semi-parametric Additive Beta (Beta-SPAM), Semi-parametric Additive Simplex (Simplex-SPAM) and Semi-parametric Additive Beta Rectangular (BR-SPAM) models.

2.1 Models

Let us consider a set of random independent variables, $Y_1, \ldots, Y_n$, such that $Y_i \overset{ind}{\sim} \text{beta}(\mu_i, \phi_i),\ \text{simplex}(\mu_i, \phi_i)$ or $\text{BR}(\mu_i, \phi_i, \alpha), i = 1, \ldots, n$. For any of these distributions, the regression models Beta-SPAM, Simplex-SPAM and BR-SPAM have the following systematic component

$$\boldsymbol{\eta}_1 = g_1(\boldsymbol{\mu}) = \boldsymbol{X\beta} + \sum_{j=1}^{k} h_j(\boldsymbol{z}_j),\ \boldsymbol{\eta}_2 = g_2(\boldsymbol{\phi}) = \boldsymbol{E\kappa}, \tag{2.1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^\top$, $g_1(.)$ and $g_2(.)$ are strictly monotone and twice differentiable link functions, such that, $g_1 : (0,1) \to \mathbb{R}$, $g_2 : \mathbb{R}^+ \to \mathbb{R}$, $\boldsymbol{\eta}_1 = (g_1(\mu_1), \ldots, g_1(\mu_n))^\top$, $\boldsymbol{\eta}_2 = (g_2(\phi_1), \ldots, g_2(\phi_n))^\top$. Also, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)^\top \in \mathbb{R}^s$ are, respectively, a p- and a s-vectors of unknown parameters, $\boldsymbol{X}_{(n\times p)} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ and $\boldsymbol{E}_{(n\times s)} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)^\top$ are (known) design matrices, respectively, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$, $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{is})^\top$ for $i = 1, \ldots, n$, $\boldsymbol{z}_j = (z_{1j}, \ldots, z_{nj}), j = 1, \ldots, k$ is a vector (known) of continuous explanatory variables associated with non-parametric components and $h_1, \ldots, h_k$ are an unknown smooth functions.

Additionally, as discussed in Section 1.2.1, we have that

$$h_j(\boldsymbol{z}_j) = \sum_{l=1}^{q_j} \gamma_{lj} b_{lj}(\boldsymbol{z}_j), \tag{2.2}$$

where γ_{lj} are the coefficients to be estimated, b_{lj} are the l-th cubic B-spline related to the j-th P-spline, evaluated at $\boldsymbol{z}_j$, $j = 1, \ldots, k$ and $l = 1, \ldots, q_j$. Since we adopted cubic B-splines, we set $b_{lj} = B_{l,d_l}$ with $d_l = 4$, as defined in Section 1.2.1, $k_j = d_l + q_j$ is the number of knots. Furthermore, it is possible to rewrite Equation (2.1), for $i = 1, \ldots, n$, as follows,

$$\eta_{1i} = g_1(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{b}_{i1}^\top \boldsymbol{\gamma}_1 + \cdots + \boldsymbol{b}_{ik}^\top \boldsymbol{\gamma}_k,$$

where $\boldsymbol{b}_{ij} = (b_{1j}(z_{ij}), \ldots, b_{q_j j}(z_{ij}))^\top$ and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{jq_j})^\top$. Therefore, in a vectorial form, we have that

$$\boldsymbol{\eta}_1 = g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j\boldsymbol{\gamma}_j,\ \boldsymbol{\eta}_2 = g_2(\boldsymbol{\phi}) = \boldsymbol{E}\boldsymbol{\kappa}, \tag{2.3}$$

where $\boldsymbol{B}_j$ is a $n \times q_j$ matrix, defined as $\boldsymbol{B_j} = (\boldsymbol{b}_{1j}, \ldots, \boldsymbol{b}_{nj})^\top$ $j = 1, \ldots, k$.

2.2 Penalized log-likelihood function

The respective log-likelihood, with systematic components given in Equation (2.3), and defining $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ for the Simplex-SPAM and the Beta-SPAM and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$ for the BR-SPAM is given by $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$, where, $l_i(\boldsymbol{\theta})$, for each one the three models, is defined as:

- Beta-SPAM,

$$\begin{aligned} l_i(\boldsymbol{\theta}) = {} & \log(\Gamma(\phi_i)) - \log(\Gamma(\mu_i\phi_i)) - \log(\Gamma([1-\mu_i]\phi_i)) + [\mu_i\phi_i - 1]\log(y_i) \\ & + ([1-\mu_i]\phi_i - 1)\log[1-y_i], \end{aligned}$$

- Simplex-SPAM,

$$l_i(\boldsymbol{\theta}) = -\frac{1}{2}\left\{\log(2\pi\phi_i) + 3\log[y_i(1-y_i)]\right\} - \frac{1}{2\phi_i}d(y_i;\mu_i),$$

  with $d(y_i;\mu_i)$ (see Equation (1.7)), in this case, defined as,

$$d(y_i;\mu_i) = \frac{(y_i-\mu_i)^2}{2\phi_i y_i(1-y_i)\mu_i^2(1-\mu_i)^2}.$$

- BR-SPAM,

$$l_i(\boldsymbol{\theta}) = \log[\varepsilon_i + (1-\varepsilon_i)b_{Y_i}(y_i;\delta_i,\phi_i)], \tag{2.4}$$

  with $\varepsilon_i = 1 - \sqrt{1 - 4\alpha\mu_i(1-\mu_i)}$, $b_Y(y;.,.)$ is the beta density defined in Equation (1.1), $\delta_i = \dfrac{\mu_i - \frac{1}{2} + \frac{1}{2}\sqrt{1-4\alpha\mu_i(1-\mu_i)}}{\sqrt{1-4\alpha\mu_i(1-\mu_i)}}$,

and for all of them, $\mu_i = g_1^{-1}(\eta_{1i})$, $\eta_{1i} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{k}\boldsymbol{b}_{ij}^\top\boldsymbol{\gamma}_j$, $\phi_i = g_2^{-1}(\eta_{2i})$ and $\eta_{2i} = \boldsymbol{e}_i^\top\boldsymbol{\kappa}$.

As discussed in the Section 1.2.2, a penalty function has to be added to the log-likelihood to guarantee, in the estimation process, the smoothness of the fitted (non-parametric) curve (see Wood (2017) and Vanegas and Paula (2016), for example). Since we adopted the P-splines, the penalty function based on differences, defined in the Equation (1.11), will be used. Hence, the penalized log-likelihood is given by

$$l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j, \tag{2.5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^\top$ is the vector of smooth parameters and $\lambda_j > 0,\ \forall j, j = 1, \ldots, k$. These parameters control the smoothness of the P-splines that will approximate the non-parametric component $h_j(.)$, such that, the larger the value of λ_j, the smoother the fitted curve will be. For ease of the notation, we will drop the superscript d of $\boldsymbol{\Lambda}_j^d$ in the follow sections, however, we emphasize that this penalty continues depending on the order d.

2.2.1 Complete log-likelihood the for BR-SPAM

From Equation (2.4) we can see that the log-likelihood of the BR-SPAM is not easily tractable (see also, Silva et al. (2020)). In order to make calculations simpler it will be used the EM algorithm (Dempster et al., 1977) to obtain the maximum likelihood estimates.

In Section 2.1, we have that if $Y_i \sim \text{BR}(\mu_i, \phi_i, \alpha)$, then, the following structure is valid:

$$v_i = \begin{cases} 0, & \text{if } Y_i \sim \text{beta}(\delta_i, \phi_i), \text{ with probability } 1 - \varepsilon_i \\ 1, & \text{if } Y_i \sim \text{uniform}(0, 1), \text{ with probability } \varepsilon_i. \end{cases}$$

where, given that the beta rectangular distribution is a mixing of a beta and a standard uniform distributions, then V_i is a non-observed variable that works as an indicator function of which distribution the data comes from. That is, $V_i \overset{ind}{\sim} \text{Bernoulli}(\varepsilon_i)$, where $\varepsilon_i = 1 - \sqrt{1 - 4\alpha\mu_i(1 - \mu_i)}$. The jointly distribution of $\boldsymbol{Y}_i^c = (Y_i, V_i)^\top$ given $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$, is given by

$$f(\boldsymbol{y}_i^c|\boldsymbol{\theta}) = \varepsilon_i^{v_i}(1 - \varepsilon_i)^{1-v_i} b(y_i; \delta_i, \phi_i)^{1-v_i} \mathbb{I}_{(0,1)}(y_i) \mathbb{I}_{\{0,1\}}(v_i), \tag{2.6}$$

where ε_i, δ_i, μ_i and ϕ_i are defined in Equation (2.4).

For the BR-SPAM, the complete likelihood is given by $L^c(\boldsymbol{\theta}; \boldsymbol{y}^c) = \prod_{i=1}^{n} f(\boldsymbol{y}_i^c|\boldsymbol{\theta})$, where $f(\boldsymbol{y}_i^c|\boldsymbol{\theta})$ is defined in Equation (2.6). Also, the complete associated log-likelihood is given by

$$l^c(\boldsymbol{\theta}; \boldsymbol{y}^c) = \sum_{i=1}^{n} l_i^*(\boldsymbol{\theta}), \tag{2.7}$$

where $l_i^*(\boldsymbol{\theta}) = v_i \log \varepsilon_i + (1 - v_i) \log(1 - \varepsilon_i) + (1 - v_i)[\log(\Gamma(\phi_i)) - \log(\Gamma(\delta_i\phi_i)) - \log(\Gamma((1 - \delta_i)\phi_i)) + (\delta_i\phi_i - 1) \log y_i + ((1 - \delta_i)\phi_i - 1) \log(1 - y_i)]$. Finally, adding a penalty term with respect to the non-parametric component, to control the smoothness, we have that

$$l_p^c(\boldsymbol{\theta}) = l^c(\boldsymbol{\theta}; \boldsymbol{y}^c) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j. \tag{2.8}$$

Then, each iteration of EM algorithm involves two steps, an E step (expectation) and a M step (maximization), as follows:

- E step: To compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \mathbb{E}[l_p^c(\boldsymbol{\theta}|\boldsymbol{y}^c)|\boldsymbol{y}, \boldsymbol{\theta}^{(m)}]$, where the expectation is taken with respect to the conditional distribution: $f(\boldsymbol{y}|\boldsymbol{v}, \boldsymbol{\theta}^{(m)})$;
- M step: To obtain $\boldsymbol{\theta}^{(m+1)} = \arg\max_{\theta\in\Theta} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$.

Then, let $\boldsymbol{\theta}^{\top(m)} = (\hat{\boldsymbol{\beta}}^{\top(m)}, \hat{\boldsymbol{\gamma}}_1^{\top(m)}, \ldots, \hat{\boldsymbol{\gamma}}_k^{\top(m)}, \hat{\boldsymbol{\kappa}}^{\top(m)}, \hat{\alpha}^{(m)})$ be the estimate of $\boldsymbol{\theta}$ in the m-th iteration of the EM algorithm. The conditional expectation with respect to $\boldsymbol{v}$ given $\boldsymbol{y}$, of the penalized complete log-likelihood defined in Equation (2.8) is given by

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}) = \mathbb{E}[l_p^c(\boldsymbol{\theta})|\boldsymbol{y}, \hat{\boldsymbol{\theta}}^{(m)}] &= \sum_{i=1}^{n}\{\hat{v}_i^{(m)} \log \varepsilon_i + (1-\hat{v}_i^{(m)}) \log(1-\varepsilon_i) \\ &+ (1-\hat{v}_i^{(m)})[\log(\Gamma(\phi_i)) - \log(\Gamma(\delta_i\phi_i)) - \log(\Gamma((1-\delta_i)\phi_i)) \\ &+ (\delta_i\phi_i - 1)\log y_i + ((1-\delta_i)\phi_i - 1)\log(1-y_i)]\} - \sum_{j=1}^{k} \frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j \\ &= \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}), \end{aligned} \tag{2.9}$$

where

$$\hat{v}_i = \mathbb{E}(V_i|y_i, \hat{\boldsymbol{\theta}}^{(m)}) = \mathbb{P}(V_i = 1|y_i, \hat{\boldsymbol{\theta}}^{(m)}) = \left(\frac{\varepsilon_i}{\varepsilon_i + (1-\varepsilon_i)b(y_i; \delta_i, \phi_i)}\right), \tag{2.10}$$

$$\varepsilon_i = 1 - \sqrt{1 - 4\alpha\mu_i(1-\mu_i)} \text{ and } \delta_i = \frac{\mu_i - \frac{1}{2} + \frac{1}{2}\sqrt{1-4\alpha\mu_i(1-\mu_i)}}{\sqrt{1-4\alpha\mu_i(1-\mu_i)}}.$$

2.3 Penalized score and gradient functions

Under Equation (2.5) for the Beta-SPAM and Simplex-SPAM, the penalized score function, say $\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta})$, for $\boldsymbol{\beta}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k$ and $\boldsymbol{\kappa}$ is given by $\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta}) = (\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top)^\top$. On the other hand, for BR-SPAM, the conditional expectation of $l_p^c(\boldsymbol{\theta})$ (Equation (2.8)) will be used to obtain the penalized gradient function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$, defined as $\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta}) = (\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top, \dot{U}_p^{\alpha}(\boldsymbol{\theta}))^\top$.

2.3.1 Beta-SPAM

For the Beta-SPAM, we have that

$$\begin{aligned} \dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*, \dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j} = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j, \\ \dot{\boldsymbol{U}}_p^{\kappa}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a}, \end{aligned} \tag{2.11}$$

where $j = 1, \dots, k$, $\boldsymbol{T}_1 = \text{diag}\{(\partial g_1(\mu_1)/\partial\mu_1)^{-1}, \dots, (\partial g_1(\mu_n)/\partial\mu_n)^{-1}\}$, $\boldsymbol{f}^* = (f_i^*, \dots, f_n^*)$, $f_i^* = \phi_i(y_i^* - \mu_i^*)$, $\mu_i^* = \Psi(\mu_i\phi_i) - \Psi[(1-\mu_i)\phi_i]$, $y_i^* = \log[y_i/(1-y_i)]$, $i = 1, \dots, n$, with $\Psi(.)$ is the digamma function, $\Psi(z) = \partial \log\Gamma(z)/\partial z$, $z > 0$, $\boldsymbol{\Phi} = \text{diag}\{\phi_1, \dots, \phi_n\}$, $\boldsymbol{a} = (a_1, \dots, a_n)^\top$, $a_i = \mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \Psi[(1-\mu_i)\phi_i] + \Psi(\phi_i)$, and $\boldsymbol{T}_2 = \text{diag}\{(\partial g_2(\phi_1)/\partial\phi_1)^{-1}, \dots, (\partial g_2(\phi_n)/\partial\phi_n)^{-1}\}$. The related demonstrations can be found in Appendix A.1.1.

2.3.2 Simplex-SPAM

The respective elements of the score vector for the Simplex-SPAM are given by (see the Appendix A.2.1 for the demonstration):

$$\begin{aligned} \dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*, \dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j} = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j, \\ \dot{\boldsymbol{U}}_p^{\kappa}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a}, \end{aligned} \tag{2.12}$$

where $j = 1, \dots, k$, $\boldsymbol{T}_1 = \text{diag}\{(\partial g_1(\mu_1)/\partial\mu_1)^{-1}, \dots, (\partial g_1(\mu_n)/\partial\mu_n)^{-1}\}$, $\boldsymbol{a} = (a_1, \dots, a_n)^\top$, $a_i = -\dfrac{1}{2\phi_i} + \dfrac{1}{2\phi_i^2} d(y_i; \mu_i)$, $d(y;\mu)$ is defined in Equation (1.7),
$\boldsymbol{T}_2 = \text{diag}\{(\partial g_2(\phi_1)/\partial\phi_1)^{-1}, \dots, (\partial g_2(\phi_n)/\partial\phi_n)^{-1}\}$, $\boldsymbol{f}^* = (f_1^*, \dots, f_n^*)$, $f_i^* = (1/\phi_i)u_i(y_i - \mu_i)$ and $u_i = \dfrac{1}{\mu_i(1-\mu_i)}\left[d(y_i;\mu_i) + \dfrac{1}{\mu_i^2(1-\mu_i)^2}\right]$, $i = 1, \dots, n$.

2.3.3 BR-SPAM

Based on Equation (2.8), $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \mathbb{E}[l_p^c(\boldsymbol{\theta})|\boldsymbol{y}, \boldsymbol{\theta}^{(m)}]$ will be used to obtain the respective penalized gradient vector. Furthermore, the penalized gradient vector is given by

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}} = \left(\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \dots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\kappa}(\boldsymbol{\theta})^\top, \dot{U}_p^{\alpha}(\boldsymbol{\theta})\right)^\top,$$

where

$$\begin{aligned} &\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*, \ \dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j, \\ &\dot{\boldsymbol{U}}_p^{\kappa}(\boldsymbol{\theta}) = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a}, \ \dot{U}_p^{\alpha}(\boldsymbol{\theta}) = \text{trace}(\boldsymbol{D}), \end{aligned} \tag{2.13}$$

$j = 1, \ldots, k$, $\boldsymbol{T}_1 = \text{diag}\{(\partial g_1(\mu_1)/\partial \mu_1)^{-1}, \ldots, (\partial g_1(\mu_n)/\partial \mu_n)^{-1}\}$, $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, with $p_i = (1 - \hat{v}_i)\{\delta_i(y_i^* - \delta_i^*) + \log(1 - y_i) - \Psi[(1 - \delta_i)\phi_i] + \Psi(\phi_i)\}$, $\delta_i^* = \Psi(\delta_i \phi_i) - \Psi[(1 - \delta_i)\phi_i]$, $y_i^* = \log[y_i/(1 - y_i)]$, for $i = 1, \ldots, n$. Also, $\boldsymbol{f}^* = (f_1^*, \ldots, f_n^*)^\top$, where

$$f_i^* = \frac{2\alpha(1 - 2\mu_i)}{1 - \varepsilon_i}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1 - \hat{v}_i)}{(1 - \varepsilon_i)}\right] + (1 - \hat{v}_i)(1 - \alpha)\frac{\phi_i}{(1 - \varepsilon_i)^3}(y_i^* - \delta_i^*),$$

$\boldsymbol{D} = \text{diag}\{d_1, \ldots, d_n\}$,

$$d_i = \frac{2\mu_i(1 - \mu_i)}{(1 - \varepsilon_i)}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1 - \hat{v}_i)}{(1 - \varepsilon_i)}\right] - (1 - \hat{v}_i)\left[\frac{\phi_i \mu_i(1 - \mu_i)(1 - 2\mu_i)}{(1 - \varepsilon_i)^3}\right](y_i^* - \delta_i^*),$$

$i = 1, \ldots, n$ and $\boldsymbol{T}_2 = \text{diag}\{(\partial g_2(\phi_1)/\partial \phi_1)^{-1}, \ldots, (\partial g_2(\phi_n)/\partial \phi_n)^{-1}\}$. The related demonstration can be found in Appendix A.3.1.

2.4 Penalized Hessian and Fisher information matrices

The necessary matrices of second derivatives realted to $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$, for Beta-SPAM and Simplex-SPAM, and related to $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ for Beta-SPAM, with respect to $\boldsymbol{\theta}$, are discussed in this section. For the two first models, we have that

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) \end{pmatrix}, \tag{2.14}$$

where $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_j}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_j}(\boldsymbol{\theta})^\top$ and $\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_{j'}\gamma_j}(\boldsymbol{\theta})^\top$, $j = 1, \ldots, k$. On the other hand, for the third model we have

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\alpha}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\alpha}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\alpha}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\alpha}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\alpha\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\alpha\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\alpha\alpha}(\boldsymbol{\theta}) \end{pmatrix}, \tag{2.15}$$

where $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_j}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_j}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_{j'}\gamma_j}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\alpha}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\alpha}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\alpha\gamma_j}(\boldsymbol{\theta})^\top$ and $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\alpha}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top$, $j = 1, \ldots, k$.

Finally, the Fisher information for the Beta-SPAM and the Simplex-SPAM is given by

$$\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right] = \begin{pmatrix} \boldsymbol{K}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{K}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) \end{pmatrix}, \tag{2.16}$$

where $\boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\gamma_j\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\boldsymbol{K}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta})^\top$, $\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\gamma_{j'}\gamma_j}(\boldsymbol{\theta})^\top$ and $\boldsymbol{K}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_j}(\boldsymbol{\theta})^\top$, $j = 1, \dots, k$.

2.4.1 Beta-SPAM

The elements of the Hessian matrix (Equation 2.14), are given by (see Appendix A.1.2 for the demonstration of the results),

$$\begin{aligned} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{X}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j, \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{E}^\top\boldsymbol{S}\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}^*\boldsymbol{E}, \\ \ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}^*\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_{j'}, \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{B}_j, \end{aligned} \tag{2.17}$$

$j = 1, \dots, k$, $\boldsymbol{Q} = \text{diag}\{q_1, \dots, q_n\}$,

$$q_i = \phi_i\left[\phi_i\{\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)\} + (y_i^* - \mu_i^*)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\right]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2},$$

$\boldsymbol{Q}^* = \text{diag}\{q_1^*, \dots, q_n^*\}$,

$$q_i^* = \{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)] - (y_i^* - \mu_i^*)\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1},$$

$\boldsymbol{S} = \text{diag}\{s_1, \dots, s_n\}$, and

$$\begin{aligned} s_i = &\left[\Psi'(\mu_i\phi_i)\mu_i^2 + \Psi'((1-\mu_i)\phi_i)(1-\mu_i)^2 - \Psi'(\phi_i) + a_i\left(\frac{\partial g_2(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial^2 g_2(\mu_i)}{\partial\mu_i^2}\right)\right] \\ &\times\left(\frac{\partial g_2(\mu_i)}{\partial\mu_i}\right)^{-2}, \end{aligned}$$

where $\Psi'(.)$ is the trigamma function, $\Psi'(z) = \partial^2 \log\Gamma(z)/\partial z^2$, for $z > 0$ and $a_i = \mu_i(y_i^* - \mu_i^*) + \log(1-y_i) - \Psi[(1-\mu_i)\phi_i] + \Psi(\phi_i)$, $i = 1, \ldots, n$.

On the other hand, the elements of the Fisher information (Equation (2.16)) are given by (see Appendix A.1.2 for the demonstration of the results).

$$\begin{aligned}
&\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}, \quad \boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{B}_j,\\
&\boldsymbol{K}_p^{\kappa\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{E}^\top\boldsymbol{P}\boldsymbol{E}, \quad \boldsymbol{K}_p^{\beta\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}^*\boldsymbol{E},\\
&\boldsymbol{K}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}^*\boldsymbol{E}, \quad \boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}\boldsymbol{B}_{j'},\\
&\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}\boldsymbol{B}_j + \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned} \tag{2.18}$$

where $\boldsymbol{W} = \mathrm{diag}\{w_1, \ldots, w_n\}$,

$$w_i = \phi_i^2[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2},$$

$\boldsymbol{P} = (p_1, \ldots, p_n)^\top$, and

$$p_i = \left[\Psi'(\mu_i\phi_i)\mu_i^2 + \Psi'((1-\mu_i)\phi_i)(1-\mu_i)^2 - \Psi'(\phi_i)\right]\left(\frac{\partial g_2(\mu_i)}{\partial\mu_i}\right)^{-2}, \ i = 1, \ldots, n.$$

Also, $\boldsymbol{W}^* = \mathrm{diag}\{w_1^*, \ldots, w_n^*\}$, where

$$w_i^* = \{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)]\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}, \ i = 1, \ldots, n.$$

2.4.2 Simplex-SPAM

The elements of the respective Hessian matrix (Equation 2.14) are given by (see Appendix A.2.2 for the respective demonstration),

$$\begin{aligned}
&\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{X}, \quad \ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{B}_j,\\
&\ddot{\boldsymbol{U}}_p^{\kappa\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{E}^\top\boldsymbol{S}\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\beta\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}^*\boldsymbol{E},\\
&\ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}^*\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_{j'},\\
&\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned} \tag{2.19}$$

where $j = 1, \dots, k$, $\boldsymbol{Q}^* = (q_1^*, \dots, q_n^*)^\top$,

$$q_i^* = [(1/\phi_i^2)u_i(y_i - \mu_i)] \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-1}$$

and u_i is defined in Equation (2.12). Also, $\boldsymbol{S} = \text{diag}\{s_1, \dots, s_n\}$, where

$$s_i = \left[\frac{1}{2\phi_i^2} - \frac{1}{\phi_i^3} d(y_i;\mu_i) + a_i \left(\frac{\partial^2 g_2(\phi_i)}{\partial \phi_i^2}\right)\left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-1}\right]\left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-2}, \ i = 1, \dots, n,$$

$d(y;\mu)$ is defined in Equation (1.7) and a_i is defined in Equation (2.12). Furthermore, $\boldsymbol{Q} = \text{diag}\{q_1, \dots, q_n\}$, where

$$q_i = \frac{1}{\phi_i}\left\{u_i + (y_i - \mu_i)\left[\frac{2(y_i-\mu_i)u_i}{\mu_i(1-\mu_i)} + \frac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4} + \frac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right] + u_i(y_i - \mu_i) \right.$$
$$\left. \times \left(\frac{\partial^2 g_1(\mu_i)}{\partial \mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1}\right\}\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-2}.$$

On the other hand, the elements of the Fisher information (Equation (2.16)) are given by (see Appendix A.2.2 for the respective demonstrations).

$$\begin{aligned}
\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{X}, \quad \boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{B}_j,\\
\boldsymbol{K}_p^{\kappa\kappa}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{E}^\top\boldsymbol{P}\boldsymbol{E}, \quad \boldsymbol{K}_p^{\beta\phi}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\phi}\right] = \boldsymbol{0},\\
\boldsymbol{K}_p^{\gamma_j\phi}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\phi}\right] = \boldsymbol{0}, \quad \boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}\boldsymbol{B}_{j'},\\
\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}\boldsymbol{B}_j + \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned} \tag{2.20}$$

$j = 1, \dots, k$, $\boldsymbol{P} = \text{diag}\{p_1, \dots, p_n\}$, $p_i = 1/(2\phi_i^2)$, $i = 1, \dots, n$, $\boldsymbol{W} = \text{diag}\{w_1, \dots, w_n\}$, where

$$w_i = \frac{1}{\phi_i}\left[\frac{3\phi_i}{\mu_i(1-\mu_i)} + \frac{1}{\mu_i^3(1-\mu_i)^3}\right]\left(\frac{\partial g(\mu_i)}{\partial \mu_i}\right)^{-2}, \ i = 1, \dots, n.$$

2.4.3 BR-SPAM

The main diagonal elements of the Hessian matrix (Equation (2.15)) related to the BR-SPAM, are defined below (see Appendix A.3.2 for the respective demonstration),

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = \boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{X}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} = \boldsymbol{B}_j^\top\boldsymbol{R}\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j,\\
\ddot{\boldsymbol{U}}_p^{\kappa\kappa}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top} = \boldsymbol{E}^\top\boldsymbol{S}\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\alpha\alpha}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\alpha^2} = \text{trace}(\boldsymbol{J}),
\end{aligned} \tag{2.21}$$

$j = 1, \ldots, k$, $\boldsymbol{R} = \mathrm{diag}\{r_1, \ldots, r_n\}$,

$$r_i = \left\{ \left[\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^2} \right] \left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2} \right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)} \right] \left[-\frac{4\alpha}{(1-\varepsilon_i)} + \frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^3} \right] \right.$$
$$+ (1-\alpha)(1-\hat{v}_i) \left[\frac{-(1-\alpha)\phi_i^2}{(1-\varepsilon_i)^6} w_i^* + \frac{6\alpha\phi_i(1-2\mu_i)}{(1-\varepsilon_i)^5}(y_i^* - \delta_i^*) \right] - f_i^* \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1}$$
$$\left. \times \left(\frac{\partial^2 g_1(\mu_i)}{\partial \mu_i^2} \right) \right\} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-2}, \ i = 1, \ldots, n,$$

where f_i^* is defined in Equation (2.13), $w_i^* = \Psi'(\delta_i\phi_i) + \Psi'((1-\delta_i)\phi_i)$, $\boldsymbol{S} = \mathrm{diag}\{s_1, \ldots, s_n\}$,

$$s_i = \left[-(1-\hat{v}_i)[(1-\delta_i)^2\Psi'((1-\delta_i)\phi_i) + \delta^2\Psi'(\delta_i\phi_i) - \Psi'(\phi_i)] - a_i \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \right.$$
$$\left. \times \left(\frac{\partial^2 g_2(\phi_i)}{\partial \phi_i^2} \right) \right] \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-2},$$

a_i is given in Equation (2.13), $i = 1, \ldots, n$. Also, $\boldsymbol{J} = \mathrm{diag}\{j_1, \ldots, j_n\}$, where

$$j_i = \left\{ \left[\frac{4\mu_i^2(1-\mu_i)^2}{(1-\varepsilon_i)^2} \right] \left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2} \right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)} \right] \left[\frac{4\mu_i^2(1-\mu_i)^2}{(1-\varepsilon_i)^3} \right] \right.$$
$$\left. - (1-\hat{v}_i) \left[\frac{\phi\mu_i^2(1-\mu_i)^2(1-2\mu_i)}{(1-\varepsilon_i)^5} \right] \left[\frac{\phi_i(1-2\mu_i)}{(1-\varepsilon_i)} w_i^* + 6(y_i^* - \delta_i^*) \right] \right., \ i = 1, \ldots, n.$$

The elements outside the main diagonal of the Hessian matrix, related to Equation (2.15), are given by (see Appendix A.3.2 for the respective demonstration),

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\kappa}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\kappa}^\top} = -\boldsymbol{X}^\top \boldsymbol{R}^* \boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\kappa}^\top} = -\boldsymbol{B}_j^\top \boldsymbol{R}^* \boldsymbol{E}, \\
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_{j'}^\top} = \boldsymbol{B}_j^\top \boldsymbol{R} \boldsymbol{B}_{j'}, \quad \ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_j^\top} = \boldsymbol{X}^\top \boldsymbol{R} \boldsymbol{B}_j, \\
\ddot{\boldsymbol{U}}_p^{\beta\alpha}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\beta} \partial \alpha} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{s}^*, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\alpha}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\gamma}_j \partial \alpha} = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{s}^*, \\
\ddot{\boldsymbol{U}}_p^{\kappa\alpha}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial \boldsymbol{\kappa} \partial \alpha} = -\boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{c}^*,
\end{aligned} \tag{2.22}$$

where $\boldsymbol{T}_1 = \mathrm{diag}\{(\partial g_1(\mu_1)/\partial\mu_1)^{-1}, \ldots, (\partial g_1(\mu_n)/\partial\mu_n)^{-1}\}$, $\boldsymbol{R}^* = \mathrm{diag}\{r_1^*, \ldots, r_n^*\}$,

$$r_i^* = -(1-\hat{v}_i) \left\{ \frac{(1-\alpha)}{(1-\varepsilon_i)^3} [\phi_i(\delta_i w_i^* - \Psi'((1-\delta_i)\phi_i)) - (y_i^* - \delta_i^*)] \right\}$$
$$\times \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1}.$$

Furthermore, $\boldsymbol{T}_2 = \mathrm{diag}\{(\partial g_2(\phi_1)/\partial\phi_1)^{-1}, \ldots, (\partial g_2(\phi_n)/\partial\phi_n)^{-1}\}$, $\boldsymbol{s}^* = (s_1^*, \ldots, s_n^*)^\top$, where

$$\begin{aligned} s_i^* = & \left[\frac{4\alpha\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right] \\ & + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\left\{2(1-2\mu_i)\left[\frac{1}{(1-\varepsilon_i)} + \frac{2\alpha\mu_i(1-\mu_i)}{(1-\varepsilon_i)^3}\right]\right\} \\ & + (1-\hat{v}_i)\left\{\left[\frac{(1-\alpha)\phi_i^2\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^6}w_i^*\right] + (y_i^* - \delta_i^*) \right. \\ & \left. \times \left\{-\phi_i\left[\frac{1}{(1-\varepsilon_i)^3} - \frac{6(1-\alpha)\mu_i(1-\mu_i)}{(1-\varepsilon_i)^5}\right]\right\}\right\} \; i = 1, \ldots, n, \end{aligned}$$

$\boldsymbol{c}^* = (c_1^*, \ldots, c_n^*)^\top$, and

$$c_i^* = (1-\hat{v}_i)\left\{\left[\frac{\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right][\phi_i(\delta_i w_i^* - \Psi'((1-\delta_i)\phi_i)) - (y_i^* - \delta_i^*)]\right\},$$

for $i = 1, \ldots, n$.

2.5 Parameter estimation

Beta-SPAM and Simplex-SPAM

For estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j, j = 1, \ldots, k$ in Beta-SPAM and Simplex-SPAM we consider the Fisher scoring and backfitting algorithms. For estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j, j = 1, \ldots, k$ in Beta-SPAM and Simplex-SPAM we consider the Fisher scoring and backfitting algorithms. Given that, under certain regularity conditions (Green and Silverman, 1993), the penalized maximum likelihood estimator (PMLE) ($\hat{\boldsymbol{\theta}}$) exists and is unique and considering $\boldsymbol{\lambda}$ fixed, is obtained as solution of $\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}} = \dot{\boldsymbol{U}}_p^{\gamma_1} = \cdots = \dot{\boldsymbol{U}}_p^{\gamma_1} = \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}} = \boldsymbol{0}$.

Since the respective solution have no-closed form, some optimization algorithm should be employed to find the respective PMLE estimates, then we adopted the Fisher scoring algorithm. Indeed, the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$, $j = 1, \ldots, k$ at the $(u+1)$-th step of iterative process to maximization the penalized likelihood function is defined as

$$\begin{pmatrix} \boldsymbol{K}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_1} & \ldots & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_k} \\ \boldsymbol{K}_p^{\gamma_1\boldsymbol{\beta}} & \boldsymbol{K}_p^{\gamma_1\gamma_1} & \ldots & \boldsymbol{K}_p^{\gamma_1\gamma_k} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{K}_p^{\gamma_k\boldsymbol{\beta}} & \boldsymbol{K}_p^{\gamma_k\gamma_1} & \ldots & \boldsymbol{K}_p^{\gamma_k\gamma_k} \end{pmatrix}^{(u)} \begin{pmatrix} \boldsymbol{\beta}^{(u+1)} - \boldsymbol{\beta}^{(u)} \\ \boldsymbol{\gamma}_1^{(u+1)} - \boldsymbol{\gamma}_1^{(u)} \\ \vdots \\ \boldsymbol{\gamma}_k^{(u+1)} - \boldsymbol{\gamma}_k^{(u)} \end{pmatrix} = \begin{pmatrix} \dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}} \\ \dot{\boldsymbol{U}}_p^{\gamma_1} \\ \vdots \\ \dot{\boldsymbol{U}}_p^{\gamma_k} \end{pmatrix}^{(u)}$$

Then, the backfitting (Gauss–Seidel) (Breiman and Friedman, 1985) iterations, that are used to solve the equations of above system, take the form:

$$\boldsymbol{\beta}^{(u+1)} = (\boldsymbol{X}^\top\boldsymbol{W}^{(u)}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{W}^{(u)}\left[\boldsymbol{z}^{(u)} - \sum_{i=1}^{k}\boldsymbol{B}_j\boldsymbol{\gamma}_j^{(u)}\right] \tag{2.23a}$$

$$\boldsymbol{\gamma}_j^{(u+1)} = (\boldsymbol{B}_j^\top\boldsymbol{W}^{(u)}\boldsymbol{B}_j + \lambda_j\boldsymbol{\Lambda}_j)^{-1}\boldsymbol{B}_j^\top\boldsymbol{W}^{(u)}\left[\boldsymbol{z}^{(u)} - \boldsymbol{X}\boldsymbol{\beta}^{(u+1)} - \sum_{j^*\neq j}\boldsymbol{B}_{j^*}\boldsymbol{\gamma}_{j^*}^{(u+1)}\right], \tag{2.23b}$$

$u = 0, 1, 2, \ldots,\ j = 1, \ldots, k$, where, $\boldsymbol{z} = \boldsymbol{X\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j\boldsymbol{\gamma}_j + (\boldsymbol{W}^{(u)})^{-1}\boldsymbol{T}_1\boldsymbol{f}^*$. For more details about backfitting see, for example, Hastie and Tibshirani (1990), Rigby and Stasinopoulos (2005) and Ibacache-Pulgar and Reyes (2017).

On the other hand, the PMLE of $\boldsymbol{\kappa}$, says $\hat{\boldsymbol{\kappa}}$, can be obtained, following Ibacache-Pulgar and Reyes (2017), using the Fisher scoring algorithm. Then, given $\boldsymbol{\beta}^{(u+1)}, \boldsymbol{\gamma}_1^{(u+1)}, \ldots, \boldsymbol{\gamma}_k^{(u+1)}$, in the $(u+1)$-th step of estimation algorithm, $\hat{\boldsymbol{\kappa}}$ estimates is obtained by the following equation,

$$\boldsymbol{\kappa}^{(u+1)} = \boldsymbol{\kappa}^{(u)} - (\boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}})^{-1}\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}\Big|_{(\boldsymbol{\beta}^{(u+1)}, \boldsymbol{\gamma}_1^{(u+1)}, \ldots, \boldsymbol{\gamma}_k^{(u+1)}, \boldsymbol{\kappa}^{(u)})}. \tag{2.24}$$

In summary, the PMLE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ can be obtained by iterating between the backfitting and the Fisher score algorithms to obtain PMLE of $\boldsymbol{\kappa}$, which is equivalent to the following iterative process,

(i) To choose suitable starting values, say $\boldsymbol{\beta}^{(0)}, \boldsymbol{\gamma}_1^{(0)}, \boldsymbol{\gamma}_2^{(0)}, \ldots, \boldsymbol{\gamma}_k^{(0)}$ and $\boldsymbol{\kappa}^{(0)}$;

(ii) To obtain $\boldsymbol{\beta}^{(u+1)}$ and $\boldsymbol{\gamma}_j^{(u+1)},\ j = 1, \ldots, k$ using the backfitting algorithm (Equations (2.23a) and (2.23b)), considering $\boldsymbol{\kappa}^{(u)}$;

(iii) Then, for current values $\boldsymbol{\beta}^{(u+1)}, \boldsymbol{\gamma}_1^{(u+1)}, \boldsymbol{\gamma}_2^{(u+1)}, \ldots, \boldsymbol{\gamma}_k^{(u+1)}$, to obtain $\boldsymbol{\kappa}^{(u+1)}$ through Equation (2.24);

(iv) Iterating between (ii) and (iii) until reach some convergence criterion, that is until $||\boldsymbol{\theta}^{(u+1)} - \boldsymbol{\theta}^{(u)}|| > \epsilon, \epsilon > 0$.

BR-SPAM

As presented in Section 2.3.3, the EM algorithm iterates between the E step, which consist in to compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(u)})$ (see Equation (2.9)), and the M step, which will be presented here. The penalized maximum likelihood estimator (PMLE) of $\hat{\boldsymbol{\theta}}$ is obtained as the solution of $\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}} = \dot{\boldsymbol{U}}_p^{\gamma_1} = \cdots = \dot{\boldsymbol{U}}_p^{\gamma_1} = \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}} = \boldsymbol{0}$. Since such equations do not lead to an analytical solution, it is necessary to use some numerical optimization algorithm. Among others, the L-BFGS-B algorithm (Byrd et al., 1995), optimization using PORT routines (Gay, 1990), the Nelder-Mead algorithm (Nelder and Mead, 1965), are interesting options.

Then, we subdivided the M step into two other steps: in the first one, the parameters related to the non-parametric part were estimated using one of the aforementioned algorithms, whereas, in the second one, given the estimates already obtained, the parameters related to the parametric part were estimated using the same former approach.

The numerical optimization procedures require to provide initial values. For $\boldsymbol{\beta}, \boldsymbol{\gamma}_1 \ldots, \boldsymbol{\gamma}_k$ and $\boldsymbol{\kappa}$, based on penalized least squares estimates, we choose $\boldsymbol{\beta}^{(0)} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}$

$\boldsymbol{X}^\top \boldsymbol{y}$, $\boldsymbol{\gamma}_j^{(0)} = (\boldsymbol{B}_j^\top \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j)^{-1} \boldsymbol{B}_j^\top (\boldsymbol{y} - \boldsymbol{X}\beta^{(0)}), j = 1, \ldots, k$ and $\boldsymbol{\kappa}^{(0)} = (\boldsymbol{E}^\top \boldsymbol{E})^{-1} \boldsymbol{E}^\top \boldsymbol{y}$. For α, we follow Silva et al. (2020), considering that it controls the tails of the beta rectangular distribution. Therefore we use the method of the moments estimator of the degrees of freedom, says ν, of the Student t distribution. Specifically, we have that $\hat{\nu}_{MM} = -2\sum_{i=1}^{n} y_i^2 / \left(\sum_{i=1}^{n} y_i^2 - n\right), n < \sum_{i=1}^{n} y_i^2$. Then, we consider the following reparameterization (since $\alpha \in [0, 1]$) $\hat{\nu}_R = \hat{\nu}_{MM}/(\hat{\nu}_{MM} + 1)$, such that $\alpha^{(0)} = 2\sum_{i=1}^{n} y_i^2 / \left(\sum_{i=1}^{n} y_i^2 + n\right)$.

The joint iterative process, in order to find the PMLE of $\boldsymbol{\theta}$, is given by

(i) To choose suitable starting values, say $\boldsymbol{\beta}^{(0)}, \boldsymbol{\gamma}_1^{(0)}, \boldsymbol{\gamma}_2^{(0)}, \ldots, \boldsymbol{\gamma}_k^{(0)}, \boldsymbol{\kappa}^{(0)}$ and $\alpha^{(0)}$;

(ii) To compute $\hat{\boldsymbol{v}}$ (E step in Equation (2.10));

(iii) To use some optimization algorithm to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(u)})$ in relation to $\boldsymbol{\gamma}^{(u)}$ to find $\boldsymbol{\gamma}_j^{(u+1)}, j = 1, \ldots, k$;

(iv) For current values $\boldsymbol{\gamma}_j^{(u+1)}, j = 1, \ldots, k$, to obtain $\boldsymbol{\beta}^{(u+1)}$, $\boldsymbol{\kappa}^{(u+1)}$ and $\alpha^{(u+1)}$, through some optimization algorithm to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(u)})$, concerning those parameters;

(v) Iterating between (ii) and (iii) until reach some convergence criterion, that is until $||\boldsymbol{\theta}^{(u+1)} - \boldsymbol{\theta}^{(u)}|| > \epsilon, \epsilon > 0$.

2.6 Effective degrees of freedom

In additive linear models, the degrees of freedom correspond to approximately the number of effective parameters related to the non-parametric components. Then, following Hastie and Tibshirani (1990) and Eilers and Marx (1996) the effective degrees of freedom corresponding to the j-th non-parametric component is given by: $df_{\gamma_j} = \text{trace}\{\boldsymbol{B}_j^\top \widehat{\boldsymbol{W}} \boldsymbol{B}_j (\boldsymbol{B}_j^\top \widehat{\boldsymbol{W}} \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j)^{-1}\}$, for the Beta-SPAM and the Simplex-SPAM, whereas, for the BR-SPAM, it is given by: $df_{\gamma_j} = \text{trace}\{\boldsymbol{B}_j^\top \widehat{\boldsymbol{R}} \boldsymbol{B}_j (\boldsymbol{B}_j^\top \widehat{\boldsymbol{R}} \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j)^{-1}\}, \forall j = 1, \ldots, k$ (for more details about the development for the degrees of freedom of the BR-SPAM, see Appendix A.5). In this case, the total effective degrees of freedom are approximately given by $df(\boldsymbol{\lambda}) \cong p + \sum_{j=1}^{k} df_{\gamma_j}(\boldsymbol{\lambda})$.

2.7 Model selection criteria

The Information Criteria (IC) adopted in this work are: the Akaike Information Criterion (Akaike, 1974), AIC $= -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2[s^* + df(\boldsymbol{\lambda})]$, the Bayesian Information Criterion (Schwarz et al., 1978), BIC $= -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + \log(n)[s^* + df(\boldsymbol{\lambda})]$, the Hannan-Quinn Information (Hannan and Quinn, 1979), HQIC $= -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2[s^* + df(\boldsymbol{\lambda})] \log(\log(n))$, the corrected AIC, AICc $= \text{AIC} + (2k^{**}(k^{**}+1))/(n-k^{**}-1)$, the sample adjusted BIC, SABIC $= -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + [s^* + df(\boldsymbol{\lambda})]\log((n+2)/24)$, and the Generalized Akaike Information Criterion GAIC $= -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + k^*[s^* + df(\boldsymbol{\lambda})]$, where $k^* = 1, \ldots, \log(n)$, $k^{**} = [s^* + df(\boldsymbol{\lambda})]$, s^* is $s^* = s$, for the Beta-SPAM and the Simplex-SPAM, whereas $s^* = s + 1$ for the BR-SPAM, where s is the number of precision/dispersion parameters.

2.8 Estimation of smoother parameters

For selecting the smooth parameter, it can be considered the generalized cross-validation method (see Wahba and Wold (1975) and Wood (2017), for example). Indeed, for the Beta-SPAM and the Simplex-SPAM, $\boldsymbol{\lambda}$ is chosen such

$$\widehat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}} \text{GCV}(\boldsymbol{\lambda}) = \text{argmin}_{\boldsymbol{\lambda}} \sum_{i=1}^{n} \frac{\hat{w}_i(\hat{z}_i^{**} - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2}{[1 - n^{-1}\text{trace}\{\widehat{\boldsymbol{H}}(\boldsymbol{\lambda})\}]^2},$$

where w_i is the i-th component of the diagonal of the respective Fisher information, z_i^{**} is the i-th component of the vector $\boldsymbol{z}^{**} = \boldsymbol{X\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j + (\boldsymbol{W})^{-1}\boldsymbol{T}_1\boldsymbol{f}^*$, and, for the BR-SPAM,

$$\widehat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}} \text{GCV}(\boldsymbol{\lambda}) = \text{argmin}_{\boldsymbol{\lambda}} \sum_{i=1}^{n} \frac{\hat{r}_i(\hat{z}_i^{**} - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2}{[1 - n^{-1}\text{trace}\{\widehat{\boldsymbol{H}}(\boldsymbol{\lambda})\}]^2},$$

where r_i is the i-th component of the diagonal of the respective Hessian matrix, and z_i^{**} is the i-th component of the vector $\boldsymbol{z}^{**} = \boldsymbol{X\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j - (\boldsymbol{R})^{-1}\boldsymbol{T}_1\boldsymbol{f}^*$ and $\boldsymbol{H}(\boldsymbol{\lambda})$ is defined as $\boldsymbol{H}(\boldsymbol{\lambda}) = \boldsymbol{N}^* \left[\boldsymbol{N}^{*\top}\dot{\boldsymbol{W}}\boldsymbol{N}^* - \boldsymbol{\Lambda}(\boldsymbol{\lambda})\right]^{-1} \boldsymbol{N}^{*\top}\dot{\boldsymbol{W}}$, where $\boldsymbol{\Lambda}(\boldsymbol{\lambda}) = \oplus_{j=1}^{n} \lambda_j \boldsymbol{\Lambda}_j^d$, $\boldsymbol{N}^*\boldsymbol{\xi} = [\boldsymbol{X}, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_k]\boldsymbol{\xi}$, $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top)^\top$ and $\dot{\boldsymbol{W}}$ is the respective Fisher information for Simplex-SPAM and Beta-SPAM, whereas, for BR-SPAM is the respective Hessian matrix. Another possibility is to select $\boldsymbol{\lambda}$ through Information Criteria, which was described in Section 2.7, for example doing $\widehat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}}\text{AIC}$.

2.9 Obtaining the standard errors

Beta-SPAM and Simplex-SPAM

Following Wood (2017), we consider, as the estimator of the covariance matrix, $(\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}))^{-1}$ and for standard error, $\text{diag}\{(\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}))^{-1/2}\}$, where $\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})$ defined in Equation (2.16).

BR-SPAM

Since we employed the EM algorithm, the standard errors will be obtained from the empirical information matrix. The Louis' principle (Louis, 1982) relates the score function of the incomplete data log-likelihood with the score function of the complete data log-likelihood, through the $\bigtriangledown_o(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[\bigtriangledown_c(\boldsymbol{\theta}; \boldsymbol{Y}_c|\boldsymbol{Y}_{obs})]$, where $\bigtriangledown_o(\boldsymbol{\theta}) = \partial l_o(\boldsymbol{\theta}; \boldsymbol{Y}_{obs})/\partial\boldsymbol{\theta}$ and $\bigtriangledown_c(\boldsymbol{\theta}) = \partial l_c(\boldsymbol{\theta}; \boldsymbol{Y}_c)/\partial\boldsymbol{\theta}$ are the score function of incomplete and complete data, respectively. As defined in Meilijson (1989), the empirical information matrix, can be calculated as:

$$\boldsymbol{I}_e(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})\boldsymbol{s}(y_i|\boldsymbol{\theta})^\top - \frac{1}{n}\boldsymbol{S}(\boldsymbol{y}|\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{y}|\boldsymbol{\theta})^\top, \tag{2.25}$$

where $\boldsymbol{S}(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})$, where $\boldsymbol{s}(y_i|\boldsymbol{\theta}),\ i = 1, \ldots, n$ is the score function for the i-th observation, given by:

$$\boldsymbol{s}(y_i|\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial l_p^c(\boldsymbol{\theta}; y_i, v_i)}{\partial\boldsymbol{\theta}}|y_i, \boldsymbol{\theta}\right],$$

and $l_p^c(\boldsymbol{\theta}|y_i, v_i)$ is the penalized complete log-likelihood given in Equation (2.8).

Replacing in Equation (2.25) $\boldsymbol{\theta}$ by its maximum likelihood estimator, says $\hat{\boldsymbol{\theta}}$ and considering $\bigtriangledown_o(\boldsymbol{\theta}) = \mathbf{0}$, Equation (2.25) can be rewritten as

$$\boldsymbol{I}_e(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})\boldsymbol{s}(y_i|\boldsymbol{\theta})^\top.$$

Specifically, the score vector for the i-th observation can be decomposed into:

$$\boldsymbol{s}(y_i|\boldsymbol{\theta}) = \left(s_{\boldsymbol{\beta}}(y_i|\boldsymbol{\theta}), s_{\boldsymbol{\gamma}_1}(y_i|\boldsymbol{\theta}), \ldots, s_{\boldsymbol{\gamma}_k}(y_i|\boldsymbol{\theta}), s_{\boldsymbol{\kappa}}(y_i|\boldsymbol{\theta}), s_\alpha(y_i|\boldsymbol{\theta})\right)^\top,$$

where

$$\begin{aligned}
s_{\boldsymbol{\beta}}(y_i|\boldsymbol{\theta}) &= \left\{\frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\right. \\
&\quad \left. + (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^* - \delta_i^*)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{x}_i, && (2.26a)\\
s_{\boldsymbol{\gamma}_j}(y_i|\boldsymbol{\theta}) &= \left\{\frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\right. \\
&\quad \left. + (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^* - \delta_i^*)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{b}_{ij} \\
&\quad - \lambda_j\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_j, && (2.26b)\\
s_{\boldsymbol{\kappa}}(y_i|\boldsymbol{\theta}) &= (1-\hat{v}_i)\{\delta_i(y_i^* - \delta_i^*) + \log(1-y_i) \\
&\quad - \Psi[(1-\delta_i)\phi_i] + \Psi(\phi_i)\}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{e}_i, && (2.26c)\\
s_\alpha(y_i|\boldsymbol{\theta}) &= \frac{2\mu_i(1-\mu_i)}{(1-\varepsilon_i)}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] - \left[\frac{\phi_i\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right](y_i^* - \delta_i^*), && (2.26d)
\end{aligned}$$

$j = 1, \ldots, k$ and $\hat{v}_i$ is given by Equation (2.10). Thus, the empirical information matrix can be calculated through Equations (2.26a) to (2.26d). Therefore, $\text{Cov}(\hat{\boldsymbol{\theta}})$ and $\text{SE}(\hat{\boldsymbol{\theta}})$ are estimated by $(\boldsymbol{I}_e(\hat{\boldsymbol{\theta}}|\boldsymbol{y}))^{-1}$ and $\text{diag}\{(\boldsymbol{I}_e(\hat{\boldsymbol{\theta}}|\boldsymbol{y}))^{-1/2}\}$, respectively.

2.10 Hypothesis testing

For testing linear statistical hypothesis, that is, $\boldsymbol{C}\varsigma = \boldsymbol{d}$, where rank$(\boldsymbol{C}) = l$, $l \geq p$ or $l \geq q$, it can be used the following Wald-type statistic

$$\xi_W = (\boldsymbol{C}\hat{\boldsymbol{\varsigma}} - \boldsymbol{d})^\top (\boldsymbol{C}\boldsymbol{V}_\varsigma \boldsymbol{C}^\top)^{-1}(\boldsymbol{C}\hat{\boldsymbol{\varsigma}} - \boldsymbol{d}).$$

Under H_0, $\xi_W \xrightarrow[n\to\infty]{D} \chi^2_l$, where $\xrightarrow[n\to\infty]{D}$ means convergence in distribution when the sample size tends to infinity. Also, $\boldsymbol{\varsigma}$ can be either a subvector of $\boldsymbol{\beta}$ or a subvector of $\boldsymbol{\kappa}$, $\boldsymbol{V}_\varsigma$ is the matrix of variance-covariance related to $\boldsymbol{\varsigma}$, obtainable from $\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}))^{-1}$.

2.11 Model fit diagnostic tools

Statistical models, in a general way, are sensitive to the lack of the underlying assumptions. Also, influential observations can lead to misleading conclusions. Therefore it is important to check the goodness of model fit, at an overall level and for each specific assumption, as well as to check the presence of influential observations. Here we developed residual and influence analysis tools.

2.11.1 Leverage

The main idea underlying leverage analysis consists on measuring the variation of predicted values, under some perturbation scheme of respective observed values. Within this framework, one of the most usual approach is the generalized leverage, introduced by Wei et al. (1998), defined as

$$\text{GL}(\boldsymbol{\theta}) = \left\{ \boldsymbol{D}_{\boldsymbol{\theta}} \left(-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}} \right)^{-1} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta y}} \right\} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where, for all three models, $\boldsymbol{D}^{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top = [\boldsymbol{T}_1\boldsymbol{X}, \boldsymbol{T}_1\boldsymbol{B}_1, \ldots, \boldsymbol{T}_1\boldsymbol{B}_k, \boldsymbol{0}_{(n\times s)}]$, $\boldsymbol{0}_{(n\times s)}$ is a matrix of zeros. Also, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}$ and $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta y}}$ assume the following structure, for each one of the developed models:

- Beta-SPAM: The elements of $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}} = \partial^2 l_p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are defined in Equation (2.17) and

  $$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta y}} = \frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top} = [\boldsymbol{X}^\top\boldsymbol{T_1}\boldsymbol{\Phi}\boldsymbol{M}, \boldsymbol{B}_1^\top\boldsymbol{T_1}\boldsymbol{\Phi}\boldsymbol{M}, \ldots, \boldsymbol{B}_k^\top\boldsymbol{T_1}\boldsymbol{\Phi}\boldsymbol{M}, \boldsymbol{E}^\top\boldsymbol{T}_2\boldsymbol{B}^*]^\top,$$

  where $\boldsymbol{B}^* = \text{diag}\{b_1^*, \ldots, b_n^*\}$, $b_i^* = -(y_i - \mu_i)/[y_i(1-y_1)]$, $\boldsymbol{M} = \text{diag}\{m_1, \ldots, m_n\}$ and $m_i = 1/[y_i(1-y_i)]$ for $i = 1, \ldots, n$.

- Simplex-SPAM: The elements of $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}} = \partial^2 l_p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are defined in Equation (2.19) and

  $$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta y}} = \frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top} = [\boldsymbol{X}^\top\boldsymbol{\Phi}^*\boldsymbol{T}_1\boldsymbol{M}, \boldsymbol{B}_1^\top\boldsymbol{\Phi}^*\boldsymbol{T}_1\boldsymbol{M}, \ldots, \boldsymbol{B}_k^\top\boldsymbol{\Phi}^*\boldsymbol{T}_1\boldsymbol{M}, \boldsymbol{E}^\top\boldsymbol{T}_2\boldsymbol{B}^*]^\top,$$

where $\boldsymbol{B}^* = \text{diag}\{b_1^*, \ldots, b_n^*\}$,

$$b_i^* = \frac{1}{2\phi_i y_i(1-y_i)}\left[\frac{2(y_i-\mu_i)}{\mu_i^2(1-\mu_i)^2} - d(y_i;\mu_i)(1-2y_i)\right], \tag{2.27}$$

$d(y;\mu)$ is given in Equation (1.7), $\boldsymbol{M} = \text{diag}\{m_1, \ldots, m_n\}$ and

$$m_i = \frac{1}{\phi_i\mu_i^3(1-\mu_i)^3}\left[\frac{3(y_i-\mu_i)^2}{y_i(1-y_i)} - \frac{(1-2y)(y_i-\mu_i)^3}{y_i^2(1-y_i)^2} + 1\right]. \tag{2.28}$$

- BR-SPAM: The elements of $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}} = \partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are defined in Equations (2.21) and (2.22) and

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta y}} = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top} = [\boldsymbol{X}^\top\boldsymbol{T}_1\boldsymbol{M}, \boldsymbol{B}_1^\top\boldsymbol{T}_1\boldsymbol{M}, \ldots, \boldsymbol{B}_k^\top\boldsymbol{T}_1\boldsymbol{M}, \boldsymbol{E}^\top\boldsymbol{T}_2\boldsymbol{B}^*, \boldsymbol{h}^\top]^\top,$$

where $\boldsymbol{B}^* = \text{diag}\{b_1^*, \ldots, b_n^*\}$, $b_i^* = (1-\hat{v}_i)[(\delta_i - y_i)/(y_i(1-y_i))]$, $i = 1, \ldots, n$, $\boldsymbol{M} = \text{diag}\{m_1, \ldots, m_n\}$,

$$m_i = (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}\frac{1}{y_i(1-y_i)}, \; i = 1, \ldots, n, \tag{2.29}$$

$\boldsymbol{h} = (h_1, \ldots, h_n)^\top$, and

$$h_i = -(1-\hat{v}_i)\left[\frac{\phi_i\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right]\frac{1}{y_i(1-y_i)}, \; i = 1, \ldots, n. \tag{2.30}$$

The index plot of $\text{GL}(\hat{\boldsymbol{\theta}})_{ii}$ and $\text{GL}(\hat{\boldsymbol{\theta}})_{ii}$ against the fitted values can be used to detect observations with high leverage values.

2.11.2 Residual analysis

Residual analysis aims to check the validity of model assumptions and/or the goodness of model fit at an overall level and for specific model assumptions. A useful approach is to so-called the quantile residuals (see Stasinopoulos et al. (2007)). Given a continuous probability density, says $f(y_i|\boldsymbol{\theta})$, $i = 1, \ldots, n$, the associated cumulative distribution function, says $F(Y_i;\boldsymbol{\theta})$, are uniformly distributed on the unit interval, if the model is well fitted to the data. Indeed, the quantile residuals are defined as

$$r_i = \Phi^{-1}(F(Y_i|\hat{\boldsymbol{\theta}})) = \Phi^{-1}\left(\int_0^{Y_i} f_Y(y|\hat{\boldsymbol{\theta}})dy\right),$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution and $f_Y(y)$ is defined in Equations (1.1), (1.6) and (1.3) for the Beta-SPAM, Simplex-SPAM and BR-SPAM, respectively. Thus, if $\widehat{\boldsymbol{\theta}}$ are consistent estimators, the distribution of r_i converges to the standard normal. Therefore, we can build appropriate quantile quantile plots with envelopes, as well as other usual plots as the residuals against the index observations and the fitted values.

2.11.3 Cook's Distance

Case deletion diagnostics is commonly employed for measuring the influence of each observation in the fitted model and is built considering the model without the i-th observation, i. e., $t = 1, \ldots, n$, with $t \neq i$ being the new index of variables, the penalized log-likelihood, $l_{p(i)}(\boldsymbol{\theta})$, is given by

$$l_{p(i)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l_{(i)}(\boldsymbol{\theta}) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_{j(i)}^{\top} \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_{j(i)} = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_{j(i)}^{\top} \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_{j(i)}.$$

Let $\hat{\boldsymbol{\theta}}_{(i)} = (\boldsymbol{\beta}_{(i)}^{\top}, \hat{\gamma}_{1(i)}^{\top}, \ldots, \hat{\gamma}_{k(i)}^{\top}, \boldsymbol{\kappa}_{(i)}^{\top})^{\top}$ be the MPLE of $\boldsymbol{\theta}$, says $\boldsymbol{\theta}$ based on $l_{p(i)}(\boldsymbol{\theta}, \boldsymbol{\lambda})$. To evaluate the influence of i-th observation on the MPLE estimate of $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^{\top}, \hat{\gamma}_1^{\top}, \ldots, \hat{\gamma}_k^{\top}, \hat{\boldsymbol{\kappa}}^{\top})^{\top}$, we may compare $\hat{\boldsymbol{\theta}}_{(i)}$ with $\hat{\boldsymbol{\theta}}$. The higher is the influence of a given observation, the more attention it is necessary. If the distance between these two sets of estimates are sufficiently large, such observation is considered as influential.

It is possible to see that it is necessary to obtain $\hat{\boldsymbol{\theta}}_{(i)}$ for all observations, which can demand a large amount of time, mainly when the number of observations is large. For this reason, it is usual to consider instead of $\hat{\boldsymbol{\theta}}_{(i)}$, $\hat{\boldsymbol{\theta}}_{(i)}^* = \hat{\boldsymbol{\theta}} + [-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})]^{-1} \dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$, where

$$\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \left. \frac{\partial l_{p(i)}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{and} \quad \ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \left. \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\top}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

which was proposed by Cook and Weisberg (1982). The generalized Cook's distance is then defined as the Euclidean norm of $(\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})$, which is given by

$$\text{GCD} = (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}})^{\top} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) (\hat{\boldsymbol{\theta}}_{(i)} - \hat{\boldsymbol{\theta}}), \tag{2.31}$$

where $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ is the observed information matrix. Then, replacing $\hat{\boldsymbol{\theta}}_{(i)}$ by $\hat{\boldsymbol{\theta}}_{(i)}^*$ in Equation (2.31), we have that

$$\begin{aligned} \text{GCD} &= \{[-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})]^{-1} \dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\}^{\top} [-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})] \{[-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})]^{-1} \dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\} \\ &= [\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}]^{\top} \{[-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})]^{-1}\}^{\top} \dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}} \\ &= [\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}]^{\top} [-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})]^{-1} \dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}} = [\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}]^{\top} \hat{\boldsymbol{\theta}}_{(i)}^*, \end{aligned}$$

where $\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}$ is the gradient vector without the i-th observation, i.e., it is defined for $t = 1, \ldots, n$ with $t \neq i$. To identify influential observations through the Cook's distance, the index plot of GCD_i is recommend.

For the Beta-SPAM and the Simplex-SPAM, $\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}$ is defined as in Equations (2.11) and (2.12), respectively, and the components of the vector $\hat{\boldsymbol{\theta}}_{(i)}^* = (\hat{\boldsymbol{\beta}}_{(i)}^{*\top}, \hat{\gamma}_{1(i)}^{*\top}, \ldots, \hat{\gamma}_{k(i)}^{*\top},$

$\hat{\boldsymbol{\kappa}}_{(i)}^{*\top})^\top$, are given by,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{(i)}^{*} &= \hat{\boldsymbol{\beta}} + (\boldsymbol{X}^\top \boldsymbol{Q} \boldsymbol{X})^{-1} \boldsymbol{X}_{(i)}^\top \boldsymbol{T}_{1(i)} \boldsymbol{f}_{(i)}^{*},\\
\hat{\boldsymbol{\gamma}}_{1(i)}^{*} &= \hat{\boldsymbol{\gamma}}_1 + (\boldsymbol{B}_1^\top \boldsymbol{Q} \boldsymbol{B}_1 + \lambda_1 \boldsymbol{\Lambda}_1)^{-1} [\boldsymbol{B}_{1(i)}^\top \boldsymbol{T}_{1(i)} \boldsymbol{f}_{(i)}^{*} - \lambda_1 \boldsymbol{\Lambda_1} \hat{\boldsymbol{\gamma}}_1],\\
&\vdots\\
\hat{\boldsymbol{\gamma}}_{k(i)}^{*} &= \hat{\boldsymbol{\gamma}}_k + (\boldsymbol{B}_k^\top \boldsymbol{Q} \boldsymbol{B}_k + \lambda_k \boldsymbol{\Lambda}_k)^{-1} [\boldsymbol{B}_{k(i)}^\top \boldsymbol{T}_{1(i)} \boldsymbol{f}_{(i)}^{*} - \lambda_k \boldsymbol{\Lambda_k} \hat{\boldsymbol{\gamma}}_k],\\
\hat{\boldsymbol{\kappa}}_{(i)}^{*} &= \hat{\boldsymbol{\kappa}} + (\boldsymbol{E}^\top \boldsymbol{S} \boldsymbol{E})^{-1} \boldsymbol{E}_{(i)}^\top \boldsymbol{T}_{2(i)} \boldsymbol{a}_{(i)},
\end{aligned}
$$

where $\boldsymbol{B}_{j(i)}$ for $j = 1, \ldots, k$, $\boldsymbol{E}_{(i)}$, $\boldsymbol{X}_{(i)}$, $\boldsymbol{T}_{1(i)}$, $\boldsymbol{T}_{2(i)}$, $\boldsymbol{f}_{(i)}^{*}$, $\boldsymbol{a}_{(i)}$ are given in Equations (2.1) for Beta-SPAM and (2.11) for Simplex-SPAM, without the i-th observation.

For the BR-SPAM, instead of the log-likelihood, we do the development based on $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, then, let $Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \mathbb{E}[l_{p(i)}^c(\boldsymbol{\theta})|\boldsymbol{y}_{(i)}, \hat{\boldsymbol{\theta}}]$, with $l_{p(i)}^c(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t^*(\boldsymbol{\theta}) - \boldsymbol{\gamma}^\top \boldsymbol{\Lambda} \boldsymbol{\gamma}$, and $l_t^*(\boldsymbol{\theta})$ defined in Equation (2.7). Additionally $\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ and $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\hat{\boldsymbol{\theta}})$ are defined, respectively, as,

$$
\dot{\boldsymbol{U}}_{p(i)}^{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q_{(i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{and} \quad \ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\hat{\boldsymbol{\theta}}) = \left.\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.
$$

Finally, the components of the vector $\hat{\boldsymbol{\theta}}_{(i)}^{*} = (\hat{\boldsymbol{\beta}}_{(i)}^{*\top}, \hat{\boldsymbol{\gamma}}_{1(i)}^{*\top}, \ldots, \hat{\boldsymbol{\gamma}}_{k(i)}^{*\top}, \hat{\boldsymbol{\kappa}}_{(i)}^{*\top}, \hat{\alpha}_{(i)}^{*})^\top$, are given by

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{(i)}^{*} &= \hat{\boldsymbol{\beta}} + \left(-\boldsymbol{X}^\top \boldsymbol{R} \boldsymbol{X}\right)^{-1} \boldsymbol{X}_{(i)}^\top \boldsymbol{T}_{(i)} \boldsymbol{f}_{(i)}^{*},\\
\hat{\boldsymbol{\gamma}}_{1(i)}^{*} &= \hat{\boldsymbol{\gamma}}_1 + \left(-\boldsymbol{B}_1^\top \boldsymbol{R} \boldsymbol{B}_1 + \lambda_1 \boldsymbol{\Lambda}_1\right)^{-1} \left[\boldsymbol{B}_{1(i)}^\top \boldsymbol{T}_{(i)} \boldsymbol{f}_{(i)}^{*} - \lambda_1 \boldsymbol{\Lambda_1} \hat{\boldsymbol{\gamma}}_1\right],\\
&\vdots\\
\hat{\boldsymbol{\gamma}}_{k(i)}^{*} &= \hat{\boldsymbol{\gamma}}_k + \left(-\boldsymbol{B}_k^\top \boldsymbol{R} \boldsymbol{B}_k + \lambda_k \boldsymbol{\Lambda}_k\right)^{-1} \left[\boldsymbol{B}_{k(i)}^\top \boldsymbol{T}_{(i)} \boldsymbol{f}_{(i)}^{*} - \lambda_k \boldsymbol{\Lambda_k} \hat{\boldsymbol{\gamma}}_k\right],\\
\hat{\boldsymbol{\kappa}}_{(i)}^{*} &= \hat{\boldsymbol{\kappa}} + (\boldsymbol{E}^\top \boldsymbol{S} \boldsymbol{E})^{-1} \boldsymbol{E}_{(i)}^\top \boldsymbol{T}_{2(i)} \boldsymbol{a}_{(i)},\\
\hat{\alpha}_{(i)}^{*} &= \hat{\alpha} + [\mathrm{trace}(\boldsymbol{J})]^{-1} \mathrm{trace}(\boldsymbol{D}_{(i)}),
\end{aligned}
$$

where $\boldsymbol{X}_{(i)}$, $\boldsymbol{f}_{(i)}^{*}$, $\boldsymbol{T}_{1(i)}$, $\boldsymbol{T}_{2(i)}$, $\boldsymbol{D}_{(i)}$, $\boldsymbol{a}_{(i)}$, $\boldsymbol{B}_{j(i)}$, for $j = 1, \ldots, k$ are given in Equations (2.1) and (2.13), without the i-th observation.

2.11.4 Local influence analysis

The likelihood displacement for Simplex-SPAM and Beta-SPAM, which defined as $\mathrm{LD}(\boldsymbol{\omega}) = 2[l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) - l_p(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}, \boldsymbol{\lambda})] \geq 0$, can be used to assess the influence of some perturbation scheme on the MPLE estimates $(\hat{\boldsymbol{\theta}})$. Here, $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ is the MPLE estimate under a perturbed model, and the perturbed penalized log-likelihood function is denoted by $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})$, where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is an n-dimensional vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$. Furthermore, it is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}_0) = l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

When the complete log-likelihood, namely $l_p^c(\boldsymbol{\theta}|\boldsymbol{y}_c)$, is used as in BR-SPAM, it is possible to define $l_p^c(\boldsymbol{\theta}, \boldsymbol{\omega}|\boldsymbol{y}_c)$ as the complete log-likelihood of the perturbed model and to use the Q-displacement function defined as $\mathrm{QD}(\boldsymbol{\omega}) = 2[Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}_\omega|\hat{\boldsymbol{\theta}})]$ for local influence analysis, where $\hat{\boldsymbol{\theta}}_\omega$ is the maximum of $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \mathbb{E}[l_p^c(\boldsymbol{\theta}|\boldsymbol{y}, \hat{\boldsymbol{\theta}})]$ and it is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $Q(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\boldsymbol{y}_c) = Q(\boldsymbol{\theta}|\boldsymbol{\omega}_0)$.

The idea of local influence consists on to study the behavior of $\mathrm{LD}(\boldsymbol{\omega})(\mathrm{QD}(\boldsymbol{\omega}))$ around $\boldsymbol{\omega}_0$. The procedure seeks to select an arbitrary direction $\boldsymbol{l}$ of norm 1 ($\| \boldsymbol{l} \| = 1$), and inspecting the plot of $\mathrm{LD}(\boldsymbol{\omega} + a\boldsymbol{l})(\mathrm{QD}(\boldsymbol{\omega} + a\boldsymbol{l}))$ against a, where $a \in \mathbb{R}$, which is known as projected line. In particular, when $\mathrm{LD}(\boldsymbol{\omega}_0) = 0(\mathrm{QD}(\boldsymbol{\omega}_0) = 0)$, we have that $\mathrm{LD}(\boldsymbol{\omega}_0 + a\boldsymbol{l})(\mathrm{QD}(\boldsymbol{\omega}_0 + a\boldsymbol{l}))$ has a local minimum at $a = 0$. Each projected line can be characterized by a normal curvature $\boldsymbol{C}_l(\boldsymbol{\theta})$ around $a = 0$. This curvature is interpreted as the the inverse of the radius of the best adjusted circle at $a = 0$. Then, instead of study all possibles for $\boldsymbol{l}$, a suggestion is to consider the direction $\boldsymbol{l}_{max}$ that corresponds to the highest curvature, denoted by $C_{\boldsymbol{l}_{max}}$. Cook (1986) used the concepts of differential geometry to show that the normal curvature in the $\boldsymbol{l}$ direction assumes the form $C_l(\boldsymbol{\theta})$, and based on that, Zhu and Lee (2001) proposed a normal curvature using $\mathrm{QD}(\boldsymbol{\omega})$, showing that the normal curvature in the $\boldsymbol{l}$ direction is given by

$$C_l(\boldsymbol{\theta}) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top [-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta} \boldsymbol{l}|,$$

where $-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})$ is the observed Fisher information matrix, and $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})$ is defined in the Equation (2.15). For the curvature based on $\mathrm{LD}(\boldsymbol{\theta})$ and $\mathrm{QD}(\boldsymbol{\theta})$, $\boldsymbol{\Delta}$, is given, respectively, by:

$$\boldsymbol{\Delta} = \left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \text{ and } \boldsymbol{\Delta} = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

The maximum of $\boldsymbol{l}^\top \boldsymbol{Z} \boldsymbol{l}$, where $\boldsymbol{Z} = \boldsymbol{\Delta}^\top [-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta}$, corresponds to the highest eigenvalue (in absolute value) of $\boldsymbol{Z}$. Thus, $C_{\boldsymbol{l}_{max}}$ corresponds to the highest eigenvalue of the $\boldsymbol{Z}$ matrix and $\boldsymbol{l}_{max}$ denotes the corresponding eigenvector.

The index plot of $|\boldsymbol{l}_{\max}|$ may reveal the observations with the highest influence at the neighborhood of $\mathrm{LD}(\boldsymbol{\omega}_0)(\mathrm{QD}(\boldsymbol{\omega}_0))$. Those observations may lead to substantial changes in the estimates of parameters under little perturbations in the model and/or the data. If the interest lies on studying the local influence on a subvector $\boldsymbol{\theta}_1$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, the respective normal curvature at the direction $\boldsymbol{l}$ is given by $C_l(\boldsymbol{\theta}_1) = 2|\boldsymbol{l}^\top \boldsymbol{\Delta}^\top ([\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})]^{-1} - \boldsymbol{Z}_1) \boldsymbol{\Delta} \boldsymbol{l}|$, where

$$\boldsymbol{Z}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\theta})]^{-1} \end{pmatrix},$$

and $-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\theta})$ denoting the observed penalized Fisher information matrix for $\boldsymbol{\theta}_2$. The index plot of the eigenvector $\boldsymbol{l} = \boldsymbol{l}_{max}$, which corresponds to the largest absolute eigenvalue of the matrix $\boldsymbol{\Delta}^\top ([\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})]^{-1} - \boldsymbol{Z}_1) \boldsymbol{\Delta}$, may indicate the points with large influence on $\hat{\boldsymbol{\theta}}_1$.

In this section, three different perturbation schemes are considered, namely: case-weight, response perturbation and covariate perturbation. To obtain local influence measures it is necessary to obtain $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}$ and $\boldsymbol{\Delta}$, which are presented, for each perturbation schemes, below.

Case-weight perturbation

In this case the perturbed penalized log-likelihood function is given by

$$l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = l(\boldsymbol{\theta}|\boldsymbol{\omega}) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j,$$

where $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i l_i(\boldsymbol{\theta})$, with $0 \leq \omega_i \leq 1$. Hence, considering $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$, the elements of the vector $\boldsymbol{\Delta}$, for the Beta-SPAM and Simplex-SPAM, are expressed as

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}^\top \hat{\boldsymbol{T}}_1 \mathrm{diag}\{\hat{\boldsymbol{f}}^*\},$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{B}_j^\top \hat{\boldsymbol{T}}_1 \mathrm{diag}\{\hat{\boldsymbol{f}}^*\} - \hat{\boldsymbol{\mu}}^*\}, \quad j = 1, \ldots, k,$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{E}^\top \hat{\boldsymbol{T}}_2 \mathrm{diag}\{\hat{\boldsymbol{a}}\},$$

where $\boldsymbol{a}$, $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, $\boldsymbol{f}^*$ are all defined in the Equation (2.11) for Beta-SPAM and in Equation (2.12) for Simplex-SPAM.

For the BR-SPAM, the perturbed Q-function is given by

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}),$$

with $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Hence, the elements of the vector $\boldsymbol{\Delta}$ are expressed as

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}^\top \hat{\boldsymbol{T}}_1 \mathrm{diag}\{\hat{\boldsymbol{f}}^*\},$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \gamma_j \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{B}_j^\top \hat{\boldsymbol{T}}_1 \mathrm{diag}\{\hat{\boldsymbol{f}}^*\}, \quad j = 1, \ldots, k,$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{E}^\top \hat{\boldsymbol{T}}_2 \mathrm{diag}\{\hat{\boldsymbol{a}}\},$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \alpha \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \hat{\boldsymbol{d}}^*,$$

where $\boldsymbol{T}$, $\boldsymbol{a}$ and $\boldsymbol{f}^*$ are defined in the Equation (2.13). Additionally, $\boldsymbol{d}^* = (d_1^*, \ldots, d_n^*)^\top$, with

$$d_i^* = \frac{2\mu_i(1-\mu_i)}{(1-\varepsilon_i)} \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] - \left[\frac{\phi\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right](y_i^* - \delta_i^*), \tag{2.32}$$

for $i = 1, \ldots, n$.

Response variable perturbation

Let us consider an additive perturbation on the i-th response by making $y_{i\omega} = y_i + \omega_i$. Then, the perturbed penalized log-likelihood is given by Equation (2.5) with y_i replaced by $y_{i\omega}$. Then, considering $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$, the elements of the vector $\boldsymbol{\Delta}$, for Beta-SPAM and for Simplex-SPAM, are expressed as

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}^\top \hat{\boldsymbol{T}}_1 \hat{\boldsymbol{M}},$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{B}_j^\top \hat{\boldsymbol{T}}_1 \hat{\boldsymbol{M}}, \quad j = 1, \ldots, k,$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{E}^\top \hat{\boldsymbol{T}}_2 \hat{\boldsymbol{B}}^*,$$

where, for Beta-SPAM, $\boldsymbol{M} = \mathrm{diag}\{m_1, \ldots, m_n\}$ and $\boldsymbol{B}^* = (b_1^*, \ldots, b_n^*)^\top$, $m_i = \phi_i/[y_i(1 - y_i)]$ and $b_i^* = -[y_i - \mu_i]/[y_i(1 - y_i)]$, $i = 1, \ldots, n$, whereas, for Simplex-SPAM, $\boldsymbol{M} = \mathrm{diag}\{\phi_i m_1, \ldots, \phi_i m_n\}$, $m_i,\ i = 1, \ldots, n$ defined in Equation (2.28) and $\boldsymbol{B}^* = \mathrm{diag}\{b_1^*, \ldots, b_n^*\}^\top$ and $b_i^*,\ i = 1, \ldots, n$ defined in Equation (2.27).

For the BR-SPAM, the perturbed Q-function, under the same additive perturbation on the i-th response, is constructed from Equation (2.9) with y_i replaced by $y_{i\omega}$. The elements of the vector $\boldsymbol{\Delta}$ are expressed as

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}^\top \hat{\boldsymbol{T}}_1 \hat{\boldsymbol{M}},$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{B}_j^\top \hat{\boldsymbol{T}}_1 \hat{\boldsymbol{M}}, \quad j = 1, \ldots, k,$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{E}^\top \hat{\boldsymbol{T}}_2 \hat{\boldsymbol{B}}^*$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \alpha \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \hat{\boldsymbol{h}}^\top,$$

where $\boldsymbol{M} = \mathrm{diag}\{m_1, \ldots, m_n\}$, $m_i,\ i = 1, \ldots, n$, $\boldsymbol{B}^*$ and $\boldsymbol{h}$ are all defined in Equation (2.29). Also $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ are presented in Equation (2.13).

Continuous covariate perturbation

Consider an additive perturbation on a continuous covariate by making $\boldsymbol{x}_d + \boldsymbol{\omega} s_{x_d}$, where $\boldsymbol{\omega} \in \mathbb{R}$ is a vector of small perturbations and s_{x_d} is the standard deviation of $\boldsymbol{x}_d$ (Thomas and Cook, 1990). For the linear predictor related to the mean, the perturbation will be adding only on the covariates related to the parametric components. Then, under the scheme of covariate perturbation, we have that $\eta_{1i\omega} = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_d(x_{id} + \omega_i s_{x_d}) + \cdots + \beta_p x_{ip} + \sum_{j=1}^{k} \boldsymbol{b}_{ij} \boldsymbol{\gamma}_j$ and $g(\mu_{i\omega}) = \eta_{1i\omega}$. Then, based on this new linear predictor, all the process to obtain the perturbed penalized log-likelihood function and the Q-function remains the same as discussed for each model in the Section 2.2.

As the dispersion/precision parameter has also a linear predictor related to it, the design matrices $\boldsymbol{X}$ (related with mean) and $\boldsymbol{E}$ (related with dispersion/precision) could be functionally dependents. Then, there are some cases to be considered when we add some perturbation on $\boldsymbol{X}$ that could interfere with $\boldsymbol{E}$ also. These cases are described below with the implications related to each one of them.

Case: X and E are functionally independent

In this case, $\phi_{i,\omega} = \phi_i$. Hence, considering $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$, the elements of $\boldsymbol{\Delta}$ for the Beta-SPAM and the Simplex-SPAM, are given by

$$\begin{aligned}
\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -s_{x_d}[\hat{\beta}_d \boldsymbol{X}^\top \hat{\boldsymbol{Q}} - \mathcal{J}\hat{\boldsymbol{T}}_1 \text{diag}\{\hat{\boldsymbol{f}}^*\}],\\
\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -s_{x_d}\hat{\beta}_d \boldsymbol{B}_j^\top \hat{\boldsymbol{Q}}, \quad j = 1, \ldots, k,\\
\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\beta}_d s_{x_d} \hat{\boldsymbol{E}}^\top \hat{\boldsymbol{Q}}^*,
\end{aligned}$$

where β_d is the estimator of the parameter related to the d-th covariate, $\mathcal{J}$ is a $p \times n$ matrix of zeros except for the d-th line, which contains ones. For the BR-SPAM, the above elements are given by

$$\begin{aligned}
\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= s_{x_d}[\hat{\beta}_d \boldsymbol{X}^\top \hat{\boldsymbol{R}} + \mathcal{J}\hat{\boldsymbol{T}}_1 \text{diag}\{\hat{\boldsymbol{f}}^*\}],\\
\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\beta}_d s_{x_d} \boldsymbol{B}_j^\top \hat{\boldsymbol{R}}, \quad j = 1, \ldots, k,\\
\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\beta}_d s_{x_d} \hat{\boldsymbol{E}}^\top \hat{\boldsymbol{R}}^*,\\
\left.\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial \alpha \partial \boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\beta}_d s_{x_d} \hat{\boldsymbol{T}}_1 \text{diag}\{\hat{\boldsymbol{d}}^*\}.
\end{aligned}$$

Case: X and E are equal

When $\boldsymbol{X} = \boldsymbol{E}$, we have that $\eta_{2i\omega} = \kappa_1 + \kappa_2 x_{i2} + \cdots + \kappa_d(x_{id} + \omega_i s_{x_d}) + \cdots + \kappa_p x_{ip}$, the i-th precision parameter, $\phi_{i\omega}$, is such that $g_2(\phi_{i\omega}) = \eta_{2i\omega}$, Additionally, for the Beta-

SPAM and the Simplex-SPAM, we have that

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{X}^\top(\hat{\beta}_d\hat{\boldsymbol{Q}}+\hat{\boldsymbol{Q}}^*\hat{\kappa}_d)-\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}],$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{B}_j^\top(\hat{\beta}_d\hat{\boldsymbol{Q}}+\hat{\boldsymbol{Q}}^*\hat{\kappa}_d)-\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}], \quad j=1,\ldots,k,$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{E}^\top(\hat{\kappa}_d\hat{\boldsymbol{S}}+\boldsymbol{Q}^*\hat{\beta}_d)-\mathcal{J}\hat{\boldsymbol{T}}_2\text{diag}\{\boldsymbol{a}\}],$$

where κ_d is the estimator of the parameter related to the d-th covariate, and for the BR-SPAM, we have that

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{X}^\top(\hat{\beta}_d\hat{\boldsymbol{R}}-\hat{\boldsymbol{R}}^*\hat{\kappa}_d)+\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}],$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{B}_j^\top(\hat{\beta}_d\hat{\boldsymbol{R}}-\hat{\boldsymbol{R}}^*\hat{\kappa}_d)+\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}], \quad j=1,\ldots,k,$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{E}^\top(\hat{\kappa}_d\hat{\boldsymbol{S}}-\hat{\boldsymbol{R}}^*\hat{\beta}_d)+\mathcal{J}\hat{\boldsymbol{T}}_2\text{diag}\{\boldsymbol{a}\}],$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\alpha\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\hat{\beta}_d\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{s}}^*\}-\hat{\kappa}_d\hat{\boldsymbol{T}}_2\text{diag}\{\hat{\boldsymbol{c}}^*\}].$$

Case: X and E share some covariate

When some perturbed covariates in $\boldsymbol{X}$ are also present in $\boldsymbol{E}$ two approaches can be considered. First, it can be considered that $e_{id'} = x_{id}$, for some pair (d, d') with $d = 2, \ldots, p$ and $d' = 2, \ldots, s$. In this case, $\eta_{2i\omega} = \kappa_1 + \kappa_2 n_{i2} + \cdots + \kappa_{d'}(x_{id} + \omega_i s_{x_d}) + \cdots + \kappa_p n_{ip}$ and, for the Beta-SPAM and the Simplex-SPAM, we have that

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{X}^\top(\hat{\beta}_d\hat{\boldsymbol{Q}}+\hat{\boldsymbol{Q}}^*\hat{\kappa}_{d'})-\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}],$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{B}_j^\top(\hat{\beta}_d\hat{\boldsymbol{Q}}+\hat{\boldsymbol{Q}}^*\hat{\kappa}_{d'})-\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}], \quad j=1,\ldots,k,$$

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{E}^\top(\hat{\kappa}_{d'}\hat{\boldsymbol{S}}+\boldsymbol{Q}^*\hat{\beta}_d)-\mathcal{P}\hat{\boldsymbol{T}}_2\text{diag}\{\hat{\boldsymbol{a}}\}],$$

where $\mathcal{P}$ is a $s \times n$ matrix of zeros except for the d'-th line, which corresponds to a unity vector, for $d' = 1, \ldots, s$. Furthermore, for the BR-SPAM, we have that

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{X}^\top(\hat{\beta}_d\hat{\boldsymbol{R}}-\hat{\boldsymbol{R}}^*\hat{\kappa}_{d'})+\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}],$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{B}_j^\top(\hat{\beta}_d\hat{\boldsymbol{R}}-\hat{\boldsymbol{R}}^*\hat{\kappa}_{d'})+\mathcal{J}\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}], \quad j=1,\ldots,k,$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{E}^\top(\hat{\kappa}_{d'}\hat{\boldsymbol{S}} - \hat{\boldsymbol{R}}^*\hat{\beta}_d) + \mathcal{P}\hat{\boldsymbol{T}}_2\mathrm{diag}\{\hat{\boldsymbol{a}}\}],$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\alpha\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\hat{\beta}_d\hat{\boldsymbol{T}}_1\mathrm{diag}\{\hat{\boldsymbol{s}}^*\} - \hat{\kappa}_{d'}\hat{\boldsymbol{T}}_2\mathrm{diag}\{\hat{\boldsymbol{c}}^*\}],$$

In the second approach, let $e_{id'} = \mathcal{G}(x_{id})$, where $d = 2,\ldots,p$ and $d' = 2,\ldots,s$, such that $\mathcal{G}$ is twice differentiable. In this case, $\eta_{2i\omega} = \kappa_1 + \kappa_2 n_{i2} + \cdots + \kappa_{d'}\mathcal{G}(x_{id} + \omega_i s_{x_d}) + \cdots + \kappa_p n_{ip}$ and for the Beta-SPAM and the Simplex-SPAM, we have that

$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{X}^\top(\hat{\beta}_d\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^*\dot{\boldsymbol{G}}\hat{\kappa}_{d'}) - \mathcal{J}\hat{\boldsymbol{T}}_1\mathrm{diag}\{\hat{\boldsymbol{f}}^*\}],$$
$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{B}_j^\top(\hat{\beta}_d\hat{\boldsymbol{Q}} + \hat{\boldsymbol{Q}}^*\dot{\boldsymbol{G}}\hat{\kappa}_{d'}) - \mathcal{J}\hat{\boldsymbol{T}}_1\mathrm{diag}\{\hat{\boldsymbol{f}}^*\}], \quad j = 1,\ldots,k,$$
$$\left.\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -s_{x_d}[\boldsymbol{E}^\top(\hat{\kappa}_{d'}\dot{\boldsymbol{G}}\hat{\boldsymbol{S}} + \boldsymbol{Q}^*\hat{\beta}_d) - \mathcal{P}\hat{\boldsymbol{T}}_2\dot{\boldsymbol{G}}\mathrm{diag}\{\boldsymbol{a}\}],$$

where $\dot{\boldsymbol{G}} = \mathrm{diag}\{\partial\mathcal{G}(x_{1d})/\partial x_{1d},\ldots,\partial\mathcal{G}(x_{nd})/\partial x_{nd}\}$. Additionally, for the BR-SPAM, we have that

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{X}^\top(\hat{\beta}_d\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^*\dot{\boldsymbol{G}}\hat{\kappa}_{d'}) + \boldsymbol{J}\hat{\boldsymbol{T}}_1\mathrm{diag}\{\hat{\boldsymbol{\mu}}\}],$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{B}_j^\top(\hat{\beta}_d\hat{\boldsymbol{R}} - \hat{\boldsymbol{R}}^*\dot{\boldsymbol{G}}\hat{\kappa}_{d'}) + \boldsymbol{J}\hat{\boldsymbol{T}}_1\mathrm{diag}\{\hat{\boldsymbol{f}}^*\}], \quad j = 1,\ldots,k,$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\boldsymbol{E}^\top(\hat{\kappa}_{d'}\dot{\boldsymbol{G}}\hat{\boldsymbol{S}} - \hat{\boldsymbol{R}}^*\hat{\beta}_d) + \mathcal{P}\hat{\boldsymbol{T}}_2\dot{\boldsymbol{G}}\mathrm{diag}\{\boldsymbol{a}\}],$$
$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\alpha\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s_{x_d}[\hat{\beta}_d\hat{\boldsymbol{T}}_1\mathrm{diag}\{\hat{\boldsymbol{s}}^*\} - \hat{\kappa}_{d'}\dot{\boldsymbol{G}}\hat{\boldsymbol{T}}_2\mathrm{diag}\{\hat{\boldsymbol{c}}^*\}].$$

2.12 Link functions choice

In this section we present the link functions considered for each parameter. For the dispersion (ϕ) and probabilities of augmented data (p_0 and p_1) we considered the log/sqrt and logit links, respectively, following the McCullagh and Nelder (1989). For the mean we considered: the logit, probit, c-log-clog, loglog and cauchit links. The first two are symmetric, the third and fourth are, respectively, negative and positive skewed and the last one is symmetric with heavy tails. See McCullagh and Nelder (1989) and Koenker and Yoon (2009), for example. In Figure 10 we can see the behavior of these links. The probit goes to zero and one at the same rate. The same occurs with the logit and cauchit, but the respective rates are smaller than those of probit link, being even smaller for the cauchit link. On the other hand, for the cloglog and loglog, those rates are not the same, where for the cloglog the rate of convergence to one is higher than to the zero, occurring the opposite to the loglog.

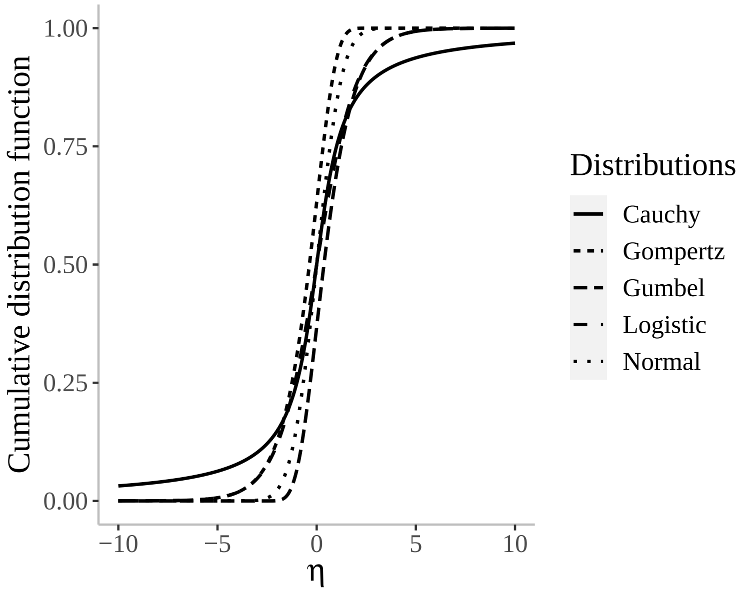


Figure 10 – Comparison among probit, cauchit, logit and cloglog link functions: Plot of their cumulative probability function vs the linear predictor η.

2.13 Computational tools

All methods proposed in this work were implemented in the `R` program (R Core Team, 2020). The Beta-SPAM and Simplex-SPAM were implemented within the R package, `gamlss` (Stasinopoulos et al., 2007), which handles with the Generalized Additive Models for Location, Scale and Shape. The framework of the gamlss accommodates the developed methodology for these two models. In such package it is available the so-called RS algorithm, which allows for fitting models similar to these two, under the presented estimation process.

For the BR-SPAM, the maximization process was done by using the function `optim` of the software `R`, which considers the L-BFGS-B algorithm (Byrd et al., 1995) to obtain the maximum of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$, since the Hessian matrix extracted from the `optim` is a good approximation for the one obtained analytically in the previous section. And the function `nlminb` (Gay, 1990) that uses, which they call as optimization using PORT routines. The last function, present same performance of `optim`, but in some cases, is faster than `optim`. We intend to let all routines available in a `R` package within a near future.

2.14 Simulation Studies

In this section, we present the results of four simulation studies. Section 2.14.1 is related to the a parameter recovery study. In Section 2.14.2, we perform a link misspecification study. In Section 2.14.3, we evaluate the response distribution misspecification. Finally, in Section 2.14.4 is presented a study of the performance of the proposed quantile residuals and the local influence measures. The results for the first study are presented in Appendix, whereas the other studies are available at `https://github.com/aureaflg/Simulations-study.git`, where we present reproducible codes as well as related plots and tables.

2.14.1 Study 1: Parameter recovery

We considered several scenarios of interest defined by the combination of the levels of some important factors. They are (with the respective levels within parenthesis): sample size (n) (50, 100, 500), regression model (beta, simplex, BR), link function (probit, logit, cauchit, cloglog, loglog) and modeled parameter(s) (mean and dispersion). For each scenario, $R = 300$ replicates of Monte Carlo were generated, using a given model (as we explain ahead). The general scenarios are described in Table 2.

Table 2 – Scenarios of simulation study 1

Regression model (modelling mean and dispersion)	n	Link functions
betabeta(μ_i, ϕ_i)	50 100 500	probit, logit, cauchit, cloglog and loglog.
simplexsimplex(μ_i, ϕ_i)	50 100 500	probit, logit, cauchit, cloglog and loglog.
BR(μ_i, ϕ_i, α)	50 100 500	probit, logit, cauchit, cloglog and loglog.

In all scenarios the following structure for linear predictors was considered to both simulated and fit the data sets: $g(\mu_i) = \beta_1 X_i + cos(Z_i)$, $log(\phi_i) = \kappa_0 + \kappa_1 E_i$ for $i = 1, \ldots, n$, where $X_i \overset{iid}{\sim}$ uniform$(0, 1)$, $E_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $Z_i \overset{iid}{\sim}$ uniform$(0, 3\pi)$, $g(.)$ is either the probit, logit, cauchit, cloglog. The actual values for the parameters were: loglog link ($\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$) and other links ($\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$). This change to the loglog link function is justified by the fact that, depending on the parameters chosen, generating from the models using this link function can generate values very close to one or zero, which could compromise the fit of the models to the simulated data even for the BR-SPAM model. For all models, the knots number and the smoothing parameter $\boldsymbol{\lambda}$ were set on 40 and 100, respectively, because these assigned values were efficient in making the fitted curves smooth enough.

The related results are presented in Appendix B.1, where Figures 34 to 38 are related to Beta-SPAM, Figures 39 to 43 are related to Simplex-SPAM, and, Figures 44 to 48 concern to BR-SPAM. Generally, for all models, the parameters were properly recovered, even under small sample sizes. Also, as the sample size increases, the estimates become more accurate. For the BR-SPAM under the cloglog and loglog links, the results were less accurate, compared with the other scenarios. Indeed, the associated bias tend to slightly increase, as the sample size increases, despite of the their low magnitude. However, such issue does not compromise the overall accuracy of the estimates.

2.14.2 Study 2: Link function misspecification

Only the scenarios varying the regression models (beta, simplex, BR) and the link functions (probit, logit, cauchit, cloglog,loglog) are considered. We set the sample size in 500, modeling the mean and the dispersion. We generate only one replica from each model, setting a distribution and a link function for it, and fit the simulated data using a model with the same distribution but varying the link function, using all the different ones from the one used to generate the simulated data. In Table 8 we present the 60 scenarios of interest.

Table 3 – Scenarios of simulation study 2

Distribution that will generate and fit the simulated data	Actual link function	Link function considered in fitted model
beta(μ_i, ϕ_i), simplex(μ_i, ϕ_i) or BR(μ_i, ϕ_i, α)	logit	probit, cauchit, cloglog and loglog
	probit	logit, cauchit, cloglog and loglog
	cauchit	logit, probit, cloglog and loglog
	cloglog	logit, probit, cauchit and loglog
	loglog	logit, probit, cloglog and cauchit

The considered linear predictors and link functions for related scenarios are given by: $g(\mu_i) = \beta_1 x_i + cos(z_i)$, $log(\phi_i) = \kappa_0 + \kappa_1 e_i$, for $i = 1, \ldots, n$, where $X_i \overset{iid}{\sim}$ uniform$(0, 1)$, $E_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $Z_i \overset{iid}{\sim}$ uniform$(0, 3\pi)$, is either the probit, logit, cauchit, cloglog. The actual parameter values considered loglog link were $\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$ and $\alpha = 0.7$, whereas for the other cases were $\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$ and $\alpha = 0.7$. This change considered for loglog link is done, because depend on the actual parameter values, to simulate and adjust considering this link function may be difficult, given that the simulated data are too close of one and zero some times. We considered the accuracy in terms of model prediction and parameter recovery.

In general, for the Beta-SPAM and the Simplex-SPAM the actual mean was properly recovered. The BR-SPAM presented more problems with the misspecification of link function, mainly for the cases in which the cauchit were the link of the fitted model, also in the cases where the model was generated from the cloglog and was fitted with loglog, the points presented a great dispersion. The parameter recovery, in general, was good, when the link function that generating was symmetric, the adjust with loglog and cloglog, which are asymmetric link, did not recovery so well the parameter related with mean. While, that for parameters related with dispersion/precision, they were all well estimated.

2.14.3 Study 3: Response distribution misspecification

In this section we vary only the response distribution, setting the sample size in 100 and the link function as the logit. We modeled only the mean and we generated only R = 1 replica.

Table 4 – Scenarios of simulation study 3

Generating data model ($n = 100$)	Fitted model
beta(μ_i, ϕ)	simplex BR
simplex(μ_i, ϕ)	beta BR
BR(μ_i, ϕ, α)	beta simplex

The considered linear predictor for related scenarios presented in Table 4, have the following structure $g(\mu_i) = \beta_1 x_i + cos(z_i)$ for $i = 1, \ldots, n$, where $X_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $Z_i \overset{iid}{\sim}$ uniform$(0, 3\pi)$, $g(.)$ is the logit. The real values considered for the parameters were $\beta_1 = -1$, $\phi = 20$, $\alpha = 0.1$.

The parameters were well recovered for all model, in general. The difference can be seen, when it is compared the estimated curve for them, in general, the curve is better estimated for the model whose generating it. Note that for this study the value of parameter α of BR was considering very small to the generating data could be adjust for the others models, because when the α increasing, the adjust of others models become poor and for Simplex-SPAM, it is impossible the adjust in same cases. Thus, the conclusion is that when data presents values too close to the limits of the interval $(0, 1)$, the BR-SPAM is the best model, in contrast to when the data is distributed further away from the limits, the models present a similar behaviour.

2.14.4 Study 4: Performance of residuals tools and local influence

To evaluate the performance of the model fit assessment tools, we considered scenarios where we generating data from two regression models (Simplex/Beta SPAM), fixing the link function as the cauchit and the sample size in 100 and, then, we fitting all models (Simplex/Beta/BR SPAM) under the same conditions. For the residual analysis we aim to check if the residuals are able to detect any depart from the model assumptions, mainly the distribution assumption for response variable distribution.

For the local influence analysis, first we generating from on of the three regression model (Simplex/Beta/BR SPAM) and one selected observation was "contaminated" among those that were generated as the response. Then, we fit using the same model that the data came from. For the covariate perturbation scheme, the same process was made, except that now the "contamination" is done in a covariate of a selected observation related with the mean, of the parametric part of the predictor. The "contamination" was done by adding 10 times of the standard deviation of this covariate.

The related models of the scenarios concerning the model fit assessment tools, have this following structure $Y_i \overset{ind}{\sim} \text{beta}(\mu_i, \phi_i)$ or $\text{simplex}(\mu_i, \phi_i)$, $g(\mu_i) = \beta_1 X_i + cos(Z_i)$, $log(\phi_i) = \kappa_0 + \kappa_1 E_i$, for $i = 1, \ldots, n$, where $X_i \overset{iid}{\sim} \text{uniform}(0, 1)$, $E_i \overset{iid}{\sim} \text{uniform}(1, 2)$, $Z_i \overset{iid}{\sim} \text{uniform}(0, 3\pi)$ and $g(.)$ is the cauchit. The real values considered for the parameters were $\beta_1 = -2$, $\kappa_0 = -1$ and $\kappa_1 = 2$.

The related models of the scenarios concerns the local influence analysis, is given by: $Y_i \overset{ind}{\sim} \text{beta}(\mu_i, \phi_i), \text{BR}(\mu_i, \phi_i, \alpha)$ or $\text{simplex}(\mu_i, \phi_i)$, $g(\mu_i) = \beta_1 X_i + cos(Z_i)$, $log(\phi_i) = \kappa_0 + \kappa_1 E_i$, for $i = 1, \ldots, n$, where $X_i \overset{iid}{\sim} \text{uniform}(0, 1)$, $E_i \overset{iid}{\sim} \text{uniform}(0, 1)$ and $Z_i \overset{iid}{\sim} \text{uniform}(0, 3\pi)$, besides that, for each scenario, $g(.)$ is logit. The real values considered for the parameters were $\beta_1 = -2$ $\kappa_0 = 1$ and $\kappa_1 = -4$.

The residuals behave as expected. Indeed when the true underlying and fitted models match, the points are well within the confidence bands, for the QQ plots, and present a random behavior with no outliers, in the index plot. On the other hand, when these two models do not match, the points are no longer well within the confidence bands (QQ-plot) and a systematic behavior can be seen (index plot). Therefore, we have evidence that the developed tools work properly.

For the influence analysis, in both perturbation schemes, the contaminated observations where properly flagged through the $\boldsymbol{l}_{max}$ measure. Therefore, we have indications that the influence analysis tools are useful for such purpose.

2.15 Real data analysis

This data set is related to reading skills of 44 individuals recruited from primary schools in the Australian Capital Territory. They were presented in Smithson and Verkuilen (2006). The response variable is the score on a test of reading accuracy, measured in the (0,1) interval and the higher the score, the better the performance. The covariates are Dyslexia versus Non-dyslexia status (a factor with two levels) and non-verbal IQ converted to z-scores (continuous). This data was previously analyzed by Espinheira et al. (2008) and Ferrari et al. (2011) using beta regression models considering, respectively, with constant and varying dispersion. The data set can be obtained in the `betareg` package available in `R` program (R Core Team, 2020) under the name `ReadingSkills`.

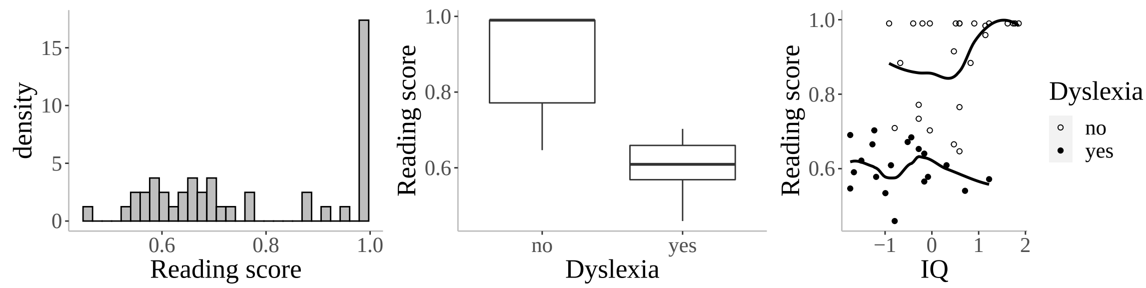


Figure 11 – Explanatory analysis plots of reading skills data set: (1) Histogram of Reading score (2) Boxplot of Reading score by categories of Dyslexia covariate (3) Scatter plot of Reading score versus IQ with two non-parametric curve fitted by LOESS for categories of variable Dyslexia.

In Figure 11, the third plot, which was built by fitting two non-parametric curves for categories of Dyslexia using the method LOESS as an explanatory analysis, suggests a non-linear relation between Reading score and IQ, for both categories of variable Dyslexia. Also, we have that the response variability is related to the Dyslexia, since the boxplot of this covariate suggests a different variability through the categories. We fitted the three following models to the data:

$$\begin{aligned} Y_{il} &\overset{ind}{\sim} \text{beta}(\mu_{il}, \phi_{il})\text{or simplex}(\mu_{il}, \phi_{il}) \text{ or BR}(\mu_{il}, \phi_{il}, \alpha), \\ g_1(\mu_{il}) &= \beta_0 + (\beta_1)_l + h_1(z_i), \\ g_2(\phi_{il}) &= \kappa_0 + (\kappa_1)_l,\ i = 1, \ldots, 44, \end{aligned}$$

where, $l = 1, 2$ and $(\beta_1)_1 = (\kappa_1)_1 = 0$. The parameters (β_1) and (κ_1) are related with Dyslexia ($(\beta_1)_1$ and $(\kappa_1)_1$ related with the kid does not have dyslexia) and $\boldsymbol{z}$ is the variable QI. Also, $h_1(.)$ is an unknown function, which will be approximated by P-splines with cubic B-splines and considering quadratic differences for the penalty function with 40 knots. The choice for $g_1(.)$ will be made by comparing the information criteria for the models. For $g_2(.)$ was adopted the log-link for the Beta-SPAM and BR-SPAM, and the square root for the Simplex-SPAM, because this last model presented better results using this link function. For the Beta-SPAM and Simplex-SPAM all covariates were significant to 5% level, while under the BR-SPAM, we reduce the original structure to the:

$$\begin{aligned} Y_{il} &\overset{ind}{\sim} \text{BR}(\mu_{il}, \phi_{il}, \alpha), \\ g_1(\mu_{il}) &= h_1(z_i), \\ g_2(\phi_{il}) &= \kappa_0 + (\kappa_1)_l,\ i = 1, \ldots, 44, \end{aligned}$$

where, $l = 1, 2$ and $(\boldsymbol{\kappa}_1)_1 = 0$.

All information criteria (IC), except SABIC, see Table 5 indicate, for each distribution, that the cauchit link is the best one. Therefore we compared the three respective models (Beta-SPAM, Simplex-SPAM and BR-SPAM), using the model fit assessment tools. Even though the IC indicates the BR-SPAM as the best model, the related QQ plot with envelope for the quantile residuals indicates the Beta-SPAM as the one with the best fit (see Figure 12). We decided to select the model based on the QQ plot, then, the model selected were Beta-SPAM

Table 5 – Information Criteria for the three models.

Model	Link function	AIC	AICC	SABIC	HQIC	BIC
BR-SPAM	cauchit	**-190.19**	**-175.24**	**-216.31**	**-180.81**	**-164.88**
	logit	-43.15	-29.85	-68.02	-34.21	-19.04
	probit	-41.41	-22.96	-69.85	-31.19	-13.84
	cloglog	-78.39	-57.09	-108.49	-67.57	-49.21
	loglog	10.57	20.74	-11.62	18.55	32.08
Beta-SPAM	cauchit	**-109.58**	**-96.47**	-127.71	**-100.69**	**-85.61**
	logit	-108.19	-94.41	-126.70	-99.12	-83.72
	probit	-106.57	-86.80	**-128.00**	-96.06	-78.23
	cloglog	-105.38	-84.64	-127.22	-94.67	-76.51
	loglog	-108.21	-92.57	-127.71	-98.65	-82.43
Simplex-SPAM	cauchit	**-125.78**	**-72.64**	-156.44	**-110.75**	**-85.25**
	logit	-120.03	-24.93	-156.53	-102.14	-71.78
	probit	-118.25	4.37	**-157.26**	-99.13	-66.69
	cloglog	-117.33	15.61	-157.11	-97.82	-64.73
	loglog	-120.03	-49.71	-153.49	-103.62	-75.78

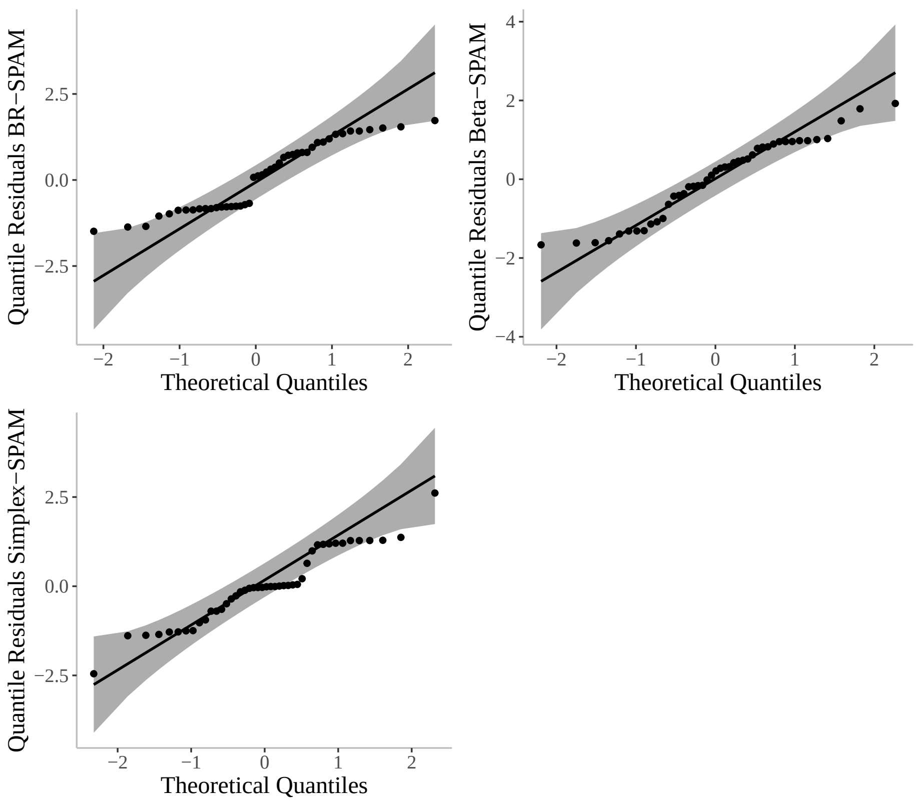


Figure 12 – Quantile-quantile plot with 99% envelopes for all models.

Table 6 – Inferential results for the Beta-SPAM.

Parameters	Estimate	SE	p-value	$IC_{95\%}$
β_0	2.94	0.73	0.0001	[1.52;4.36]
$(\beta_1)_2$	-2.61	0.72	0.0003	[-4.02;-1.21]
κ_0	1.76	0.31	<0.0001	[1.15;2.38]
$(\kappa_1)_2$	3.19	0.45	<0.0001	[2.31;4.07]

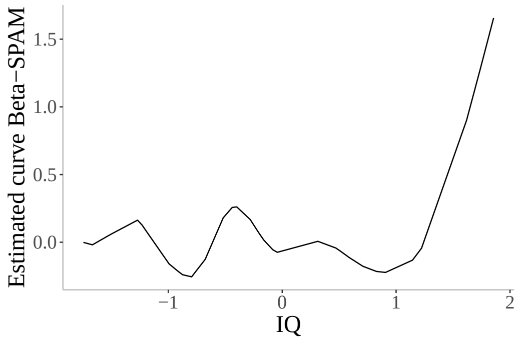
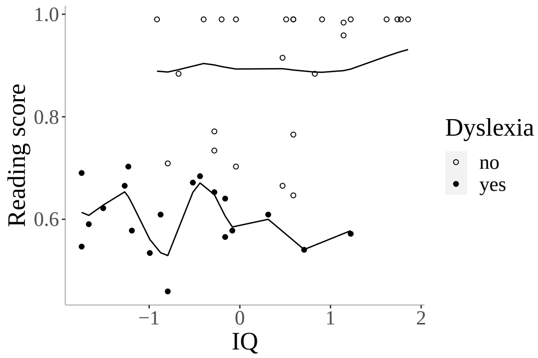


Figure 13 – Estimated curve for Beta-SPAM.

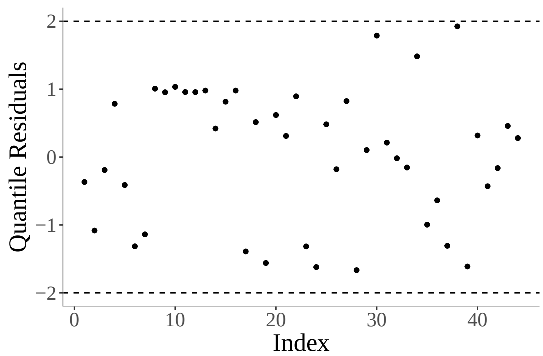
Figure 13 shows the fitted curve for non-parametric part. The effective degrees of freedom were 10.43382 Beta-SPAM, with the smooth parameter, $\lambda = 10$. Some adjustment results can be seen in Figure 14. First, the adjusted mean for model Beta-SPAM is shown in the plot 14a with the variables Reading score and IQ by Dyslexia. All the residuals are well within the the confidence bands of the QQ plot, even though there is some no random pattern. The other residuals plots show no tendency (Figure 14b) as well the normality seems to be reasonable according to the histogram (Figures 14d).

The results of Cook's distance and Generalized Leverage are presented in Figure 15. We can see that the observations #31,#32 and #33 were flagged as high leverage, while the observations #14,#18,#20 and #25 , were flagged as high Cook's distance. Concerning the local influence analysis, from Figure 16, for the response variable perturbation scheme, no observations were flagged. However, for the case-weight perturbation scheme, the observations #38 e #39 were flagged.

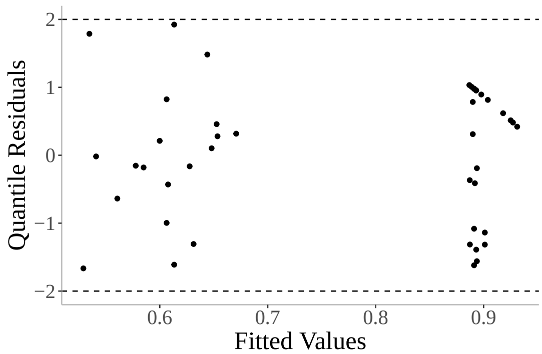
After removing the observations #31,#32 and #33, no one of the following quantities change: significance of the parameters, Cook's distance and the generalized leverage. Also, the behavior of the residuals remained the same. Removing the observations #14,#18,#20 and #25, the significance of the parameters did not change as well as the behavior of the residuals. However, the Cook's distance and the generalized leverage flagged some observations that had not been flagged before. Finally, the exclusion of the observations #38 e #39, did not affect neither the significance of the regression parameters, neither the behavior of the residuals. However, according to the local influence analysis, a new observations appeared as influential. That is, the first set of influential observations,



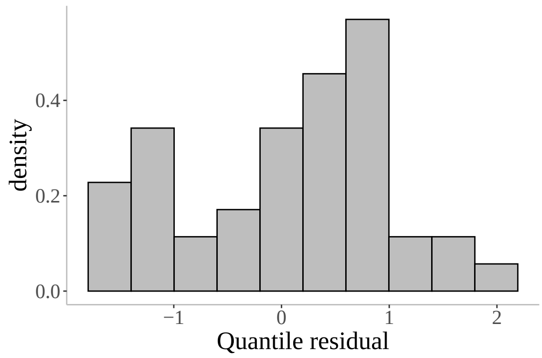
(a) Observed and predicted response mean as a function of IQ.



(b) Index plot of quantile residuals.



(c) Fitted values versus quantile residuals.



(d) Histogram of quantile residuals.

Figure 14 – Results of fitted Beta-SPAM.

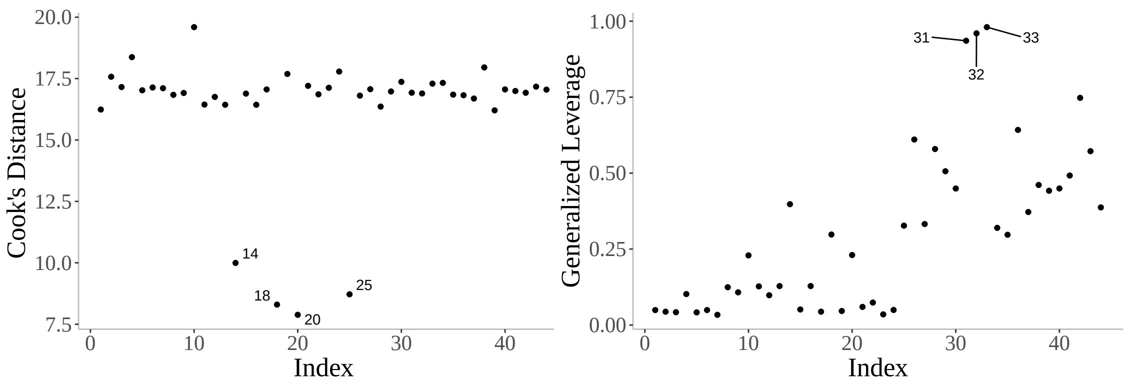


Figure 15 – Cook's Distance and Generalized Leverage measure for the Beta-SPAM.

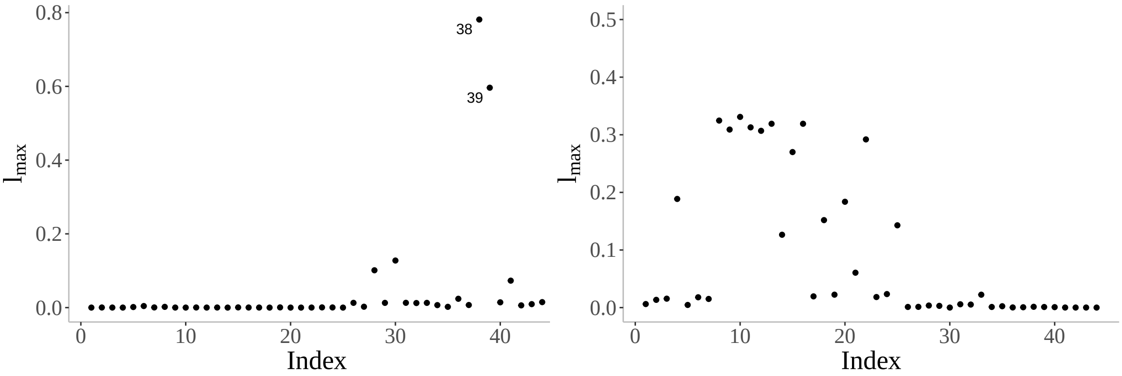


Figure 16 – Index plot of $\boldsymbol{l}_{max}$ under case-weight perturbation (left plot) and under response perturbation (right plot) for Beta-SPAM.

somehow, was probably masking the behavior of this new set. In conclusion, once that the exclusion of the flagged observations did not lead to significant changes on the model fit, no further changes are necessary.

From the selected model we infer that the kids with dyslexia have lower Reading scores than the kid without this condition. Also, it indicates that the Dyslexia variable, is significantly different between the categories for the dispersion parameter.

3 Semi-parametric Additive Models for independent and augmented data

In this Chapter we present extensions of the models developed in Chapter 2, allowing the response belongs to the [0,1] interval (as well as the respective particular cases). The regression structures and probability distributions are those presented before. Therefore, the model class developed in the previous Chapter is a particular case of this developed in this Chapter.

3.1 Models

Let us consider, $Y_1, \ldots, Y_n$, such that $Y_i \overset{ind}{\sim}$ ZOAB$(p_{0i}, p_{1i}, \mu_i, \phi_i)$, or ZOAS$(p_{0i}, p_{1i}, \mu_i, \phi_i)$ or ZOABR$(p_{0i}, p_{1i}, \mu_i, \phi_i, \alpha)$, $i = 1, \ldots, n$. In addition, we assume that

$$\begin{aligned} \eta_{1i} &= g_1(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{k} h_j(\boldsymbol{z}_j), \\ \eta_{2i} &= g_2(\phi_i) = \boldsymbol{e}_i^\top \boldsymbol{\kappa}, \\ (\zeta_{0i}, \zeta_{1i}) &= H(p_{0i}, p_{1i}) = (h_0(p_{0i}, p_{1i}), h_1(p_{0i}, p_{1i})), \end{aligned} \tag{3.1}$$

where $\mu_i = \mathbb{E}(Y_i|Y_i \in (0,1))$, $p_{0i} = \mathbb{P}(Y_i = 0)$, $p_{1i} = \mathbb{P}(Y_i = 1)$ and $1 - p_{0i} - p_{1i} = \mathbb{P}(Y_i \in (0,1))$. Also, $\eta_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, $\eta_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$, $\zeta_{0i} = \boldsymbol{f}_i^\top \boldsymbol{\rho}$, $\zeta_{1i} = \boldsymbol{m}_i^\top \boldsymbol{\tau}$ are linear predictors, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)^\top \in \mathbb{R}^s$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{s_0})^\top \in \mathbb{R}^{s_0}$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{s_1})^\top \in \mathbb{R}^{s_1}$ are vectors of unknown (regression) parameters. Also, $\boldsymbol{X}_{(n\times p)} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$, $\boldsymbol{E}_{(n\times s)} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)^\top$, $\boldsymbol{F}_{(n\times s_0)} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^\top$ and $\boldsymbol{M}_{(n\times s_1)} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n)^\top$ are known design matrix, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$, $\boldsymbol{e}_i = (n_{i1}, \ldots, n_{is})^\top$, $\boldsymbol{f}_i = (f_{i1}, \ldots, f_{is_0})^\top$, $\boldsymbol{m}_i = (m_{i1}, \ldots, m_{is_1})^\top$, $i = 1, \ldots, n$. Finally $\boldsymbol{z}_j = (z_{1j}, \ldots, z_{nj})$, $j = 1, \ldots, k$ is a specification vector associated with non-parametric components and $h_1, \ldots, h_k$ are unknown smooth functions.

We also assume that $g_1(.)$ and $g_2(.)$ are strictly monotone and twice differentiable link functions, such that, $g_1 : (0,1) \to \mathbb{R}$ and $g_2 : \mathbb{R}^+ \to \mathbb{R}$. While H is a bijective transformation of the set $\mathbb{C} = \{(p_{0i}, p_{1i}) : 0 < p_{0i} < 1, 0 < p_{1i} < 1 - p_{0i}\}$ onto $\mathbb{R}^2$, twice differentiable in each component. Under the related assumptions to H, it follows that the partial derivatives of $p_{0i} = h_0^*(\zeta_{0i}, \zeta_{1i})$ and $p_{1i} = h_1^*(\zeta_{0i}, \zeta_{1i})$ are continuous in $\mathbb{R}^2$ and p_{0i}, p_{1i} can be written in terms of ζ_{0i} and ζ_{1i}, uniquely (Rudin, 1976). Following Ospina (2008), we can consider H such that

$$H(p_{0i}, p_{1i}) = (h_0(p_{0i}, p_{1i}), h_1(p_{0i}, p_{1i})) = \left(h\left(\frac{p_{0i}}{1 - p_{0i} - p_{1i}}\right), h\left(\frac{p_{1i}}{1 - p_{0i} - p_{1i}}\right) \right),$$

where $h : \mathbb{R}^+ \to \mathbb{R}$ is strictly monotone and twice differentiable. Notice also that h_0 and h_1 are functions such that, $\mathbb{R}^2 \to \mathbb{R}$. Hence, following Ospina (2008) and setting $h = \log(.)$, it results that

$$\frac{\mathbb{P}(Y_i = 0)}{\mathbb{P}(Y_t \in (0,1))} = \frac{p_{0i}}{1 - p_{0i} - p_{1i}} = \exp(\zeta_{0i}),$$
$$\frac{\mathbb{P}(Y_i = 1)}{\mathbb{P}(Y_t \in (0,1))} = \frac{p_{1i}}{1 - p_{0i} - p_{1i}} = \exp(\zeta_{1i}),$$

leading to,

$$p_{0i} = \mathbb{P}(Y_i = 0) = \frac{e^{\zeta_{0i}}}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}}, \quad p_{1i} = \mathbb{P}(Y_i = 1) = \frac{e^{\zeta_{1i}}}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}},$$
$$1 - p_{0i} - p_{1i} = \mathbb{P}(Y_i \in (0,1)) = \frac{1}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}}. \tag{3.2}$$

Furthermore, we have

$$h_j(\boldsymbol{z}_j) = \sum_{l=1}^{q_j} \gamma_{lj} b_{lj}(\boldsymbol{z}_j), \tag{3.3}$$

where γ_{lj} are the coefficients to be estimated, b_{lj} are the l-th cubic B-spline related to the j-th P-spline, evaluated at $\boldsymbol{z}_j$, $j = 1, \dots, k$ and $l = 1, \dots, q_j$. Since we adopted cubic B-splines, we set $b_{lj} = B_{l,d_l}$ with $d_l = 4$, as defined in Section 1.2.1, $k_j = d_l + q_j$ is the number of knots. Furthermore, it is possible to rewrite Equation (3.3) to simplify the notation, for the i-th observation, $i = 1, \dots, n$, such that $\eta_{1i} = g_1(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{b}_{i1}^\top \boldsymbol{\gamma}_1 + \dots + \boldsymbol{b}_{ik}^\top \boldsymbol{\gamma}_k$, where $\boldsymbol{b}_{ij} = (b_{1j}(z_{ij}), \dots, b_{q_j j}(z_{ij}))^\top$ and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \dots, \gamma_{jq_j})^\top$. Also, we have that

$$\boldsymbol{\eta}_1 = g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j, \tag{3.4}$$

where $\boldsymbol{B}_{j(n\times q_j)}$ has elements $b_{lj}(z_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, k$ and $l = 1, \dots, q_j$.

The adopted augmented distributions admit a finite mixture framework in terms of the respective non-augmented and the Bernoulli distributions. Indeed assuming Y_i, which is the response variable for the i-th subject, we have that

$$z_i^* = \begin{cases} 0, & \text{if } y_i \in (0,1) \\ 1, & \text{if } y_i \in \{0,1\} \end{cases} \tag{3.5}$$

Thus, when $Z_i^* = 0$, $Y_i \sim \text{beta}(\mu_i, \phi_i)/\text{simplex}(\mu_i, \phi_i)/\text{BR } (\mu_i, \phi_i, \alpha)$, while, when $Z_i^* = 1$, $Y_i \sim \text{Bernoulli}(\eta_i)$, where $\eta_i = p_{1i}/(p_{0i} + p_{1i})$. Additionally $P(Z_i^* = 1) = \mathbb{P}(Y_i = 0 \text{ or } Y_i = 1) = p_{0i} + p_{1i}$, and $\mathbb{P}(Z_i^* = 0) = 1 - p_{0i} - p_{1i}$. Thus, $Z_i^* \sim \text{Bernoulli}(p_{0i} + p_{1i})$.

3.2 Penalized log-likelihood function

The jointly distribution of $(y_i, z_i^*)^\top$ assuming $\boldsymbol{\theta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ for ZOAB-SPAM and ZOAS-SPAM and $\boldsymbol{\theta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$ for ZOABR-SPAM, considering the three distributions, can be written as

$$f(y_i, z_i^*; p_{0i}, p_{1i}, p_{0i}, \varpi_i) = f_1(y_i; \varrho_i)^{z_i^*} f_2(y_i; \varpi_i)^{1-z_i^*} (p_{0i} + p_{1i})^{z_i^*} \times (1 - p_{0i} - p_{1i})^{1-z_i^*} \mathbb{I}_{\{y_i, z_i^*\}}, \tag{3.6}$$

where $\mathbb{I}_{\{y_i,z_i^*\}} = \mathbb{I}_{(0,1)}(y_i)\mathbb{I}_{\{0\}}(z_i^*) + \mathbb{I}_{\{0,1\}}(y_i)\mathbb{I}_{\{1\}}(z_i^*)$ and $f_1(y_i; \eta_i)$ is the probability function of Bernoulli distribution with parameter $\varrho_i = p_{1i}/(p_{0i} + p_{1i})$. Besides that, the ϖ_i and $f_2(y_i; \varpi_i)$ are, for each distribution, defined below:

- For the ZOAB-SPAM: $\varpi_i = (\mu_i, \phi_i)$ and $f_2(y_i; \varpi_i)$ is the beta density as defined in the Equation (1.1);
- For the ZOAS-SPAM: $\varpi_i = (\mu_i, \phi_i)$ and $f_2(y_i; \varpi_i)$ is the simplex density as defined in the Equation (1.6);
- For: ZOABR-SPAM: $\varpi_i = (\mu_i, \phi_i, \alpha)$ and $f_2(y_i; \varpi_i)$ is the beta rectangular (BR) density as defined in the Equation (1.3);

Thus, the likelihood is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i, z_i^*; p_{0i}, p_{1i}, p_{0i}, \mu_i, \phi_i) = L_1(\boldsymbol{\vartheta}) L_2(\boldsymbol{\varphi}), \tag{3.7}$$

where $f(y_i, z_i^*; p_{0i}, p_{1i}, p_{0i}, \varpi_i)$ is given in the Equation (3.6), $\boldsymbol{\theta} = (\boldsymbol{\vartheta}^\top, \boldsymbol{\varphi}^\top)^\top$, where $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$ and $\boldsymbol{\varphi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ are related to the discrete and continuous part of the model, respectively. Also,

$$L_1(\boldsymbol{\vartheta}) = \prod_{i=1}^{n} (p_{0i}^{1-y_i} p_{1i}^{y_i})^{z_i^*} (1 - p_{0i} - p_{1i})^{1-z_i^*}, L_2(\boldsymbol{\varphi}) = \prod_{i=1}^{n} b_{Y_i}^*(y_i; \varpi_i)^{1-z_i^*}, \tag{3.8}$$

where $b_Y^*(y; ., .)$ is either defined: in Equation (1.1) for the ZOAB-SPAM; in Equation (1.6) for the ZOAS-SPAM; in Equation (1.3) for the ZOABR-SPAM. Also, μ_i and (ϕ_i, p_{0i}, p_{1i}) are defined in Equations (3.4) and (3.1), respectively.

From Equation (3.7) the likelihood is separable, in terms of $\boldsymbol{\vartheta}$ and $\boldsymbol{\varphi}$ (Pace and Salvan, 1997). Then, the inference about these parameters can be made independently from each other.

The respective log-likelihood is given by: $l(\boldsymbol{\theta}) = l_1(\boldsymbol{\vartheta}) + l_2(\boldsymbol{\varphi})$, where $l_1(\boldsymbol{\vartheta}) = \sum_{i=1}^{n} l_i(p_{0i}, p_{1i})$ and $l_2(\boldsymbol{\varphi}) = \sum_{i=1}^{n} l_i(\varpi_i)$. For each model we have that

- For the ZOAB-SPAM:

$$l_i(p_{0i}, p_{1i}) = z_i^*[(1-y_i)\log(p_{0i}) + y_i\log(p_{1i})] + (1-z_i^*)\log(1-p_{0i}-p_{1i}),$$
$$l_i(\mu_i, \phi_i) = (1-z_i^*)\{\log(\Gamma(\phi_i)) - \log(\Gamma(\mu_i\phi_i)) - \log(\Gamma([1-\mu_i]\phi_i)) + [\mu_i\phi_i - 1]\log(y_i) + ([1-\mu_i]\phi_i - 1)\log[1-y_i]\};$$

- For the ZOAS-SPAM:

$$l_i(p_{0i}, p_{1i}) = z_i^*[(1-y_i)\log(p_{0i}) + y_i\log(p_{1i})] + (1-z_i^*)\log(1-p_{0i}-p_{1i}),$$
$$l_i(\mu_i, \phi_i) = (1-z_i^*)\left\{-\frac{1}{2}\{\log(2\pi\phi_i) + 3\log[y_i(1-y_i)]\} - \frac{1}{2\phi_i}d(y_i;\mu_i)\right\},$$

with $d(y;\mu)$ defined in Equation (1.7);

- For the ZOABR-SPAM:

$$\begin{aligned} l_i(p_{0i}, p_{1i}) &= z_i^*[(1-y_i)\log(p_{0i}) + y_i\log(p_{1i})] + (1-z_i^*)\log(1-p_{0i}-p_{1i}), \\ l_i(\mu_i, \phi_i, \alpha) &= (1-z_i^*)\log\left\{\varepsilon_i + (1-\varepsilon_i)b_{Y_i}(y_i;\delta_i,\phi_i)\right\}, \end{aligned} \tag{3.9}$$

$$\varepsilon_i = 1 - \sqrt{1-4\alpha\mu_i(1-\mu_i)},\ \delta_i = \frac{\mu_i - \frac{1}{2} + \frac{1}{2}\sqrt{1-4\alpha\mu_i(1-\mu_i)}}{\sqrt{1-4\alpha\mu_i(1-\mu_i)}};$$

where $\mu_i = g_1^{-1}(\eta_{1i})$, $\eta_{1i} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{k}\boldsymbol{b}_{ij}^\top\boldsymbol{\gamma}_j$, $\phi_i = g_2^{-1}(\eta_{2i})$, $\eta_{2i} = \boldsymbol{e}_i^\top\boldsymbol{\kappa}$, and (p_{0i}, p_{1i}) is defined in the Equation (3.2). As discussed in the Section 2.2, a penalty function has to be adding in the log-likelihood to guarantee the smoothness of the P-spline that will approximate the non-parametric component $h_1(.)$. The penalized log-likelihood considering the differences penalty (see Equation (1.11)) is given by

$$l_p(\boldsymbol{\theta},\boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \sum_{j=1}^{k}\frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top\boldsymbol{\Lambda}_j^d\boldsymbol{\gamma}_j = l_1(\boldsymbol{\vartheta}) + l_2(\boldsymbol{\varphi}) - \sum_{j=1}^{k}\frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top\boldsymbol{\Lambda}_j^d\boldsymbol{\gamma}_j = l_1(\boldsymbol{\vartheta}) + l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda}), \tag{3.10}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^\top$ is the vector of smooth parameters, $\lambda_j > 0$, $\forall j, j = 1, \ldots, k$, and $l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda}) = l_2(\boldsymbol{\varphi}) - \sum_{j=1}^{k}\frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top\boldsymbol{\Lambda}_j^d\boldsymbol{\gamma}_j$. Since the semi-parametric predictor appears only in the continuous part, the penalization is not necessary for the discrete part. For ease of notation we define $\boldsymbol{\Lambda}_j = \boldsymbol{\Lambda}_j^d,\ j = 1, \ldots, k$.

3.2.1 Complete log-likelihood for ZOABR-SPAM

In order to make calculations simpler it will be used the EM algorithm (Dempster et al., 1977) to obtain the maximum likelihood estimates, as made in the Section

2.2.1, given that ZOABR-SPAM is an extension of the model BR-SPAM. All the calculations made there (Section 2.2.1) will be repeated here particularizing for the case of ZOABR-SPAM.

Hence, given the model defined in Section 3.1, in which $Y_i \sim \text{ZOABR}(p_{0i}, p_{1i}, \mu_i, \phi_i, \alpha)$ for $i = 1, \ldots, n$ and from definition of variable Z^* in Equation (3.5), if $Z_i^* = 0$, Y_i following a $\text{BR}(\mu_i, \phi_i, \alpha)$, which density mixture of beta distribution and standard uniform distribution. Thus, let V_i be a non-observed variable for $i = 1, \ldots, n$, such that

$$v_i = \begin{cases} 0, & \text{if } Y_i \sim \text{beta}(\delta_i, \phi_i), \text{ with probability } 1 - \varepsilon_i \\ 1, & \text{if } Y_i \sim \text{uniform}(0, 1), \text{ with probability } \varepsilon_i. \end{cases}$$

Thus, $V_i \sim \text{Bernoulli}(\varepsilon_i)$, where $\varepsilon_i = 1 - \sqrt{1 - 4\alpha\mu_i(1 - \mu_i)}$. The jointly distribution of $\boldsymbol{Y}_i^c = (Y_i, V_i, Z_i^*)$ conditional on $\boldsymbol{\theta} = (\boldsymbol{\vartheta}^\top, \boldsymbol{\varphi}^\top)^\top$, where $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$ the subvector of $\boldsymbol{\theta}$ related to the discrete part of the model and $\boldsymbol{\varphi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$, related to the continuous part,

$$f(\boldsymbol{y}_i^c|\boldsymbol{\theta}) = (p_{0i}^{1-y_i} p_{1i}^{y_i})^{z_i^*}(1 - p_{0i} - p_{1i})^{1-z_i^*}[\varepsilon_i^{v_i}(1 - \varepsilon_i)^{1-v_i} b(y_i; \delta_i, \phi_i)^{1-v_i}]^{1-z_i^*} \mathbb{I}_{\{y_i, v_i, z_i^*\}}, \tag{3.11}$$

where, $\mathbb{I}_{\{y_i,v_i,z_i^*\}} = \mathbb{I}_{(0,1)}(y_i)\mathbb{I}_{\{0,1\}}(v_i)\mathbb{I}_{\{0\}}(z_i^*) + \mathbb{I}_{\{1\}}(z_i^*)\mathbb{I}_{\{0,1\}}(y_i)$.

The complete likelihood is given by

$$L^c(\boldsymbol{\theta}; \boldsymbol{y}^c) = \prod_{i=1}^n f(y_i, v_i, z_i^*|\boldsymbol{\theta}) = L_1(\boldsymbol{\vartheta}) L_2^c(\boldsymbol{\varphi}),$$

with $f(y_i, v_i, z_i^*|\boldsymbol{\theta})$ is the jointly distribution defined in Equation (3.11), besides that $L_1(\boldsymbol{\vartheta})$ is defined in Equation (3.8) and $L_2^c(\boldsymbol{\varphi})$ is defined below:

$$L_2^c(\boldsymbol{\varphi}) = \prod_{i=1}^n [\epsilon_i^{v_i}(1 - \epsilon_i)^{1-v_i} b_{Y_i}(y_i; \mu_i, \phi_i)]^{1-z_i^*},$$

the complete log-likelihood associated is $l^c(\boldsymbol{\theta}; \boldsymbol{y}_c) = l_1(\boldsymbol{\vartheta}) + l_2^c(\boldsymbol{\varphi})$, where $l_1(\boldsymbol{\vartheta})$ defined in Equation (3.9) and $l_2^c(\boldsymbol{\varphi}) = \sum_{i=1}^n l_i^*(\boldsymbol{\theta})$, with $l_i^*(\boldsymbol{\theta}) = (1 - z_i^*)\{v_i \log \varepsilon_i + (1 - v_i)\log(1 - \varepsilon_i) + (1 - v_i)[\log(\Gamma(\phi_i)) - \log(\Gamma(\delta_i\phi_i)) - \log(\Gamma((1 - \delta_i)\phi_i)) + (\delta_i\phi_i - 1)\log y_i + ((1 - \delta_i)\phi_i - 1)\log(1 - y_i)]\}$. Now, adding a likelihood penalty with respect to non-parametric component to control the smoothness in complete log-likelihood, it leads to

$$l_p^c(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \sum_{j=1}^k \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j = l_1(\boldsymbol{\vartheta}) + l_2^c(\boldsymbol{\varphi}) - \sum_{j=1}^k \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j = l_1(\boldsymbol{\vartheta}) + l_{2p}^c(\boldsymbol{\varphi}, \boldsymbol{\lambda}), \tag{3.12}$$

where $l_{2p}^c(\boldsymbol{\varphi}, \boldsymbol{\lambda})$ is the logarithm of complete likelihood related to continuous part of the model already with the penalty.

3.3 Penalized score and gradient functions

Since the likelihood is separable we can get, independently, the score and Hessian components for the discrete and continuous parts, as we show below.

3.3.1 Penalized score function for discrete part

Since the discrete part for the three models is the same, the respective score and Hessian are the same as well. Indeed, from Equation (3.10), we have that the score vector is given by $\dot{\boldsymbol{U}}_p^{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta})^\top\right)^\top$, where (see the Appendix A.4.1 for demonstration of results related to the discrete part):

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta}) = \frac{\partial l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\rho}} = \boldsymbol{F}^\top \boldsymbol{d}_1 \quad \dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = \frac{\partial l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau}} = \boldsymbol{M}^\top \boldsymbol{d}_2, \tag{3.13}$$

$j = 1, \ldots, k$, $\boldsymbol{d}_1 = (d_{11}, \ldots, d_{1n})$, $\boldsymbol{d}_2 = (d_{21}, \ldots, d_{2n})$, where, respectively, $d_{1i} = z_i^*(1 - y_i) - p_{0i}$ and $d_{2i} = z_i^* y_i - p_{1i}$, for $i = 1, \ldots, n$.

3.3.2 Penalized score function for continuous part of models

From Equation (3.10) we have, for the ZOAB-SPAM and ZOAS-SPAM, that $\dot{\boldsymbol{U}}_p^{\boldsymbol{\varphi}}(\boldsymbol{\theta}) = (\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top)^\top$. On the other hand, for the ZOABR-SPAM, from Equation (3.12), we have that, $\dot{\boldsymbol{U}}_p^{\boldsymbol{\varphi}}(\boldsymbol{\theta}) = (\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top,$ $\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top, \dot{U}_p^{\alpha}(\boldsymbol{\theta}))^\top$. The model-specific components are presented in the following subsections.

3.3.2.1 ZOAB-SPAM

The respective elements of the score function for the ZOAB-SPAM are given by,

$$\begin{aligned}
\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*, \\
\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j} = \boldsymbol{B}_j^\top \boldsymbol{T}_1^* \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j, \\
\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a},
\end{aligned} \tag{3.14}$$

where $j = 1, \ldots, k$, $\boldsymbol{T}_1 = \text{diag}\{(\partial g_1(\mu_1)/\partial \mu_1)^{-1}, \ldots, (\partial g_1(\mu_n)/\partial \mu_n)^{-1}\}$, $\boldsymbol{f}^* = (f_1^*, \ldots, f_n^*)^\top$, $f_i^* = (1 - z_i^*)\phi_i(y_i^* - \mu_i^*)$, $\mu_i^* = [\Psi(\mu_i \phi_i) - \Psi[(1 - \mu_i)\phi_i]$, $y_i^* = \log[y_i/(1 - y_i)]$, $i = 1, \ldots, n$. Also, $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$,

$$a_i = (1 - z_i^*)\{\mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \Psi[(1 - \mu_i)\phi_i] + \Psi(\phi_i)\}, \ i = 1, \ldots, n,$$

and $\boldsymbol{T}_2 = \text{diag}\{(\partial g_2(\phi_1)/\partial \phi_1)^{-1}, \ldots, (\partial g_2(\phi_n)/\partial \phi_n)^{-1}\}$.

3.3.2.2 ZOAS-SPAM

For the ZOAS-SPAM, we have that

$$\begin{aligned}\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*, \quad \dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j} = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j,\\ \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a},\end{aligned} \tag{3.15}$$

$j = 1,\dots,k$, where $\boldsymbol{T}_1 = \mathrm{diag}\{(\partial g_1(\mu_1)/\partial \mu_1)^{-1},\dots,(\partial g_1(\mu_n)/\partial \mu_n)^{-1}\}$, $\boldsymbol{a} = (a_1,\dots,a_n)^\top$,

$$a_i = (1 - z_i^*)\left[-\frac{1}{2\phi_i} + \frac{1}{2\phi_i^2} d(y_i;\mu_i)\right],\ i = 1,\dots,n,$$

$d(y;\mu)$ is defined in Equation (1.7), $\boldsymbol{T}_2 = \mathrm{diag}\{(\partial g_2(\phi_1)/\partial \phi_1)^{-1},\dots,(\partial g_2(\phi_n)/\partial \phi_n)^{-1}\}$ and $\boldsymbol{f}^* = (f_1^*,\dots,f_n^*\}$, $f_i^* = \phi_i^{-1} u_i(y_i - \mu_i)$, where

$$u_i = \frac{(1 - z_i^*)}{\mu_i(1-\mu_i)}\left[d(y_i;\mu_i) + \frac{1}{\mu_i^2(1-\mu_i)^2}\right],\ i = 1,\dots,n.$$

3.3.2.3 ZOABR-SPAM

Let $\boldsymbol{\varphi}^{\top(m)} = (\hat{\boldsymbol{\beta}}^{\top(m)}, \hat{\boldsymbol{\gamma}}_1^{\top(m)},\dots,\hat{\boldsymbol{\gamma}}_k^{\top(m)}, \hat{\boldsymbol{\kappa}}^{\top(m)}, \hat{\alpha}^{(m)})$ be the estimate of $\boldsymbol{\varphi}$ in the m-th iteration. The conditional expectation with respect to $\boldsymbol{v}$ conditional on $\boldsymbol{y}$, of Equation (3.12) is given by

$$\begin{aligned}Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)}) &= \mathbb{E}[l_{2p}^c(\boldsymbol{\varphi})|\boldsymbol{y},\boldsymbol{z}^*,\hat{\boldsymbol{\varphi}}^{(m)}] = \sum_{i=1}^n \{(1 - z_i^*)\{\hat{v}_i^{(m)} \log \varepsilon_i + (1 - \hat{v}_i^{(m)}) \log(1 - \varepsilon_i)\\ &+ (1 - \hat{v}_i^{(m)})[\log(\Gamma(\phi_i)) - \log(\Gamma(\delta_i \phi_i)) - \log(\Gamma((1 - \delta_i)\phi_i))\\ &+ (\delta_i \phi_i - 1)\log y_i + ((1 - \delta_i)\phi_i - 1)\log(1 - y_i)]\}\} - \sum_{j=1}^k \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j\\ &= \sum_{i=1}^n Q_i(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)}),\end{aligned} \tag{3.16}$$

where

$$\hat{v}_i = \mathbb{E}(V_i|y_i, z_i^*, \hat{\boldsymbol{\varphi}}^{(m)}) = \mathbb{P}(V_i = 1|y_i, z_i^*, \hat{\boldsymbol{\varphi}}^{(m)}) = \left(\frac{\varepsilon_i}{\varepsilon_i + (1 - \varepsilon_i) b(y_i;\delta_i,\phi_i)}\right)^{1 - z_i^*}, \tag{3.17}$$

$\varepsilon_i = 1 - \sqrt{1 - 4\alpha\mu_i(1-\mu_i)}$, $\delta_i = \dfrac{\mu_i - \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\alpha\mu_i(1-\mu_i)}}{\sqrt{1 - 4\alpha\mu_i(1-\mu_i)}}$ and $b(y_i;.,.)$ is given in Equation (1.1).

The penalized gradient vector, $\dot{\boldsymbol{U}}_p^{\varphi}(\boldsymbol{\theta})$ is, the, given by

$$\dot{\boldsymbol{U}}_p^{\varphi}(\boldsymbol{\theta}) = \frac{\partial Q(\boldsymbol{\varphi}|\boldsymbol{\varphi}^{(m)})}{\partial \boldsymbol{\varphi}} = \left(\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top,\dots,\dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top, \dot{U}_p^{\alpha}(\boldsymbol{\theta})\right)^\top,$$

where

$$\begin{aligned} \dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) &= \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*, \quad \dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda_j} \boldsymbol{\gamma}_j, \\ \dot{\boldsymbol{U}}_p^{\kappa}(\boldsymbol{\theta}) &= \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a}, \quad \dot{U}_p^{\alpha}(\boldsymbol{\theta}) = \text{trace}(\boldsymbol{D}), \end{aligned} \tag{3.18}$$

$j = 1, \ldots, k$, where $\boldsymbol{T}_1 = \text{diag}\{(\partial g_1(\mu_1)/\partial \mu_1)^{-1}, \ldots, (\partial g_1(\mu_n)/\partial \mu_n)^{-1}\}$, $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, $a_i = (1-z_i^*)(1-\hat{v}_i)\{\delta_i(y_i^*-\delta_i^*)+\log(1-y_i)-\Psi[(1-\delta_i)\phi_i]+\Psi(\phi_i)\}$, $\delta_i^* = \Psi(\delta_i\phi_i)-\Psi[(1-\delta_i)\phi_i]$ and $y_i^* = \log[y_i/(1-y_i)]$, for $i = 1, \ldots, n$. Also, $\boldsymbol{f}^* = (f_1^*, \ldots, f_n^*)^\top$,

$$\begin{aligned} f_i^* = (1 - z_i^*)\Bigg\{ & \frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] \\ & + (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^* - \delta_i^*)\Bigg\}, \end{aligned}$$

$\boldsymbol{D} = \text{diag}\{d_1, \ldots, d_n\}$,

$$d_i = (1 - z_i^*)\left\{\frac{2\mu_i(1-\mu_i)}{(1-\varepsilon_i)}\left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] - (1-\hat{v}_i)\left[\frac{\phi_i\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right](y_i^* - \delta_i^*)\right\},$$

and $\boldsymbol{T}_2 = \text{diag}\{(\partial g_2(\phi_1)/\partial \phi_1)^{-1}, \ldots, (\partial g_2(\phi_n)/\partial \phi_n)^{-1}\}$. The related demonstration can be found in Appendix A.3.1.

3.4 Penalized Hessian and Fisher information matrices

3.4.1 Penalized Fisher information matrix for discrete part

In order to obtain the Fisher information matrix for $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, (related to the discrete part) we first obtain the second derivatives of $l_1(\boldsymbol{\vartheta})$, that is

$$\ddot{\boldsymbol{U}}_p^{\vartheta\vartheta}(\boldsymbol{\theta}) = \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\rho\rho}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\rho\tau}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\tau\rho}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix}, \tag{3.19}$$

where $\ddot{\boldsymbol{U}}_p^{\rho\tau}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\tau\rho}(\boldsymbol{\theta})^\top$. The respective elements are given by

$$\begin{aligned} \ddot{\boldsymbol{U}}_p^{\rho\rho}(\boldsymbol{\theta}) &= \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^\top} = -\boldsymbol{F}^\top \boldsymbol{D}_1^* \boldsymbol{F}, \quad \ddot{\boldsymbol{U}}_p^{\tau\tau}(\boldsymbol{\theta}) = \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^\top} = -\boldsymbol{M}^\top \boldsymbol{D}_2^* \boldsymbol{M}, \\ \ddot{\boldsymbol{U}}_p^{\rho\tau}(\boldsymbol{\theta}) &= \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\tau}^\top} = -\boldsymbol{F}^\top \boldsymbol{D}_3^* \boldsymbol{M}, \end{aligned}$$

where $\boldsymbol{D}_1^* = \text{diag}\{d_{11}^*, \ldots, d_{1n}^*\}$, $\boldsymbol{D}_2^* = \text{diag}\{d_{21}^*, \ldots, d_{2n}^*\}$ and $\boldsymbol{D}_3^* = \text{diag}\{d_{31}^*, \ldots, d_{3n}^*\}$, where, respectively, $d_{1i}^* = p_{0i}(1-p_{0i}), d_{2i}^* = p_{1i}(1-p_{1i}), d_{3i}^* = -p_{0i}p_{1i}$, for $i = 1, \ldots, n$. The, the Fisher information matrix is given by

$$\boldsymbol{K}_p^{\vartheta\vartheta}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top}\right] = \begin{pmatrix} \boldsymbol{K}_p^{\rho\rho}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\rho\tau}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\tau\rho}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix}, \tag{3.20}$$

where $\boldsymbol{K}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\boldsymbol{\tau\rho}}(\boldsymbol{\theta})^\top$. The respective elements are given below (see Appendix A.4.2 for the related demonstrations).

$$\boldsymbol{K}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}^\top}\right] = \boldsymbol{F}^\top\boldsymbol{D}_{\mathbf{1}}^*\boldsymbol{F}, \quad \boldsymbol{K}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^\top}\right] = \boldsymbol{M}^\top\boldsymbol{D}_{\mathbf{2}}^*\boldsymbol{M},$$
$$\boldsymbol{K}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\rho}\partial\boldsymbol{\tau}^\top}\right] = \boldsymbol{F}^\top\boldsymbol{D}_{\mathbf{3}}^*\boldsymbol{M}.$$

3.4.2 Penalized Hessian and Fisher information matrices for the continuous part

The general form for the respective Hessian matrix for the ZOAB-SPAM and ZOAS-SPAM,

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\varphi\varphi}}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}^\top} = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) \end{pmatrix}, \qquad (3.21)$$

where $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_j}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_j}(\boldsymbol{\theta})^\top$ and $\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_{j'}\gamma_j}(\boldsymbol{\theta})^\top$, for $j = 1, \dots, k$. On the other hand, the respective Hessian matrix for the ZOABR-SPAM, is given by

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\varphi\varphi}}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}^\top} = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\alpha}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\alpha}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\alpha}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\alpha}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\alpha\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\alpha\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\kappa}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\alpha\alpha}(\boldsymbol{\theta}) \end{pmatrix}, \qquad (3.22)$$

where $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_j}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_j}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\gamma_{j'}\gamma_j}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\alpha}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\beta}}(\boldsymbol{\theta})^\top$, $\ddot{\boldsymbol{U}}_p^{\gamma_j\alpha}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\alpha\gamma_j}(\boldsymbol{\theta})^\top$ and $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\alpha}(\boldsymbol{\theta}) = \ddot{\boldsymbol{U}}_p^{\alpha\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top$ for $j = 1, \dots, k$.

Then, the respective Fisher information matrix for the ZOAB-SPAM and

ZOAS-SPAM is given by

$$\boldsymbol{K}_p^{\varphi\varphi}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}^\top}\right] = \begin{pmatrix} \boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\beta\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\beta\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\beta\kappa}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\gamma_1\beta}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\kappa}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{K}_p^{\gamma_k\beta}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\kappa}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\kappa\beta}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\kappa\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\kappa\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\kappa\kappa}(\boldsymbol{\theta}) \end{pmatrix}, \tag{3.23}$$

where $\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\gamma_j\beta}(\boldsymbol{\theta})^\top$, $\boldsymbol{K}_p^{\beta\kappa}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\kappa\beta}(\boldsymbol{\theta})^\top$, $\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\gamma_{j'}\gamma_j}(\boldsymbol{\theta})^\top$ and $\boldsymbol{K}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \boldsymbol{K}_p^{\kappa\gamma_j}(\boldsymbol{\theta})^\top$, for $j = 1,\dots,k$.

3.4.2.1 ZOAB-SPAM

The elements of the Hessian matrix (Equation (3.21)), are given by,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}^\dagger\boldsymbol{X}, \quad \ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}^\dagger\boldsymbol{B}_j, \\
\ddot{\boldsymbol{U}}_p^{\kappa\kappa}(\boldsymbol{\theta}) &= \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{E}^\top\boldsymbol{S}^\dagger\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\beta\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}^{\dagger\dagger}\boldsymbol{E}, \\
\ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) &= \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}^{\dagger\dagger}\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}^\dagger\boldsymbol{B}_{j'}, \\
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}^\dagger\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned} \tag{3.24}$$

$j = 1,\dots,k$, $\boldsymbol{Q}^\dagger = \boldsymbol{K}^*\boldsymbol{Q}$, where $\boldsymbol{Q} = \mathrm{diag}\{q_1,\dots,q_n\}$,

$$q_i = \phi_i\left[\phi_i\{\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)\} + (y_i^* - \mu_i^*)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\right] \times\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2},$$

$\boldsymbol{Q}^{\dagger\dagger} = \boldsymbol{K}^*\boldsymbol{Q}^*$, $\boldsymbol{Q}^* = \mathrm{diag}\{q_1^*,\dots,q_n^*\}$,

$$q_i^* = \{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)] - (y_i^* - \mu_i^*)\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1},$$

$\boldsymbol{S}^\dagger = \boldsymbol{K}^*\boldsymbol{S}$, $\mathrm{diag}\{s_1,\dots,s_n\}$ and

$$s_i = \left[\Psi'(\mu_i\phi_i)\mu_i^2 + \Psi'((1-\mu_i)\phi_i)(1-\mu_i)^2 - \Psi'(\phi_i) + a_i^*\left(\frac{\partial g_2(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial^2 g_2(\mu_i)}{\partial\mu_i^2}\right)\right] \times\left(\frac{\partial g_2(\mu_i)}{\partial\mu_i}\right)^{-2}.$$

Also, $a_i^* = \mu_i(y_i^* - \mu_i^*) + \log(1-y_i) - \Psi[(1-\mu_i)\phi_i] + \Psi(\phi_i)$, $i = 1,\dots,n$ and $\boldsymbol{K}^* = \mathrm{diag}\{(1-z_1^*),\dots,(1-z_n^*)\}$.

On the other hand, the elements of Fisher information, Equation (3.23), are defined below

$$
\begin{aligned}
&\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}^\dagger\boldsymbol{X}, \quad \boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}^\dagger\boldsymbol{B}_j,\\
&\boldsymbol{K}_p^{\kappa\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{E}^\top\boldsymbol{P}^\dagger\boldsymbol{E}, \quad \boldsymbol{K}_p^{\beta\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{X}^\top\boldsymbol{W}^{\dagger\dagger}\boldsymbol{E},\\
&\boldsymbol{K}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}^{\dagger\dagger}\boldsymbol{E}, \quad \boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}^\dagger\boldsymbol{B}_{j'},\\
&\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{B}_j^\top\boldsymbol{W}^\dagger\boldsymbol{B}_j + \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned}
\tag{3.25}
$$

$j = 1,\ldots,k$, $\boldsymbol{W}^\dagger = \boldsymbol{K}^*\boldsymbol{W}$, $\boldsymbol{W} = \mathrm{diag}\{w_1,\ldots,w_n\}$ and

$$
w_i = \phi_i^2[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}.
$$

Also, $\boldsymbol{P}^\dagger = \boldsymbol{K}^*\boldsymbol{P}$, $\boldsymbol{P} = (p_1,\ldots,p_n)^\top$,

$$
p_i = \left[\Psi'(\mu_i\phi_i)\mu_i^2 + \Psi'((1-\mu_i)\phi_i)(1-\mu_i)^2 - \Psi'(\phi_i)\right]\left(\frac{\partial g_2(\mu_i)}{\partial\mu_i}\right)^{-2}.
$$

$\boldsymbol{W}^{\dagger\dagger} = \boldsymbol{W}^*\boldsymbol{K}^*$, where $\boldsymbol{W}^* = \mathrm{diag}\{w_1^*,\ldots,w_n^*\}$, and

$$
w_i^* = \{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)]\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1},
$$

$i = 1,\ldots,n$.

3.4.2.2 ZOAS-SPAM

The elements of the Hessian matrix, Equation (3.21), are defined below.

$$
\begin{aligned}
&\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{X}, \quad \ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{B}_j,\\
&\ddot{\boldsymbol{U}}_p^{\kappa\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{E}^\top\boldsymbol{S}^\dagger\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\beta\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{X}^\top\boldsymbol{Q}^*\boldsymbol{E},\\
&\ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}^*\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_{j'},\\
&\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_{2p}(\boldsymbol{\varphi},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned}
\tag{3.26}
$$

$j = 1,\ldots,k$, $\boldsymbol{Q}^* = (q_1^*,\ldots,q_n^*)^\top$,

$$
q_i^* = [(1/\phi_i^2)u_i(y_i-\mu_i)]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1},
$$

u_i is defined in Equation (3.15), $\boldsymbol{S}^\dagger = \boldsymbol{K}^*\boldsymbol{S}$, $\boldsymbol{S} = \text{diag}\{s_1, \ldots, s_n\}$,

$$s_i = \left[\frac{1}{2\phi_i^2} - \frac{1}{\phi_i^3} d(y_i; \mu_i) + a_i(1 - z_i^*)^{-1} \left(\frac{\partial^2 g_2(\phi_i)}{\partial \phi_i^2}\right) \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-1}\right] \times \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-2}, \; i = 1, \ldots, n,$$

$d(y; \mu)$ is defined in Equation (1.7) and a_i is defined in Equation (3.15). Also, $\boldsymbol{Q} = \text{diag}\{q_1, \ldots, q_n\}$,

$$q_i = \frac{1}{\phi_i}\Bigg\{u_i + (1 - z_i^*)(y_i - \mu_i)\left[\frac{2(y_i - \mu_i)u_i}{\mu_i(1 - \mu_i)} + \frac{3(1 - 2\mu_i)}{\mu_i^4(1 - \mu_i)^4} + \frac{(1 - 2\mu_i)d(y_i; \mu_i)}{\mu_i^2(1 - \mu_i)^2}\right] + u_i(y_i - \mu_i)\left(\frac{\partial^2 g_1(\mu_i)}{\partial \mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-2},$$

and $\boldsymbol{K}^* = \text{diag}\{(1 - z_1^*), \ldots, (1 - z_n^*)\}$.

The respective elements of the Fisher information, Equation (3.23), are given by

$$\begin{aligned}
\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right] = \boldsymbol{X}^\top \boldsymbol{W}^\dagger \boldsymbol{X}, \quad \boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{X}^\top \boldsymbol{W}^\dagger \boldsymbol{B}_j, \\
\boldsymbol{K}_p^{\kappa\kappa}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^\top}\right] = \boldsymbol{E}^\top \boldsymbol{P}^\dagger \boldsymbol{E}, \quad \boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_{j'}^\top}\right] = \boldsymbol{B}_j^\top \boldsymbol{W}^\dagger \boldsymbol{B}_{j'}, \\
\boldsymbol{K}_p^{\gamma_j\phi}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j \partial \phi}\right] = \boldsymbol{0}, \quad \boldsymbol{K}_p^{\beta\phi}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \phi}\right] = \boldsymbol{0}, \\
\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^\top}\right] = \boldsymbol{B}_j^\top \boldsymbol{W}^\dagger \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j,
\end{aligned} \tag{3.27}$$

$j = 1, \ldots, k$, $\boldsymbol{P}^\dagger = \boldsymbol{K}^*\boldsymbol{P}$, $\boldsymbol{P} = \text{diag}\{p_1, \ldots, p_n\}$, $p_i = 1/(2\phi_i^2)$, $i = 1, \ldots, n$, $\boldsymbol{W}^\dagger = \boldsymbol{K}^*\boldsymbol{W}$, $\boldsymbol{W} = \text{diag}\{w_1, \ldots, w_n\}$, and

$$w_i = \frac{1}{\phi_i}\left[\frac{3\phi_i}{\mu_i(1 - \mu_i)} + \frac{1}{\mu_i^3(1 - \mu_i)^3}\right]\left(\frac{\partial g(\mu_i)}{\partial \mu_i}\right)^{-2}, \; i = 1, \ldots, n.$$

3.4.2.3 ZOABR-SPAM

The elements of the diagonal of the Hessian matrix (Equation (2.15)) are defined below.

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \boldsymbol{X}^\top \boldsymbol{R}^\dagger \boldsymbol{X}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^\top} = \boldsymbol{B}_j^\top \boldsymbol{R}^\dagger \boldsymbol{B}_j - \lambda_j \boldsymbol{\Lambda}_j, \\
\ddot{\boldsymbol{U}}_p^{\kappa\kappa}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^\top} = \boldsymbol{E}^\top \boldsymbol{S}^\dagger \boldsymbol{E}, \\
\ddot{\boldsymbol{U}}_p^{\alpha\alpha}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial \alpha^2} = -\text{trace}(\boldsymbol{J}^\dagger),
\end{aligned} \tag{3.28}$$

$j = 1, \ldots, k$, $\boldsymbol{R}^{\dagger} = \boldsymbol{K}^{*}\boldsymbol{R}$, $\boldsymbol{R} = \mathrm{diag}\{r_1, \ldots, r_n\}$,

$$r_i = \left\{ \left[\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^2} \right] \left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2} \right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)} \right] \left[-\frac{4\alpha}{(1-\varepsilon_i)} + \frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^3} \right] \right.$$
$$+ (1-\alpha)(1-\hat{v}_i) \left[\frac{-(1-\alpha)\phi_i^2}{(1-\varepsilon_i)^6} w_i^* + \frac{6\alpha\phi_i(1-2\mu_i)}{(1-\varepsilon_i)^5}(y_i^* - \delta_i^*) \right] - (1-z_i^*)^{-1} f_i^* \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1}$$
$$\left. \times \left(\frac{\partial^2 g_1(\mu_i)}{\partial \mu_i^2} \right) \right\} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-2}, \; i = 1, \ldots, n,$$

f_i^* is defined in Equation (2.13) and $w_i^* = \Psi'(\delta_i\phi_i) + \Psi'((1-\delta_i)\phi_i)$. Also, $\boldsymbol{S}^{\dagger} = \boldsymbol{K}^{*}\boldsymbol{S}$, $\boldsymbol{S} = \mathrm{diag}\{s_1, \ldots, s_n\}$,

$$s_i = \left[-(1-\hat{v}_i)[(1-\delta_i)^2\Psi'((1-\delta_i)\phi_i) + \delta^2\Psi'(\delta_i\phi_i) - \Psi'(\phi_i)] \right.$$
$$\left. - (1-z_i^*)^{-1} a_i \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \left(\frac{\partial^2 g_2(\phi_i)}{\partial \phi_i^2} \right) \right] \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-2},$$

and a_i is defined in Equation (2.13), for $i = 1, \ldots, n$. In addition, $\boldsymbol{J}^{\dagger} = \boldsymbol{K}^{*}\boldsymbol{J}, \boldsymbol{J} = \mathrm{diag}\{j_1, \ldots, j_n\}$,

$$j_i = \left\{ \left[\frac{4\mu_i^2(1-\mu_i)^2}{(1-\varepsilon_i)^2} \right] \left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2} \right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)} \right] \left[\frac{4\mu_i^2(1-\mu_i)^2}{(1-\varepsilon_i)^3} \right] \right.$$
$$\left. - (1-\hat{v}_i) \left[\frac{\phi\mu_i^2(1-\mu_i)^2(1-2\mu_i)^2}{(1-\varepsilon_i)^5} \right] \left[\frac{\phi_i(1-2\mu_i)}{(1-\varepsilon_i)} w_i^* - 6(y_i^* - \delta_i^*) \right] \right\}, \; i = 1, \ldots, n,$$

and $\boldsymbol{K}^{*} = \mathrm{diag}\{(1-z_1^*), \ldots, (1-z_n^*)\}$.

The elements outside the main diagonal of the Hessian matrix (Equation (2.15)) are given by

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\kappa}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^{\top}} = -\boldsymbol{X}^{\top}\boldsymbol{R}^{\dagger\dagger}\boldsymbol{E}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^{\top}} = -\boldsymbol{B}_j^{\top}\boldsymbol{R}^{\dagger\dagger}\boldsymbol{E},\\
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^{\top}} = \boldsymbol{B}_j^{\top}\boldsymbol{R}^{\dagger}\boldsymbol{B}_{j'}, \quad \ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^{\top}} = \boldsymbol{X}^{\top}\boldsymbol{R}^{\dagger}\boldsymbol{B}_j,\\
\ddot{\boldsymbol{U}}_p^{\beta\alpha}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\beta}\partial\alpha} = \boldsymbol{X}^{\top}\boldsymbol{T}_1\boldsymbol{s}^{\dagger}, \quad \ddot{\boldsymbol{U}}_p^{\gamma_j\alpha}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\alpha} = \boldsymbol{B}_j^{\top}\boldsymbol{T}_1\boldsymbol{s}^{\dagger},\\
\ddot{\boldsymbol{U}}_p^{\kappa\alpha}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}}^{(m)})}{\partial\boldsymbol{\kappa}\partial\alpha} = \boldsymbol{E}^{\top}\boldsymbol{T}_2\boldsymbol{c}^{\dagger},
\end{aligned} \tag{3.29}$$

$\boldsymbol{T}_1 = \mathrm{diag}\{(\partial g_1(\mu_1)/\partial\mu_1)^{-1}, \ldots, (\partial g_1(\mu_n)/\partial\mu_n)^{-1}\}$, $\boldsymbol{R}^{\dagger\dagger} = \boldsymbol{K}^{*}\boldsymbol{R}^{*}$, $\boldsymbol{R}^{*} = \mathrm{diag}\{r_1^*, \ldots, r_n^*\}$ and

$$r_i^* = -(1-\hat{v}_i) \left\{ \frac{(1-\alpha)}{(1-\varepsilon_i)^3} [\phi_i(\delta_i w_i^* - \Psi'((1-\delta_i)\phi_i)) - (y_i^* - \delta_i^*)] \right\}$$
$$\times \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1}.$$

Also, $\boldsymbol{T}_2 = \text{diag}\{(\partial g_2(\phi_1)/\partial\phi_1)^{-1}, \ldots, (\partial g_2(\phi_n)/\partial\phi_n)^{-1}\}$, $\boldsymbol{s}^\dagger = \boldsymbol{K}^*\boldsymbol{s}^*$, $\boldsymbol{s}^* = (s_1^*, \ldots, s_n^*)^\top$,

$$\begin{aligned} s_i^* = &\left\{\left[\frac{4\alpha\mu_i(1-\mu_i)(1-2\delta_i)}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right]\right. \\ &+ \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\left\{2(1-2\mu_i)\left[\frac{1}{(1-\varepsilon_i)} + \frac{2\alpha\mu_i(1-\mu_i)}{(1-\varepsilon_i)^3}\right]\right\} \\ &+ (1-\hat{v}_i)\left\{\left[\frac{(1-\alpha)\phi_i^2\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^6}w_i^*\right] + (y_i^* - \delta_i^*)\right. \\ &\left.\left.\times\left\{-\phi\left[\frac{1}{(1-\varepsilon_i)^4} - \frac{6(1-\alpha)\mu_i(1-\mu_i)}{(1-\varepsilon_i)^6}\right]\right\}\right\}\right\} \; i = 1, \ldots, n, \end{aligned}$$

$\boldsymbol{c}^\dagger = \boldsymbol{K}^*\boldsymbol{c}^*$, $\boldsymbol{c}^* = (c_1^*, \ldots, c_n^*)^\top$, and

$$c_i^* = (1-\hat{v}_i)\left\{\left[\frac{\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right][\phi_i(\delta_i w_i^* - \Psi'((1-\delta_i)\phi_i)) - (y_i^* - \delta_i^*)]\right\},$$

for $i = 1, \ldots, n$.

3.5 Parameter estimation

Here we present the estimation process for the three models, separated for each part (discrete and continuous), provided that the likelihood is separable, as shown before.

3.5.1 Parameter estimation of discrete part

The maximum likelihood estimates of $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$ is obtained by solving $\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}} = \dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}} = \boldsymbol{0}$. Since such equations do not have analytical solution, some numerical method should be employed. We adopt the Fisher scoring algorithm, that is

$$\boldsymbol{\vartheta}^{(u+1)} = \boldsymbol{\vartheta}^{(u)} + (\boldsymbol{K}_p^{\boldsymbol{\vartheta}\boldsymbol{\vartheta}})^{-1}\dot{\boldsymbol{U}}_p^{\boldsymbol{\vartheta}}\Big|_{\boldsymbol{\vartheta}^{(u)}},$$

$u = 0, 1, 2, ...$ where $\dot{\boldsymbol{U}}_p^{\boldsymbol{\vartheta}}$ and $\boldsymbol{K}_p^{\boldsymbol{\vartheta}\boldsymbol{\vartheta}}$ are described in Equations (3.14) and (3.20), respectively. These steps are repeated while $||\boldsymbol{\varphi}^{(u+1)} - \boldsymbol{\varphi}^{(u)}|| \geq \epsilon, \epsilon > 0$. As done in Silva et al. (2020), the function `gamlss` of R package `gamlss` (see (Rigby and Stasinopoulos, 2005)) is used to obtain the starting values. Specifically, we use the following sintax: `fit=gamlss(y∼(-1+X),sigma.formula=(-1+E),nu.formula=(-1+F), tau.formula= (-1+M), family=BEINF)`

3.5.2 Parameter estimation of continuous part

ZOAB-SPAM and ZOAS-SPAM

Similarly to the approach presented in Section 2.5, for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j, j = 1, \ldots, k$, we apply a combination of the Fisher scoring with the backfitting

(Gauss–Seidel) (Breiman and Friedman, 1985) algorithms. This combined algorithm is given by

$$\boldsymbol{\beta}^{(u+1)} = (\boldsymbol{X}^\top \boldsymbol{W}^{\dagger(u)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}^{\dagger(u)} \left[\boldsymbol{z}^{**(u)} - \sum_{i=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j^{(u)} \right] \tag{3.30a}$$

$$\boldsymbol{\gamma}_j^{(u+1)} = (\boldsymbol{B}_j^\top \boldsymbol{W}^{\dagger(u)} \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j)^{-1} \boldsymbol{B}_j^\top \boldsymbol{W}^{\dagger(u)} \left[\boldsymbol{z}^{**(u)} - \boldsymbol{X}\boldsymbol{\beta}^{(u+1)} - \sum_{j^* \neq j} \boldsymbol{B}_{j^*} \boldsymbol{\gamma}_{j^*}^{(u+1)} \right], \tag{3.30b}$$

$u = 0, 1, 2, \ldots,\ j = 1, \ldots, k$, where, $\boldsymbol{z}^{**} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j + (\boldsymbol{W}^{\dagger(u)})^{-1} \boldsymbol{T}_1 \boldsymbol{f}^*$. On the other hand, the PMLE of $\boldsymbol{\kappa}$, says $\hat{\boldsymbol{\kappa}}$, can be obtained, following Ibacache-Pulgar and Reyes (2017), that is, using the Fisher scoring algorithm. Then, given $\boldsymbol{\beta}^{(u+1)}, \boldsymbol{\gamma}_1^{(u+1)}, \ldots, \boldsymbol{\gamma}_k^{(u+1)}$, the $(u+1)$-th step of this algorithm, is given by

$$\boldsymbol{\kappa}^{(u+1)} = \boldsymbol{\kappa}^{(u)} - (\boldsymbol{K}_p^{\boldsymbol{\kappa}\boldsymbol{\kappa}})^{-1} \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}} \Big|_{(\boldsymbol{\beta}^{(u+1)}, \boldsymbol{\gamma}_1^{(u+1)}, \ldots, \boldsymbol{\gamma}_k^{(u+1)}, \boldsymbol{\kappa}^{(u)})}. \tag{3.31}$$

Then, the whole estimation process is presented below

(i) To choose suitable starting values, say $\boldsymbol{\beta}^{(0)}, \boldsymbol{\gamma}_1^{(0)}, \boldsymbol{\gamma}_2^{(0)}, \ldots, \boldsymbol{\gamma}_k^{(0)}$ and $\boldsymbol{\kappa}^{(0)}$;

(ii) To obtain $\boldsymbol{\beta}^{(u+1)}$ and $\boldsymbol{\gamma}_j^{(u+1)}$, $j = 1, \ldots, k$ using the backfitting iterations (Equations (3.30a) and (3.30b)), considering $\boldsymbol{\kappa}^{(u)}$;

(iii) Given current values, say $\boldsymbol{\beta}^{(u+1)}, \boldsymbol{\gamma}_1^{(u+1)}, \boldsymbol{\gamma}_2^{(u+1)}, \ldots, \boldsymbol{\gamma}_k^{(u+1)}$, to obtain $\boldsymbol{\kappa}^{(u+1)}$ by using Equation (3.31);

(iv) Iterating between (ii) and (iii) until reach some convergence criterion, that is until $||\boldsymbol{\theta}^{(u+1)} - \boldsymbol{\theta}^{(u)}|| > \epsilon, \epsilon > 0$.

ZOABR-SPAM

Similarly to the developments presented in Section 2.5 the EM algorithm will be employed. That is, we seek to maximize $Q(\boldsymbol{\varphi}|\boldsymbol{\varphi}^{(m)})$, regarding $\boldsymbol{\varphi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \alpha^\top)^\top$. Since the M-step has no analytical solution, some numerical method should be used. The respective details can be found in Section 2.5 for the BR-SPAM as well as the obtaining the necessary initial values. The summary of this estimation process is given by:

(i) To choose suitable starting values, say $\boldsymbol{\beta}^{(0)}, \boldsymbol{\gamma}_1^{(0)}, \boldsymbol{\gamma}_2^{(0)}, \ldots, \boldsymbol{\gamma}_k^{(0)}, \boldsymbol{\kappa}^{(0)}$ and $\alpha^{(0)}$;

(ii) To compute $\hat{\boldsymbol{v}}$ (E step in Equation (3.17));

(iii) Use some optimization algorithm to maximize $Q(\boldsymbol{\varphi}|\boldsymbol{\varphi}^{(u)})$ in relation to $\boldsymbol{\gamma}^{(u)}$ to find $\boldsymbol{\gamma}_j^{(u+1)}, j = 1, \ldots, k$;

(iv) For current values $\boldsymbol{\gamma}_j^{(u+1)}, j = 1, \ldots, k$, to obtain $\boldsymbol{\beta}^{(u+1)}$, $\boldsymbol{\kappa}^{(u+1)}$ and $\alpha^{(u+1)}$ by using some optimization algorithm to maximize $Q(\boldsymbol{\varphi}|\boldsymbol{\varphi}^{(u)})$ concerning these parameters;

(v) Iterating among (ii), (iii) and (iv) until reach some convergence criterion, that is until $||\boldsymbol{\theta}^{(u+1)} - \boldsymbol{\theta}^{(u)}|| > \epsilon, \epsilon > 0$.

3.6 Effective degrees of freedom

In additive linear models, the degrees of freedom correspond to approximately the number of effective parameters related to the non-parametric components. Then, following Hastie and Tibshirani (1990) and Eilers and Marx (1996) the effective degrees of freedom corresponding to the j-th non-parametric component is given by: $df_{\gamma_j} = \text{trace}\{\boldsymbol{B}_j^\top \widehat{\boldsymbol{W}}^\dagger \boldsymbol{B}_j (\boldsymbol{B}_j^\top \widehat{\boldsymbol{W}}^\dagger \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j)^{-1}\}$, for the ZOAB-SPAM and the ZOAS-SPAM, whereas, for the ZOABR-SPAM, it is given by: $df_{\gamma_j} = \text{trace}\{\boldsymbol{B}_j^\top \widehat{\boldsymbol{R}}^\dagger \boldsymbol{B}_j (\boldsymbol{B}_j^\top \widehat{\boldsymbol{R}}^\dagger \boldsymbol{B}_j + \lambda_j \boldsymbol{\Lambda}_j)^{-1}\}, \forall j = 1, \ldots, k$ (for more details about the development for the degrees of freedom of the ZOABR-SPAM, see Appendix A.5). In this case, the total effective degrees of freedom are approximately given by $df(\boldsymbol{\lambda}) \cong p + \sum_{j=1}^{k} df_{\gamma_j}(\boldsymbol{\lambda})$.

3.7 Model selection criteria

The Information Criteria will are those presented in Section 2.7, that is: AIC, BIC, AICc, HQIC and SABIC. Since, for the augmented models, the degrees of freedom are different from those for the non-agumented models, the quantity s^* must be be replaced by $s^{**} = s^* + s_0 + s_1$ and $l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ is given in Equation (3.10).

3.8 Estimation of smoother parameters

We can consider the generalized cross-validation method (see Wahba and Wold (1975) and Wood (2017), for example). Then, for the ZOAB-SPAM and ZOAS-SPAM, $\boldsymbol{\lambda}$ is chosen such

$$\hat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}} \text{GCV}(\boldsymbol{\lambda}) = \text{argmin}_{\boldsymbol{\lambda}} \sum_{i=1}^{n} \frac{\hat{w}_i^\dagger (\hat{z}_i^{**} - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2}{[1 - n^{-1}\text{trace}\{\widehat{\boldsymbol{H}}(\boldsymbol{\lambda})\}]^2},$$

where $w_i^\dagger$ is the i-th component of the diagonal of the respective Fisher information, z_i^{**} is the i-th component of the vector $\boldsymbol{z}^{**} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j + (\boldsymbol{W}^{\dagger(u)})^{-1} \boldsymbol{T}_1 \boldsymbol{f}^*$ and, for the ZOABR-SPAM,

$$\hat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}} \text{GCV}(\boldsymbol{\lambda}) = \text{argmin}_{\boldsymbol{\lambda}} \sum_{i=1}^{n} \frac{\hat{r}_i^\dagger (\hat{z}_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2}{[1 - n^{-1}\text{trace}\{\widehat{\boldsymbol{H}}(\boldsymbol{\lambda})\}]^2},$$

where $r_i^\dagger$ is the i-th component of the diagonal of the respective Hessian matrix, and z_i^{**} is the i-th component of the vector $\boldsymbol{z}^{**} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j\boldsymbol{\gamma}_j - (\boldsymbol{R}^\dagger)^{-1}\boldsymbol{T}_1\boldsymbol{f}^*$ and $\boldsymbol{H}(\boldsymbol{\lambda})$ is defined as $\boldsymbol{H}(\boldsymbol{\lambda}) = \boldsymbol{N}^* \left[\boldsymbol{N}^{*\top}\dot{\boldsymbol{W}}\boldsymbol{N}^* - \boldsymbol{\Lambda}(\boldsymbol{\lambda})\right]^{-1} \boldsymbol{N}^{*\top}\dot{\boldsymbol{W}}$, where $\boldsymbol{\Lambda}(\boldsymbol{\lambda}) = \oplus_{j=1}^{n}\lambda_j\boldsymbol{\Lambda}_j^d$, $\boldsymbol{N}^*\boldsymbol{\xi} = [\boldsymbol{X}, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_k]\boldsymbol{\xi}$, $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top)^\top$ and $\dot{\boldsymbol{W}}$ is the respective Fisher information for ZOAS-SPAM and ZOAB-SPAM, whereas, for ZOABR-SPAM is the respective Hessian matrix.

Another possibility is to select $\boldsymbol{\lambda}$ through Information Criteria, which was described in Section 3.7, for example doing $\widehat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}}\text{AIC}$.

3.9 Obtaining the standard errors

ZOAB-SPAM and ZOAS-SPAM

The respective variance-covariance matrices for the discrete part $(\text{Cov}(\hat{\boldsymbol{\vartheta}}))$ and continuous part $(\text{Cov}(\hat{\boldsymbol{\varphi}}))$ are obtained from $\boldsymbol{K}_p^{\boldsymbol{\vartheta}\boldsymbol{\vartheta}}(\boldsymbol{\theta})$ and $\boldsymbol{K}_p^{\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\theta})$, respectively. The are defined, respectively, by Equations (3.20) and (3.23).

ZOABR-SPAM

For the discrete part, the variance-covariance matrix are obtained from the Fisher information matrix $(\boldsymbol{K}_p^{\boldsymbol{\vartheta}\boldsymbol{\vartheta}}(\boldsymbol{\theta}))$ given in Equation (3.23).

For the parameters related to the continuous part, the EM algorithm is used to estimate the parameters. Then, we considered the empirical information matrix to obtain the standard errors, which is given (see the related discussion in Section

$$\boldsymbol{I}_e(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})\boldsymbol{s}(y_i|\boldsymbol{\theta})^\top,$$

where $\boldsymbol{s}(y_i|\boldsymbol{\varphi}),\ i = 1, \ldots, n$ is the score function for the i-th observation, which can be calculated as:

$$\boldsymbol{s}(y_i|\boldsymbol{\varphi}) = \mathbb{E}\left[\frac{\partial l_{2p}^c(\boldsymbol{\varphi}; y_i, z_i^*, v_i)}{\partial\boldsymbol{\varphi}}|y_i, z_i^*, \boldsymbol{\varphi}\right],$$

where $l_{2p}^c(\boldsymbol{\varphi}|y_i, z_i^*, v_i)$ is the penalized complete log-likelihood related with the continuous part of the model given in Equation (3.12) and

$$\boldsymbol{s}(y_i|\boldsymbol{\theta}) = (s_{\boldsymbol{\beta}}(y_i|\boldsymbol{\theta}), s_{\boldsymbol{\gamma}_1}(y_i|\boldsymbol{\theta}), \ldots, s_{\boldsymbol{\gamma}_k}(y_i|\boldsymbol{\theta}), s_{\boldsymbol{\kappa}}(y_i|\boldsymbol{\theta}), s_{\alpha}(y_i|\boldsymbol{\theta}))^\top,$$

where

$$s_{\beta}(y_i|\boldsymbol{\theta}) = (1-\hat{z}_i^*)\left\{\frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i}\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] + (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^*-\delta_i^*)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1}\boldsymbol{x}_i, \quad (3.32a)$$

$$s_{\gamma_j}(y_i|\boldsymbol{\theta}) = (1-\hat{z}_i^*)\left\{\frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i}\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] + (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^*-\delta_i^*)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1}\boldsymbol{b}_{ij} - \lambda_j\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_j, \quad (3.32b)$$

$$s_{\boldsymbol{\kappa}}(y_i|\boldsymbol{\theta}) = (1-\hat{z}_i^*)(1-\hat{v}_i)\{\delta_i(y_i^*-\delta_i^*)+\log(1-y_i) - \Psi[(1-\delta_i)\phi_i]+\Psi(\phi_i)\}\left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-1}\boldsymbol{e}_i, \quad (3.32c)$$

$$s_{\alpha}(y_i|\boldsymbol{\theta}) = (1-\hat{z}_i^*)\left\{\frac{2\mu_i(1-\mu_i)}{(1-\varepsilon_i)}\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]-\left[\frac{\phi_i\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3}\right](y_i^*-\delta_i^*)\right\}, \quad (3.32d)$$

where $j = 1, \ldots, k$ and $\hat{v}_i$ is defined ind Equation (3.17). Thus, the empirical information matrix can be calculated using Equations (3.32a) to (3.32d). Finally, Cov($\hat{\boldsymbol{\theta}}$) and SE($\hat{\boldsymbol{\theta}}$) are estimated by $(\boldsymbol{I}_e(\hat{\boldsymbol{\theta}}|\boldsymbol{y}))^{-1}$ and $\text{diag}\{(\boldsymbol{I}_e(\hat{\boldsymbol{\theta}}|\boldsymbol{y}))^{-1/2}\}$, respectively.

3.10 Hypothesis testing

As in the Section 2.10, for testing the significance linear statistical hypothesis $\boldsymbol{C}\varsigma = \boldsymbol{d}$, where rank($\boldsymbol{C}$) $= l$, $l \geq p$ or $l \geq q$, it can be used the Wald-type statistic, that is,

$$\xi_W = (\boldsymbol{C}\hat{\boldsymbol{\varsigma}} - \boldsymbol{d})^\top(\boldsymbol{C}\boldsymbol{V}_\varsigma\boldsymbol{C}^\top)^{-1}(\boldsymbol{C}\hat{\boldsymbol{\varsigma}} - \boldsymbol{d}),$$

Under H_0, $\xi_W \xrightarrow[n\to\infty]{D} \chi_l^2$, where $\xrightarrow[n\to\infty]{D}$ means convergence in distribution when the sample size tends to infinity and, besides that, $\boldsymbol{\varsigma}$ can be either $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$, $\boldsymbol{V}_\varsigma$ is the matrix of variance-covariance regards to $\boldsymbol{\varsigma}$ extract from $\widehat{\text{Cov}}(\hat{\boldsymbol{\vartheta}})$ or $\widehat{\text{Cov}}(\hat{\boldsymbol{\varphi}})$.

3.11 Model fit diagnostic tools

As done in Section 2.11, several model fit diagnostic tools are developed/adapted for the proposed model class. Specifically, we consider: residual analysis, based on the randomized quantile residual and local influence analysis, considering the likelihood-displacement and Q-displacement.

3.11.0.1 Residual analysis

Residual analysis is important the check the validity of the underlying assumptions of models. The more complex the model the more sensitivity to the depart of such assumptions. Since the the response distribution presents a mixture structure, between a discrete and a continuous distributions, we considered the randomized quantile residual (RQR), introduced by Dunn and Smyth (1996). Following Silva et al. (2020), the RQR for our model class is given by $r_i = \Phi(G_i), i = 1, \ldots, n$, where $\Phi(.)$ denotes the cumulative distribution function of the standard Normal distribution, $G_i \sim U((a_i, b_i])$, where $a_i = \lim_{y \uparrow y_i} F(y; \hat{\upsilon}, \hat{\varrho}, \hat{\mu}, \hat{\phi})$, $b_i = F(y_i; \hat{\upsilon}, \hat{\varrho}, \hat{\mu}, \hat{\phi})$, where $\hat{\varrho} = \hat{p}_{1i}/\hat{\upsilon}$ and $\hat{\upsilon} = \hat{p}_{0i} + \hat{p}_{1i}$ and

$$F(y; \upsilon, \varrho, \mu, \phi) = \upsilon \text{ Bernoulli}(y; \varrho) + (1 - \upsilon) \int_0^y f(z; \boldsymbol{\varpi}) dz,$$

where Bernoulli$(y; \varrho)$ is the cdf of a Bernoulli with parameter ϱ and $f(z; \boldsymbol{\varpi})$ is the density of the beta, simplex and BR distributions for the models ZOAB/ZOAS/ZOABR SPAM, respectively. Then, we have that Silva et al. (2020), $G_i \sim U(0, \hat{\upsilon}(1 - \hat{\varrho})]$ if $y_i = 0$, $G_i \sim U(1 - \hat{\upsilon}\hat{\varrho}, 1]$ if $y_i = 1$ and $G_i = F(Y_i; \hat{\upsilon}, \hat{\varrho}, \hat{\mu}, \hat{\phi})$ if $y_i \in (0, 1)$. The authors mention that since the response variable is no longer continuous, it is necessary simulate several sets of values for r_i. That is, the quantile residuals will vary from one realization to another, for a given result from a model fitted to a given data set. Thus, in practice, it is important to simulate four sets of values for the quantile residuals and any pattern that is not consistent across these sets is then ignored, see Dunn and Smyth (1996) for more details.

Provided that the model is well fitted to the data and that the estimators are consistent, the residuals converge to standard normal distribution. Then, plots of the residuals against the index and/or the fitted values as well as quantile-quantile plot using the theoretical N(0,1) quantiles are useful for checking the behavior of the residuals.

3.11.1 Local Influence

Again, we follow the work of Silva et al. (2020). Thus, for the ZOAB-SPAM and ZOAS-SPAM, even though the developments (see Cook (1986)) are the same for the parameters related to both discrete and continuous parts Cook (1986), they are made separated from each other. For the discrete part of the ZOABR-SPAM, the method is that used for the two other models. However, for the continuous part, we considered the approach proposed by Zhu and Lee (2001) since the estimation process were done by using the EM algorithm, separated from the discrete part.

3.11.1.1 Local influence for discrete part for all models

The likelihood displacement for discrete part of the models is defined by $LD(\boldsymbol{\omega}) = 2[l_1(\hat{\boldsymbol{\vartheta}}) - l_1(\hat{\boldsymbol{\vartheta}}_{\omega})] \geq 0$ can be used to asses the influence of perturbations on

MPLE estimates $\hat{\boldsymbol{\vartheta}}$, where $\hat{\boldsymbol{\vartheta}}_\omega$ is the MPLE estimates of $\boldsymbol{\vartheta}$ for a perturbed model, whose perturbed penalized log-likelihood function is denoted by $l_1(\boldsymbol{\theta}|\boldsymbol{\omega})$, and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is an n-dimensional vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$. For comparison, it is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $l_1(\boldsymbol{\vartheta}|\boldsymbol{\omega}_0) = l_1(\boldsymbol{\vartheta})$

We have that $\boldsymbol{l}_{max}$ denotes the corresponding eigenvector of the highest eigenvalue of the matrix $\boldsymbol{Z} = \boldsymbol{\Delta}^\top[-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})]^{-1}\boldsymbol{\Delta}$, where $-\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})$ is the observed Fisher information matrix, $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta})$ is defined in Equation (3.19) and

$$\boldsymbol{\Delta} = \left.\frac{\partial^2 l_1(\boldsymbol{\vartheta}|\boldsymbol{\omega})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

As discussed in Section 2.11.4, the index plot of $|\boldsymbol{l}_{\max}|$ can reveal observations with high influence in the neighborhood of the likelihood displacement, $LD(\boldsymbol{\omega}_0)$. Those observations may be responsible for substantial changes in the estimates of parameters, under little perturbations in the model and in the data.

Case-weight perturbation

Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ be a perturbed vector. In this case, perturbed log-likelihood function related only with discrete part is given by

$$l_1(\boldsymbol{\vartheta}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i l_i(p_{0i}, p_{1i}),$$

with $0 \leq \omega_i \leq 1$, $i = 1, \ldots, n$. Hence, considering $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$, the elements of the penalized perturbation matrix are expressed as

$$\left.\frac{\partial^2 l_1(\boldsymbol{\vartheta}|\boldsymbol{\omega})}{\partial\boldsymbol{\rho}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{F}^\top \text{diag}\{\boldsymbol{z}^*(\mathbb{1}_n - \boldsymbol{y}) - \hat{\boldsymbol{p}}_0\},$$
$$\left.\frac{\partial^2 l_1(\boldsymbol{\vartheta}|\boldsymbol{\omega})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{M}^\top \text{diag}\{\boldsymbol{z}^*\boldsymbol{y} - \hat{\boldsymbol{p}}_1\},$$

where $\mathbb{1}_n$ is a n-vector of 1's.

3.11.1.2 Local influence for continuous part

For the ZOAB-SPAM and ZOAS-SPAM we have that $LD(\boldsymbol{\omega}) = 2[l_{2p}(\hat{\boldsymbol{\varphi}}, \boldsymbol{\lambda}) - l_{2p}(\hat{\boldsymbol{\varphi}}_\omega, \boldsymbol{\lambda})] \geq 0$ can be used to asses the influence of perturbations on MPLE estimates $\hat{\boldsymbol{\varphi}}$, where $\hat{\boldsymbol{\varphi}}_\omega$ is the MPLE estimates of $\boldsymbol{\varphi}$ for a perturbed model. The perturbed penalized log-likelihood function is denoted by $l_{2p}(\boldsymbol{\varphi}, \boldsymbol{\lambda}|\boldsymbol{\omega})$ and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is an n-dimensional vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$.

For the ZOABR-SPAM we used the complete log-likelihood for parameter estimation. Then, we can use $l_{2p}^c(\boldsymbol{\varphi}, \boldsymbol{\omega}|\boldsymbol{y}_c)$ as the complete log-likelihood of the perturbed model as well as the Q-displacement function defined by $QD(\boldsymbol{\omega}) = 2[Q(\hat{\boldsymbol{\varphi}}|\hat{\boldsymbol{\varphi}}) - Q(\hat{\boldsymbol{\varphi}}_\omega|\hat{\boldsymbol{\varphi}})]$. Then, the influence of perturbations on MPLE estimates $\boldsymbol{\varphi}$, where $\hat{\boldsymbol{\varphi}}_\omega$ is the maximum of $Q(\boldsymbol{\varphi}, \boldsymbol{\omega}|\hat{\boldsymbol{\varphi}}) = \mathbb{E}[l^c_{2p}(\boldsymbol{\varphi}|\boldsymbol{y}, \hat{\boldsymbol{\varphi}})]$, can be properly assessed.

The $\boldsymbol{l}_{max}$ denotes the corresponding eigenvector of the highest eigenvalue of $\boldsymbol{Z}$ matrix, which is defined as $\boldsymbol{Z} = \boldsymbol{\Delta}^\top[-\ddot{\boldsymbol{U}}^{\boldsymbol{\theta}\boldsymbol{\theta}}_p(\boldsymbol{\theta})]^{-1}\boldsymbol{\Delta}$, where $-\ddot{\boldsymbol{U}}^{\boldsymbol{\theta}\boldsymbol{\theta}}_p(\boldsymbol{\theta})$ is the observed Fisher information matrix, and $\ddot{\boldsymbol{U}}^{\boldsymbol{\theta}\boldsymbol{\theta}}_p(\boldsymbol{\theta})$ is defined in the Equation (3.24) for ZOAB-SPAM, in (3.26) for ZOAS-SPAM and in (3.22) for ZOABR-SPAM and for curvature based on $LD(\boldsymbol{\varphi})$ and for the one based on $QD(\boldsymbol{\varphi})$, the $\boldsymbol{\Delta}$, are respectively define as:

$$\boldsymbol{\Delta} = \left.\frac{\partial^2 l_p(\boldsymbol{\varphi}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\varphi}=\hat{\boldsymbol{\varphi}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \text{ and } \boldsymbol{\Delta} = \left.\frac{\partial Q(\boldsymbol{\varphi}|\hat{\boldsymbol{\varphi}})}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\varphi}=\hat{\boldsymbol{\varphi}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

As discussed in Section 2.11.4, the index plot of $|\boldsymbol{l}_{\max}|$ can reveal those observations with greater influence in the neighborhood of the likelihood displacement, $LD(\boldsymbol{\omega}_0)$, or the Q-displacement, $QD(\boldsymbol{\omega}_0)$. Those observations may be responsible for substantial changes in the estimates of parameters under little perturbations in the model and in the data.

Case-Weight perturbation

Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ be a perturbation vector. In this case, the perturbed penalized log-likelihood is given by $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = l(\boldsymbol{\theta}|\boldsymbol{\omega}) - \sum_{j=1}^{k} \frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}^d_j \boldsymbol{\gamma}_j$, where $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i l_i(\boldsymbol{\theta})$, and $0 \leq \omega_i \leq 1$. Hence, considering $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$, the elements of $\boldsymbol{\Delta}$, for both the ZOAS-SPAM and ZOAB-SPAM, are given by

$$\begin{aligned}
\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \boldsymbol{X}^\top\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}, \\
\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \boldsymbol{B}_j^\top\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}, \quad j = 1, \ldots, k, \\
\left.\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{\omega})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \boldsymbol{E}^\top\hat{\boldsymbol{T}}_2\text{diag}\{\hat{\boldsymbol{a}}\},
\end{aligned}$$

where $\boldsymbol{a}$, $\boldsymbol{T}_1$, $\boldsymbol{T}_2$ and $\boldsymbol{f}^*$ are defined in Equation (3.14) for the ZOAB-SPAM and in Equation (3.15) for the ZOAS-SPAM.

For the ZOABR-SPAM, the perturbed Q-function is given by $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, where $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. Hence, the elements of the perturbation matrix are expressed as

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}^\top\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\},$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\gamma_j\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{B}_j^\top\hat{\boldsymbol{T}}_1\text{diag}\{\hat{\boldsymbol{f}}^*\}, \quad j=1,\dots,k,$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{E}^\top\hat{\boldsymbol{T}}_2\text{diag}\{\hat{\boldsymbol{p}}\},$$

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\alpha\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \hat{\boldsymbol{d}}^*,$$

where $\boldsymbol{T}$, $\boldsymbol{p}$ and $\boldsymbol{f}^*$ are defined in Equation (3.16). Furthermore, $\boldsymbol{d}^{**} = (\boldsymbol{1}_n - \boldsymbol{z}^*)\boldsymbol{d}^*$, where $\boldsymbol{d}^*$ is defined in Equation (2.32).

3.12 Computational tools

All methods proposed in this work were implemented in the `R` program (R Core Team, 2020). The parameter estimation for the ZOAB-SPAM and ZOAS-SPAM were implemented within the R package `gamlss` (Stasinopoulos et al., 2007). The framework of the gamlss accommodates the developed methodology for either models and estimation process of them for these two models.

For the ZOABR-SPAM, the maximization process for continuous part was done using the function `optim` of the software `R`, which uses the L-BFGS-B (Byrd et al., 1995) algorithm, for the maximization of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ and the function `nlminb` (Gay, 1990) that uses, which they call as optimization using PORT routines.

3.13 Simulation Studies

In this section, the results of four simulation studies are presented. Section 3.13.1 is related to the analysis of parameter recovery. In Section 3.13.2, we perform a link misspecification study. In Section 3.13.3, we evaluate the response distribution misspecification. Finally, in Section 3.13.4 is presented a study related to the performance of the proposed quantile residuals and the local influence measures. The results for the first study are presented in Appendix, whereas the others are available at `https://github.com/aureaflg/Simulations-study.git`, where we made reproducible codes that also show plot and tables of interest.

3.13.1 Study 1: Parameter recovery

Here we considered several relevant scenarios defined by the combination of the levels of some factors of interest. The factors (with the respective levels within parenthesis) are sample size (n) (50, 100, 500), regression model (ZOAB, ZOAS, ZOABR), link function (probit, logit, cauchit, cloglog, loglog) and modeled parameters (mean and dispersion). For each scenario, $R = 300$ replicates of Monte Carlo were generated. Each scenario are described in Table 7.

Table 7 – Scenarios of simulation study 1

Regression model (modelling mean and dispersion)	n	Link functions
ZOAB$(\mu_i, \phi_i, p_{0i}, p_{1i})$	50 100 500	probit, logit, cauchit, cloglog and loglog.
ZOAS$(\mu_i, \phi_i, p_{0i}, p_{1i})$	50 100 500	
ZOABR$(\mu_i, \phi_i, p_{0i}, p_{1i}, \alpha)$	50 100 500	

Furthermore, the considered linear predictors and link functions for related scenarios presented in Table 7, are given by $\log(p_{0i}/(1-p_{0i}-p_{1i})) = \rho_0+\rho_1 F_i$, $\log(p_{1i}/(1-p_{0i}-p_{1i})) = \tau_0+\tau_1 M_i$, $g(\mu_i) = \beta_1 X_i + cos(Z_i)$, $\log(\phi_i) = \kappa_0+\kappa_1 E_i$ for $i = 1,\ldots,n$, where $X_i \overset{iid}{\sim}$ uniform$(0, 1)$, $E_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $Z_i \overset{iid}{\sim}$ uniform$(0, 3\pi)$, $F_i \overset{iid}{\sim}$ uniform$(0, 1)$, $M_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $g(.)$ is either the probit, logit, cauchit, cloglog.

The actual parameter values considered are: loglog link ($\beta_1 = -1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$) and other links ($\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$). This change to the loglog link function is justified by the fact that, depending on the parameters chosen, generating from the models using this link function can generate values very close to one or zero, which could compromise the fit of the models to the simulated data even for the BR-SPAM model.

For all models, when $n = 50$, 40 knots were used and the smoothing parameter was fixed at $\lambda = 10$, whereas, when $n = 100$ or $n = 500$, 50 knots were used and the smoothing parameter was fixed at $\lambda = 50$. These values lead to a smooth fitted curve for non-parametric component.

The related results are presented in Appendix B.1: from Figures 49 to 53 for the ZOAB-SPAM, from Figures 54 to 58 for the ZOAS-SPAM, and, from Figures 59 to 63 for the ZOABR-SPAM.

In a general way, for the three models, all the parameters were properly recovered and as the sample size increases, the estimates become more accurate. The estimated curves are for all sample size representing the behavior of the curves, in general well. For

$n = 100$ and $n = 500$ the non-parametric curves are more accurately estimated, i.e, there is less variability around the real curve than for n = 50.

3.13.2 Study 2: Link function misspecification

In this section, we considered only the scenarios varying the regression models (ZOAB, ZOAS, ZOABR) and the link functions (probit, logit, cauchit, cloglog,loglog). We set the sample size in 500, modeling the mean and the dispersion. We generate only one replica from each model, setting a distribution and a link function for it, and fit the simulated data using a model with the same distribution but varying the link function, using all the different ones from the one used to generate the simulated data. In Table 8 we present the 60 scenarios of interest.

Table 8 – Scenarios of simulation study 2

Distribution that will generate and fit the simulated data	Actual link function	Link function considered in fitted model
ZOAB($\mu_i, \phi_i, p_{0i}, p_{1i}$), ZOAS($\mu_i, \phi_i, p_{0i}, p_{1i}$) or ZOABR($\mu_i, \phi_i, p_{0i}, p_{1i}, \alpha$)	logit	probit, cauchit, cloglog and loglog
	probit	logit, cauchit, cloglog and loglog
	cauchit	logit, probit, cloglog and loglog
	cloglog	logit, probit, cauchit and loglog
	loglog	logit, probit, cloglog and cauchit

The considered linear predictors and link functions for related scenarios, are given by $\log(p_{0i}/(1-p_{0i}-p_{1i})) = \rho_0+\rho_1 F_i$, $\log(p_{1i}/(1-p_{0i}-p_{1i})) = \tau_0+\tau_1 M_i$, $g(\mu_i) = \beta_1 X_i + cos(Z_i)$, $\log(\phi_i) = \kappa_0 + \kappa_1 E_i$ for $i = 1, \ldots, n$, where $X_i \overset{iid}{\sim}$ uniform$(0, 1)$, $E_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $Z_i \overset{iid}{\sim}$ uniform$(0, 3\pi)$, $F_i \overset{iid}{\sim}$ uniform$(0, 1)$, $M_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $g(.)$ is either the probit, logit, cauchit, cloglog.

The actual parameter values considered loglog link were $\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$, whereas for the other cases were $\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$. We considered both the comparison of the observed and predicted values as well as the parameter recovery.

The ZOAB-SPAM and ZOABR-SPAM presented bad recovery of real mean when the actual link function was symmetric and the model that fitted the simulated data was using loglog and cloglog as link function, (both are asymmetric link) and the same behaviour were seen for the recovery of regression parameters related with mean. While, for parameters related with dispersion/precision, they were all well estimated, except in the case that the generating model was using loglog as link function because only model using loglog fitted well the parameters. For the ZOAS-SPAM, the results for the mean were better than ZOAB-SPAM and ZOABR-SPAM, because for all link function that generating and fitting, it was properly recovered. Furthermore, for ZOAS-SPAM, the parameter related with mean and for parameters related with dispersion/precision, they were all well estimated.

3.13.3 Study 3: Response distribution misspecification

In this section we vary only the response distribution. We set the sample size in 100 and link function as the logit one. We model only the mean and considered only one (R = 1) replica. We simulate from a model setting a the response distribution and, then, fit using all the others different distribution from the one which simulated the data. In Table 9 we present the 6 scenarios of interest.

Table 9 – Scenarios of simulation study 3

Actual model ($n = 100$)	Fitted model
ZOAB$(\mu_i, \phi, p_{0i}, p_{1i})$	ZOAS ZOABR
ZOAS$(\mu_i, \phi, p_{0i}, p_{1i})$	ZOAB ZOABR
ZOABR$(\mu_i, \phi, p_{0i}, p_{1i}, \alpha)$	ZOAB ZOAS

Furthermore, the related linear predictors and link functions to scenarios presented in Table 9 are given by $\log(p_{0i}/(1 - p_{0i} - p_{1i})) = \rho_0 + \rho_1 F_i$, $\log(p_{1i}/(1 - p_{0i} - p_{1i})) = \tau_0 + \tau_1 M_i$, $g(\mu_i) = \beta_1 X_i + cos(Z_i)$ for $i = 1, \dots, n$, where $X_i \overset{iid}{\sim}$ uniform$(0, 1)$, $E_i \overset{iid}{\sim}$ uniform$(0, 1)$ and $Z_i \overset{iid}{\sim}$ uniform$(0, 3\pi)$, $F_i \overset{iid}{\sim}$ uniform$(0, 1)$, $M_i \overset{iid}{\sim}$ uniform$(0, 1)$, besides that, for all scenarios, $g(.)$ is logit. The actual values considered for the parameters were $\beta_1 = -1$, $\phi = 20$, $\alpha = 0.1$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$.

The parameters, in a general way, were well recovered for all models, except the non-parametric curve when the true underlying model and the fitted model do not match. Note that for this study the value of parameter α of ZOABR was considering very small to the generating data could be adjust for the others models, because when the α increasing, the adjust of others models become poor and for ZOAS, it is impossible the adjust in same cases. In summary, is that when data presents values too close to the limits of the interval $(0, 1)$, the ZOABR-SPAM is the best model, in contrast to when the data is distributed further away from the limits, the models present a similar behaviour.

3.13.4 Study 4: Performance of residuals tools and local influence

It was generated two data sets, one from the ZOAS-SPAM and another from the ZOAB-SPAM, considering the cauchit link and a sample size of 100. Then, we fit all other distributions, with the same link. The purpose is to evaluate if the residuals plots properly indicate whether the model is well fit to the data. It was considered only one replica for this study (R=1).

In the local influence analysis for the continuous part, under the case-weight scheme, only one observation was "contaminated" from each one of the simulated data set. These were generated from the ZOAS/ZOAB/ZOABR SPAM and fitting using the respective actual model. Under the same scheme, another perturbation was added to evaluate the developed measure, it was selected an observation from a covariate of the parametric part of the linear predictor related with the mean, and was added to it 4 times of the standard deviation of this covariate.

Furthermore, in the local influence analysis for the discrete part, an observation selected from a covariate related with the probability of occurrence of zero was "contaminated" by adding 20 times of the standard deviation of this covariate.

The related models of the scenarios concerning the model fit assessment tools, have this following structure $Y_i \overset{ind}{\sim}$ ZOAB$(\mu_i, \phi_i, p_{0i}, p_{1i})$, or ZOAS$(\mu_i, \phi_i, p_{0i}, p_{1i})$, $\log(p_{0i}/(1-p_{0i}-p_{1i})) = \rho_0+\rho_1 f_i$, $\log(p_{1i}/(1-p_{0i}-p_{1i})) = \tau_0+\tau_1 m_i$, $g(\mu_i) = \beta_1 x_i + cos(z_i)$, $\log(\phi_i) = \kappa_0 + \kappa_1 e_i$, for $i = 1, \ldots, n$, where $X \overset{iid}{\sim}$ uniform$(0,1)$, $E \overset{iid}{\sim}$ uniform$(2,3)$ and $Z \overset{iid}{\sim}$ uniform$(0,3\pi)$, $F \overset{iid}{\sim}$ uniform$(0,1)$, $M \overset{iid}{\sim}$ uniform$(0,1)$, $g(.)$ is cauchit. The actual values considered for the parameters were $\beta_1 = -2$ $\kappa_0 = -1$ and $\kappa_1 = 2$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$.

The related models of the scenarios concerns the local influence analysis, is given by $Y_i \overset{ind}{\sim}$ ZOAB$(\mu_i, \phi_i, p_{0i}, p_{1i})$,ZOAS$(\mu_i, \phi_i, p_{0i}, p_{1i})$ or ZOAB$(\mu_i, \phi_i, p_{0i}, p_{1i}, \alpha)$, $\log(p_{0i}/(1-p_{0i}-p_{1i})) = \rho_0+\rho_1 f_i$, $\log(p_{1i}/(1-p_{0i}-p_{1i})) = \tau_0+\tau_1 m_i$, $g(\mu_i) = \beta_1 x_i + cos(z_i)$, $\log(\phi_i) = \kappa_0 + \kappa_1 e_i$, for $i = 1, \ldots, n$. For continuous part, it was considered $X_i \overset{iid}{\sim}$ uniform$(0,1)$, $E_i \overset{iid}{\sim}$ uniform$(2,3)$ and $Z_i \overset{iid}{\sim}$ uniform$(0,3\pi)$, $F_i \overset{iid}{\sim}$ uniform$(0,1)$, $M_i \overset{iid}{\sim}$ uniform$(0,1)$, $g(.)$ is cauchit and the real values of parameters as $\beta_1 = -2$ $\kappa_0 = 1$ and $\kappa_1 = -6$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$. And, for the discrete part, the changes were to consider $X_i \overset{iid}{\sim}$ uniform$(0,1)$, $E_i \overset{iid}{\sim}$ uniform$(2,3)$ and $Z_i \overset{iid}{\sim}$ uniform$(0,3\pi)$, $F_i \overset{iid}{\sim}$ uniform$(0,1)$, $M_i \overset{iid}{\sim}$ uniform$(0,0.1)$, $g(.)$ is cauchit and to define the real values for the parameters as $\beta_1 = -2$ $\kappa_0 = 1$ and $\kappa_1 = -6$, $\rho_0 = -1.8$, $\rho_1 = 1.5$, $\tau_0 = -3$ and $\tau_1 = 1$.

The residual plots behave as expected when the actual and fitted models match. In these cases the randomized quantile residuals are well within the confidence bands, randomly distributed with no systematic behavior. A similar pattern is observed in the respective index plot. When the models do not match we can observe some residuals out of confidence bands (QQ plots) as well a systematic behavior. Therefore, we have evidence the developed tools are useful for goodness of model fit assessment

For the influence analysis, in case-weight perturbation for both continuous and discrete parts, the three contaminated observations where flagged through the $\boldsymbol{l}_{max}$

measure. Therefore, we have indications that the influence analysis tools are useful for such purpose.

3.14 Real data analysis

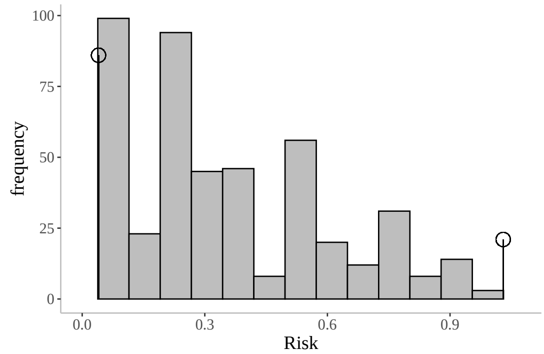
The data set here analyzed is part of a psychometric study of risk perception developed in Carlstrom et al. (2000), where the participants completed two forms of a simplified CER (Conjoint Expected Risk) questionnaire: an objective part and a subjective part. Each part described 16 health and 6 financial activities. The objective questionnaire contained hypothetical values for the five risk model variables for each 22 activities. The participants were instructed to suppose that the values were true and to provide a risk judgment for each of the activities based on these values.

For the subjective questionnaire the participants were instructed to provide their own subjective estimates of the model variables in addition to an overall judgment of risk for each activity. All participants also answered to a worldview questionnaire. The participants were recruited between 1997-1998 from five sources: UCLA psychology undergraduate classes, campus and community organizations, community and college newspaper advertisements, a paid consultant, and posted flyers.

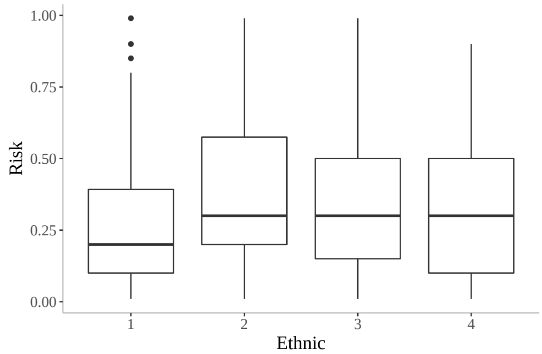
We selected the subjective part, where each subject was asked to provide a number in the [0,100] interval, such that the higher the value, the higher the perceived risk. Hence, the response is the perceived risk (Risk) of the 588 subjects related to a screening for genes that may predispose subjects to heart diseases (which is one of the 22 activities of the questionnaire) transformed to the interval [0,1]. Furthermore, as the covariates we consider: the gender (Gender), ethnic background (Ethnic), age (Age) and the worldview obtained from a related questionnaire (Wvcat). The levels of each factor are presented in Table 10. This data set was previously analyzed by Silva et al. (2020) who considered a Zero-and-One Augmented Beta Rectangular Regression Model with only parametric predictors.

Table 10 – Factor levels

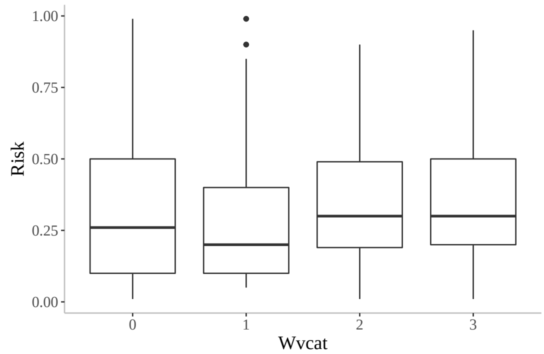
Variables	Levels
Ethnic	Caucasian (1); African-American (2) Mexican-American (3); Taiwanese-American (4)
Wvcat	unclassifiable (0); individualist (1) hierarchicalist (2); egalitarian (3)
Gender	female (0); male (1)



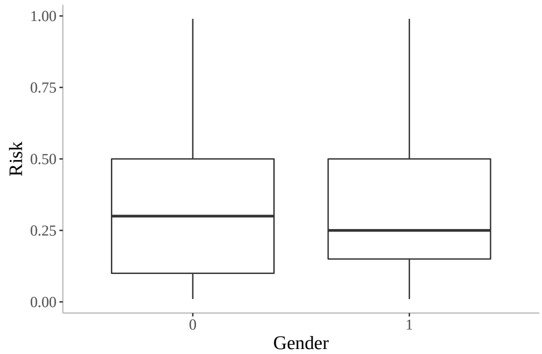
(a) Histogram of Risk by the frequency.



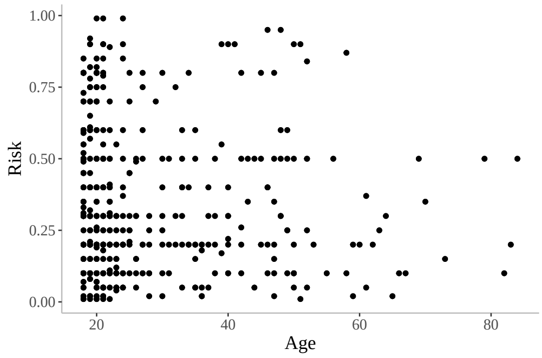
(b) Boxplot of Risk by categories of Ethnic covariate.



(c) Boxplot of Risk by categories of Wvcat covariate.



(d) Boxplot of Risk by categories of Gender covariate.



(e) Scatter plot of Risk versus Age.

Figure 17 – Explanatory analysis plots of risk perception data set.

The Figure 17 shows explonatory analysis plots of the data set for the continuous part, i.e. $0 < \text{Risk} < 1$, except the first plot, which presents the histogram of the Risk (Figure 17a). We can see that the Risk distribution seems to change over the levels of Ethnic and Wcat, whereas only the median seems to be different between the levels of Gender. That is, both Ethnic and Wcat may be affecting the location and variability of the response, whereas Gender seems to be related only to the Risk mean. Finally, we can see a nonlinear relation between Risk and Age, where a smooth function could be suitable for modeling that.

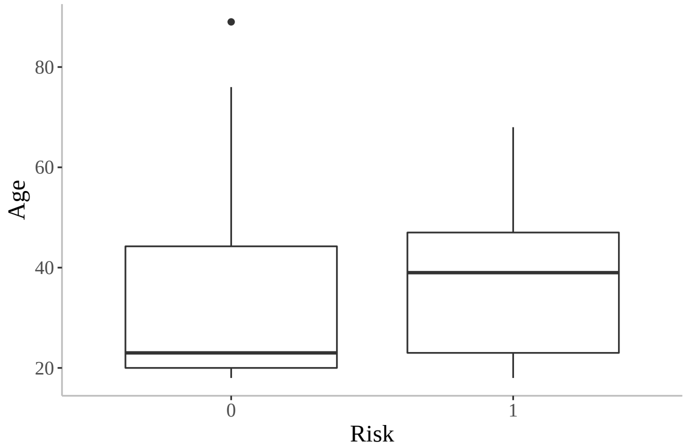


Figure 18 – Boxplots for Age by Risk=0 and Risk=1.

Table 11 – Number of discrete observations by the levels of Ethnic variable

	Risk		
Ethnic	0	1	Total
1	17	1	18
2	28	15	43
3	18	5	23
4	23	0	23
Total	86	21	107

Table 12 – Number of discrete observations by the levels of Gender variable

	Risk		
Gender	0	1	Total
0	46	15	61
1	40	6	46
Total	86	21	107

Table 13 – Number of observation by the levels of Wvcat variable

	Risk		
Wvcat	0	1	Total
0	65	16	81
1	4	2	6
2	9	2	11
3	8	1	9
Total	86	21	107

Considering, now the plot of the Figure 18 and the Tables 11, 12 and 13 related with the discrete part, i.e. Risk = 0 and Risk = 1, it is possible to notice that the covariate Wvcat is the one that seems to be the most influential for the response. Figure 18 related to Age, presents different median and dispersion for Risk=0 and Risk=1, while for the Risk equals to zero it is seen a positive skewness, for equals 1 it is seen a negative skewness, which may indicates that this variable may influence the outcome.

Initially, we fitted three models, that is, the ZOAB-SPAM, ZOAS-SPAM and ZOABR-SPAM with the following regression structures. For simplicity, we decided not to consider the possible interactions between the variables.

$$
\begin{aligned}
Y_{iljk} &\overset{ind}{\sim} \text{ZOAB}(\mu_{iljk}, \phi_{iljk}, p_{0iljk}, p_{1iljk}) \\
&\quad \text{or } \ \text{ZOABR}(\mu_{iljk}, \phi_{iljk}, p_{0iljk}, p_{1iljk}, \alpha) \\
&\quad \text{or ZOAS}(\mu_{iljk}, \phi_{iljk}, p_{0iljk}, p_{1iljk}), \\
g_1(\mu_{iljk}) &= \beta_0 + (\beta_1)_l + (\beta_2)_j + (\beta_3)_k + h_1(z_i), \\
g_2(\phi_{iljk}) &= \kappa_0 + (\kappa_1)_l + (\kappa_2)_j + (\kappa_3)_k + \kappa_4 z_i, \\
\log(p_{0iljk}/(1 - p_{0iljk} - p_{1iljk})) &= \rho_0 + (\rho_1)_l + (\rho_2)_j + (\rho_3)_k + \rho_4 z_i \\
\log(p_{1iljk}/(1 - p_{0iljk} - p_{1iljk})) &= \tau_0 + (\tau_1)_l + (\tau_2)_j + (\tau_3)_k + \tau_4 z_i,
\end{aligned}
$$

where, $i = 1, \ldots, 588$, $l = 1, 2, 3, 4$, $j = 0, 1, 2, 3$, $k = 0, 1$, $(\beta_1)_1 = (\beta_2)_0 = (\beta_3)_0 = 0$, $(\kappa_1)_1 = (\kappa_2)_0 = (\kappa_3)_0 = 0$, $(\rho_1)_1 = (\rho_2)_0 = (\rho_3)_0 = 0$, $(\tau_1)_1 = (\tau_2)_0 = (\tau_3)_0 = 0$. Also, (β_1), (κ_1), (ρ_1) and (τ_1) are related with Ethnic; (β_2), (κ_2), (ρ_2) and (τ_2) are related with Wvcat; (β_3), (κ_3), (ρ_3) and (τ_3) are related with Gender. Finally, κ_4, ρ_4 and τ_4 are associated with the variable $\boldsymbol{z}$ standardize Age.

Also, $h_1(.)$ is the unknown function that will be approximated by P-splines with cubic B-splines and quadratic differences with 50 knots for all models. The link function $g_1(.)$ will vary in cauchit, logit, probit, cloglog and loglog, and the link $g_2(.)$ was chosen the logarithm function for the ZOAB-SPAM and the ZOABR-SPAM, while for the ZOAS-SPAM, the chosen function was the square root. The difference for the link $g_2(.)$ among the models is justifying by the improvement of the fit for each model.

Therefore, we fitted all models with all possible link function for the mean and, then, only the covariates statistically significant at the level of 5% were kept in the model. Also, the equivalent levels of the categorical covariates were combined. The reduced models for each distribution are presented below. Each link function present the same significance for the covariates, then the reduced model is the same for each one of them.

ZOAB-SPAM:

$$
\begin{aligned}
Y_{il} &\overset{ind}{\sim} \text{ZOAB}(\mu_{il}, \phi_{il}, p_{0i}, p_{1i}), \\
g_1(\mu_i) &= h_1(z_i), \\
g_2(\phi_{il}) &= \kappa_0 + (\kappa_1)_l, \\
\log(p_{0i}/(1 - p_{0i} - p_{1i})) &= \rho_0 + \rho_4 z_i \\
\log(p_{1i}/(1 - p_{0i} - p_{1i})) &= \tau_0 + \tau_4 z_i,
\end{aligned}
$$

where $l = 1, 2, 3, 4$, $(\kappa_1)_1 = (\kappa_1)_3 = (\kappa_1)_4 = 0$.

ZOAS-SPAM:

$$\begin{aligned}
Y_{ilj} &\overset{ind}{\sim} \text{ZOAS}(\mu_{ilj}, \phi_i, p_{0i}, p_{1i}),\\
g_1(\mu_{ilj}) &= \beta_0 + (\beta_1)_l + (\beta_2)_j + h_1(z_i),\\
g_2(\phi_i) &= \kappa_0 + \kappa_4 z_i,\\
\log(p_{0i}/(1 - p_{0i} - p_{1i})) &= \rho_0 + \rho_4 z_i\\
\log(p_{1i}/(1 - p_{0i} - p_{1i})) &= \tau_0 + \tau_4 z_i,
\end{aligned}$$

where $l = 1, 2, 3, 4,\ j = 0, 1, 2, 3,\ (\beta_1)_1 = (\beta_1)_3 = 0,\ (\beta_2)_0 = (\beta_2)_2 = (\beta_2)_3 = 0.$

ZOABR-SPAM:

$$\begin{aligned}
Y_{il} &\overset{ind}{\sim} \text{ZOABR}(\mu_{il}, \phi_{il}, p_{0i}, p_{1i}, \alpha),\\
g_1(\mu_{il}) &= \beta_0 + (\beta_1)_l + h_1(z_i),\\
g_2(\phi_{il}) &= \kappa_0 + (\kappa_1)_l,\\
\log(p_{0i}/(1 - p_{0i} - p_{1i})) &= \rho_0 + \rho_4 z_i\\
\log(p_{1i}/(1 - p_{0i} - p_{1i})) &= \tau_0 + \tau_4 z_i,
\end{aligned} \tag{3.33}$$

where $l = 1, 2, 3, 4,\ (\beta_1)_1 = (\beta_1)_3 = (\beta_1)_4 = 0,\ (\kappa_1)_1 = (\kappa_1)_3 = 0.$

Given the reduced models presented, now, we choose the link function of the models. The information criteria (IC), see Table 14, indicate, for ZOABR-SPAM, in general, the loglog link is the best one. For the ZOAB-SPAM, that the cauchit link was the indicate and for the ZOAS-SPAM, except for AIC and AICC, indicate the logit as the best link. Then, the final three models are ZOABR-SPAM using loglog, ZOAS-SPAM using logit and ZOAB-SPAM using cauchit as link function for mean. Therefore we compared the three respective models, using the model fit assessment tools.

The QQ plots with envelope for the quantile residuals related to the three models indicate that both ZOAS-SPAM an ZOABR-SPAM did not fit well to the data. Also, seen the all the IC measures, regardless of the link function indicate the ZOABR-SPAM. Hence, the ZOABR-SPAM was the model chosen to continue with the analysis, because seems the one more appropriate for this the data set. Then, we continue with the analysis of the reduced model Equation (3.33) using $g_1(.)$ as loglog, to verify the fit quality.

Table 15 shows the estimates of parameters, the associated standard error (SE) and the associated confidence interval (IC) with a confidence of 95%. Figure 20 shows non-parametric curve. The effective degrees of freedom were 4.39 the estimated λ using GCV was 15263. From Figure 21, it is possible to see that the residuals are well within the the confidence bands of the QQ plot. The other residuals plots show no tendency as well the normality seems to be reasonable according to the histogram.

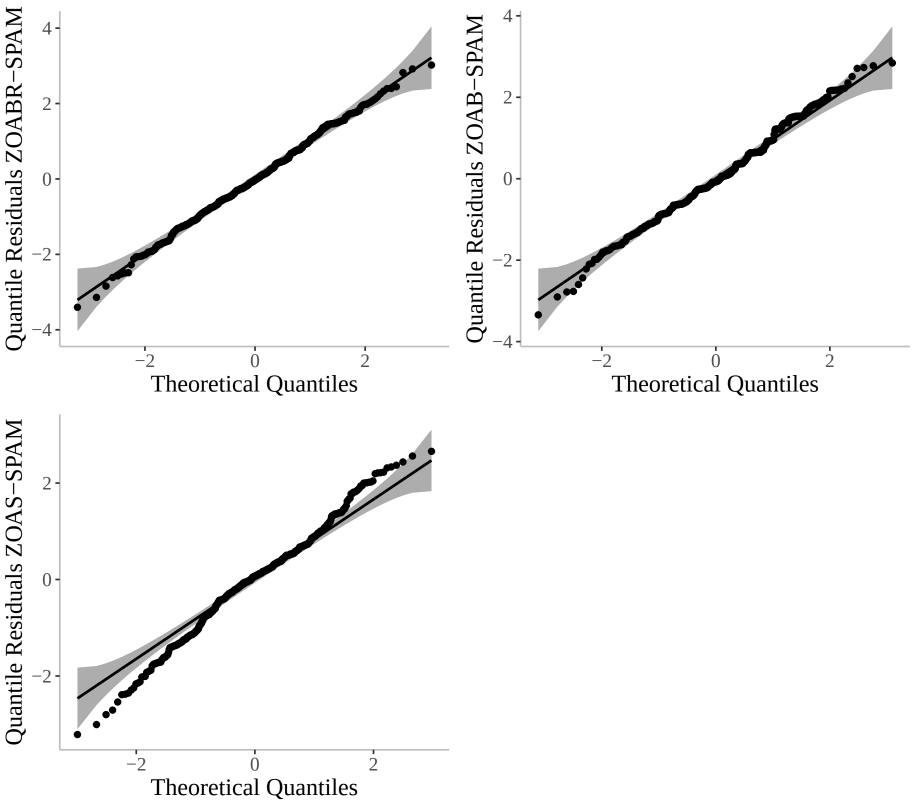


Figure 19 – Quantile-quantile plot with 95% envelopes for all models.

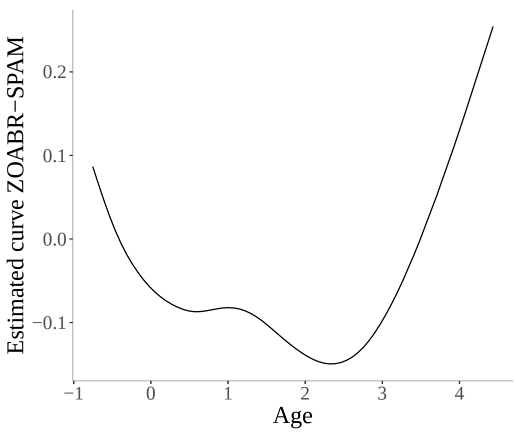


Figure 20 – Estimated curves for ZOABR-SPAM.

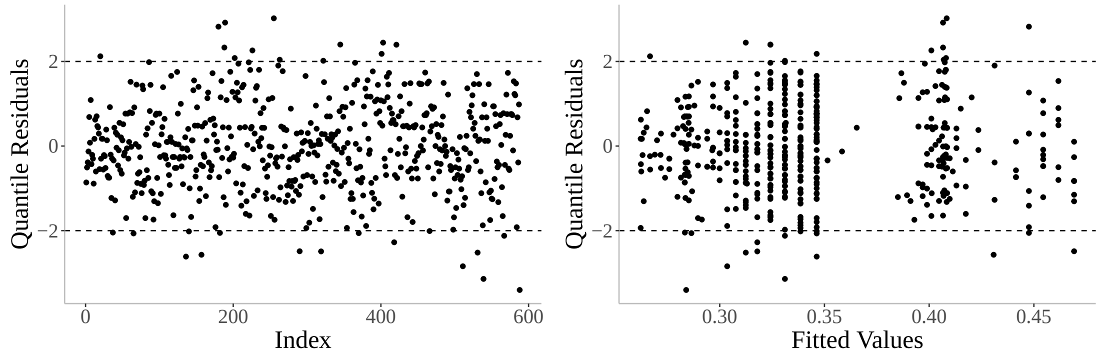
Table 14 – Information Criteria for the three models.

Models	Link function	AIC	AICC	SABIC	HQIC	BIC
ZOABR-SPAM	cauchit	284.96	285.62	261.92	307.62	343.11
	logit	269.84	270.47	247.40	291.90	326.47
	probit	274.81	275.52	250.97	298.27	335.00
	cloglog	287.47	288.17	263.72	310.82	347.39
	loglog	**269.20**	**269.97**	**244.24**	**293.73**	**332.16**
ZOAB-SPAM	cauchit	**507.42**	**507.61**	**515.84**	**519.36**	**538.08**
	logit	507.42	507.61	515.84	519.37	538.08
	probit	507.44	507.63	515.87	519.40	538.15
	cloglog	507.42	507.62	515.85	519.37	538.10
	loglog	507.44	507.64	515.88	519.41	538.15
ZOAS-SPAM	cauchit	607.29	607.86	622.09	628.28	661.17
	logit	606.07	606.64	**620.84**	**627.03**	**659.87**
	probit	606.49	607.12	622.06	628.57	663.16
	cloglog	607.25	607.84	622.30	628.60	662.05
	loglog	**605.39**	**606.02**	621.01	627.55	662.26

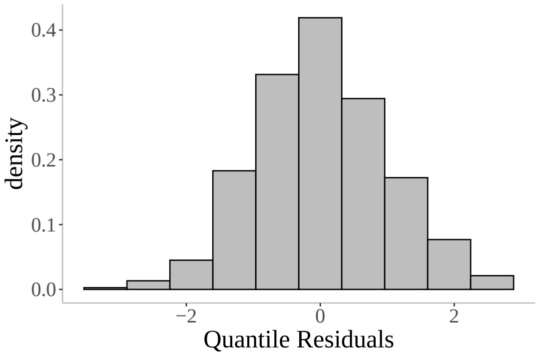
Table 15 – Inferential results for the ZOABR-SPAM.

Parameters	Estimate	SE	p-value	$IC_{95\%}$
ρ_0	-1.75	0.12	<0.0001	[-1.98;-1.51]
ρ_4	0.32	0.11	0.0021	[0.12;0.53]
τ_0	-3.30	0.25	<0.0001	[-3.79;-2.80]
τ_4	0.60	0.17	0.0003	[0.28;0.93]
β_0	-0.15	0.09	0.12308	[-0.33;0.04]
$(\beta_1)_2$	0.34	0.12	0.0040	[0.11;0.57]
κ_0	1.65	0.20	<0.0001	[1.26;2.04]
$(\kappa_1)_2$	-0.76	0.26	0.0032	[-1.28;-0.26]
$(\kappa_1)_4$	-0.42	0.19	0.0254	[-0.78;-0.05]
α	0.40	0.33	-	[-0.24;1.05]

From Figure 22 (left plot), which presents the local influence analysis for the discrete part, we see no flagged observations. On the other hand, from Figure 22 (right plot), concerning the local influence analysis of the continuous part, under the case-weight perturbation scheme, the observation #32 was flagged.



(a) Index plot of quantile residuals. (b) Fitted values versus quantile residuals.



(c) Histogram of quantile residuals.

Figure 21 – Results of fitted ZOABR-SPAM.

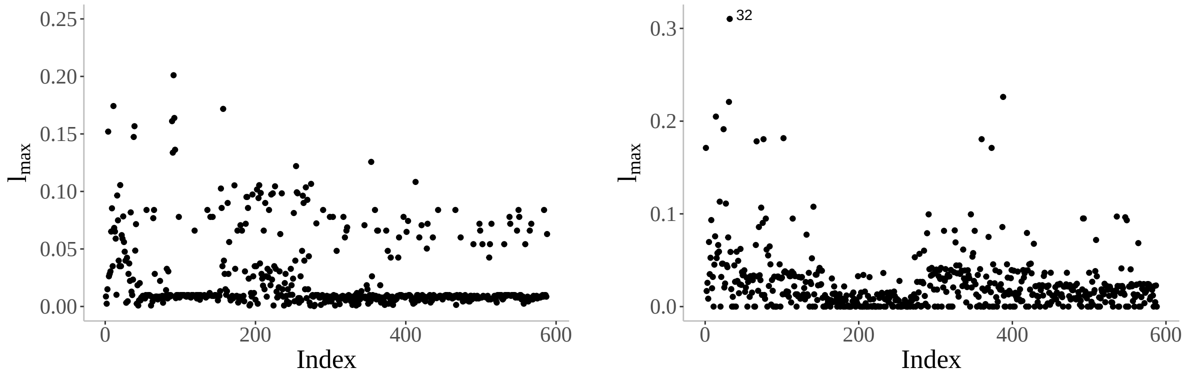


Figure 22 – Index plot of $\boldsymbol{l}_{max}$ under case-weight perturbation for discrete part (left plot) and for continuous part (right plot) for ZOABR-SPAM.

After removing such observation, the significance of the parameters did not change, neither the Cook's distance, neither the generalized leverage. Also, the behavior of the residuals remained the same.

In conclusion, once that the exclusion of the flagged observation did not lead to significant changes on the inference, no further changes are necessary. From the final model it possible to notice that the age of participant has a positive influence on the discrete values of the Risk (Risk=0 and Risk=1). The Ethnic in the categories African-American and

Taiwanese-American is a significant covariate attributed to the heterogeneous dispersion. From the fitted non-parametric curve, the Age variable presented a decreasing in risk for the younger ages and, then, an increasing in risk perception from older ages.

In the analysis of this data made made by Silva et al. (2020), only parametric components were included, and the BR distribution was considered. Their final model did not include the covariate Age. Then, comparing with our approach, our gain is include a component non-parametric in the linear predictor, which implies in more flexibility for our approach. At our work, we could include the age in the fit of the mean, and made a properly fit, given the non-linear behaviour of this covariate. The age after the analysis indicates contributing significantly to the model.

4 Semi-parametric additive models for dependent and limited data

In many fields of knowledge is usual to observe several measurements in the same experimental units, along different (so-called) evaluation conditions. These kind of data is named repeated measurement data and, when these measurements can not be mutually randomized along the evaluation conditions (as time) they are called longitudinal data. Due to this structure is expected to observe within-correlation dependence (correlation).

It is well know that to ignore such structure can lead to the misleading inference see, for example, Diggle et al. (2002). One of the most important ways to model this kind of data is the so-called Generalized Estimating Equations (GEE). This approach is based on the Estimating Equations (EE) (Godambe, 1991). The EE are function of the sample and the parameters of interest. Within the (G)EE framework conditions are seeking to guarantee that the estimators have good properties. Based on that, Liang and Zeger (1986) proposed GEE for mean estimation in models where the response belongs to Exponential Linear Family.

In this Chapter we developed a GEE approach as an extension of the Semi-parametric Additive Models for correlated bounded data, considering the beta, the simplex and the beta rectangular. Then, we will build a suitable framework that considers both the bounded nature of the response variable and the within-subject dependence. The proposal models will be the Semi-parametric Additive Beta (Beta-SPAM-GEE), the Semi-parametric Additive Simplex (Simplex-SPAM-GEE), and the Semi-parametric Additive BR (BR-SPAM-GEE) models via GEE.

4.1 Estimating functions

In this subsection, it will be reviewed some definitions and quantities related with GEE methodology required for the next sections.

The quantity $\boldsymbol{\psi}$ is defined as an estimation function of the random variable $\boldsymbol{Y}$ and the parameters of interest $\boldsymbol{\theta}$ if, for each $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\psi}(\boldsymbol{\theta}; \boldsymbol{Y}) = (\psi_1, ..., \psi_p)^\top$ is a random variable, where $\Theta \subseteq \mathbb{R}^p$ is the parameter space. Only regular cases will be considered, which means that Θ has a finite dimension p and the true parameter $\boldsymbol{\theta_0}$ is an inner point of Θ.

Assuming a random sample of n independent random vectors $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{it_i})^\top$, $i = 1, ..., n$ and each of them is related to an estimation function $\boldsymbol{\psi}_i$, then a sample estimating function $\boldsymbol{\Psi}(\boldsymbol{\theta})$ is given by $\boldsymbol{\Psi}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\boldsymbol{Y};\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\psi}_i(\boldsymbol{Y}_i;\boldsymbol{\theta})$, where $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, ..., \boldsymbol{Y}_n^\top)^\top$. Besides that, only focusing on the estimation functions whose roots are estimators of the parameters of interest, i.e., $\boldsymbol{\Psi}(\boldsymbol{Y};\hat{\boldsymbol{\theta}}) = \mathbf{0}$.

Let $Y_1, ..., Y_n$ be a random samples with $\mathbb{E}(Y_i) = \mu_i(\boldsymbol{\theta})$, μ_i doubly differentiable with relation to $\boldsymbol{\theta}$ and $\text{Var}(Y_i) = \sigma^2$, then

$$\boldsymbol{\Psi}(\boldsymbol{Y};\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}[Y_i - \mu_i(\boldsymbol{\theta})] = \mathbf{0} \tag{4.1}$$

is denoting by estimating equation.

Furthermore, if $\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\Psi}(\boldsymbol{\theta})] = \mathbf{0}$, $\forall \boldsymbol{\theta} \in \Theta$, the function $\Psi(\boldsymbol{\theta})$ is an unbiased. Thus, if all estimating functions $\boldsymbol{\psi}_i$ are unbiased, then $\boldsymbol{\Psi}(\boldsymbol{\theta})$ will also unbiased. Moreover, let $\boldsymbol{\Psi}(\boldsymbol{\theta})$ be an unbiased estimating function, then, the related variability and sensibility matrices (both $p \times p$), are given, respectively, by:

$$\boldsymbol{V}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\Psi}(\boldsymbol{\theta})\boldsymbol{\Psi}^\top(\boldsymbol{\theta})] \quad \text{and} \quad \boldsymbol{S}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{\partial}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Psi}(\boldsymbol{\theta})\right]. \tag{4.2}$$

On the other hand, the Godambe Information Matrix of $\boldsymbol{\theta}$ associated to a regular estimating function $\boldsymbol{\Psi}$ is given by $\boldsymbol{J}(\boldsymbol{\theta}) = \boldsymbol{S}^\top(\boldsymbol{\theta})\boldsymbol{V}^{-\mathbf{1}}(\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{\theta})$. The Godambe information matrix plays a similar role to the Fisher information matrix, i.e., the former is related to the information about the variability of the estimators. It can be noticed that if $\boldsymbol{S}(\boldsymbol{\theta}) = -\boldsymbol{V}(\boldsymbol{\theta})$, then the Godambe Information Matrix coincides with the Fisher Information Matrix.

Let $Q_i(\boldsymbol{\theta})$, $i = 1, ..., n$ be non stochastic matrices and $u_i = u_i(y_i;\boldsymbol{\theta})$ be a zero mean vector mutually independents, an estimating function class is said to be additive or linear if (Crowder, 1987):

$$\ell(u) = \left\{\boldsymbol{\Psi_n} \in \Re : \boldsymbol{\Psi(\theta)} = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta})u_i(y_i;\boldsymbol{\theta})\right\}. \tag{4.3}$$

A regular estimating function (see Artes (2005) for more details) is said to be optimal if its roots have minimal asymptotic variance. The optimal element within the class of linear estimating functions according to Crowder (1987), is given by:

$$\boldsymbol{\Psi}^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} Q_i^*(\boldsymbol{\theta})u_i(y_i;\boldsymbol{\theta}), \tag{4.4}$$

where $Q_i^*(\boldsymbol{\theta}) = \mathbb{E}\left(\frac{\partial u_i}{\partial \boldsymbol{\theta}^\top}\right)^\top \mathbf{Cov}(u_i)^{-1}$ and

$$\mathbf{Cov}(u_i) = \text{diag}\{\mathbf{Var}(u_i)^{1/2}\}\boldsymbol{R}^v(u_i)\text{diag}\{\mathbf{Var}(u_i)^{1/2}\}, \tag{4.5}$$

being $\boldsymbol{R}^v(u_i)$ the actual correlation matrix of u_i, for $i = 1, \ldots, n$.

It is important to notice that this previous definition also is applied to the estimating function optimality. The following theorems establish conditions that guarantee the asymptotic normality of the estimators obtained from the regular estimating functions.

According to Jørgensen and Labouriau (1995), let $\boldsymbol{\Psi} : \Omega \times \Theta \to \mathbb{R}^p$ be a regular estimating function and $\{\hat{\theta}_n\}_{n\geq 1}$ a sequence of estimators satisfying $\boldsymbol{\Psi}(\boldsymbol{y}; \hat{\boldsymbol{\theta}}) = \mathbf{0}$, and suppose that $\exists\ \theta \in \Theta : \hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$, where $\hat{\theta}_n$ is asymptotically Normal, then $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{J}^{-1}(\boldsymbol{\theta}))$, where

$$\boldsymbol{J}(\boldsymbol{\theta}) = \lim_{n\to\infty} \frac{1}{n} \{\boldsymbol{S}^\top(\boldsymbol{\theta})\boldsymbol{V}^{-1}(\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{\theta})\},$$

which replaces the asymptotic Godambe information matrix. Here the symbol "$\xrightarrow{D}$" stands for the convergence in distribution (related to $\mathbb{P}_\theta$) and "$\xrightarrow{\mathbb{P}}$", the convergence in probability.

Under certain conditions (see Sen and Singer (1994) for more details), it possible to show that $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}$ and $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{J}^{-1}(\boldsymbol{\theta}))$. A proof for this theorem can be seen in Jørgensen and Labouriau (1995), for example. It is important to cite that these conditions are generalizations of the Frechet-Cramer-Rao regularity conditions Sen et al. (2010).

In practice, the actual correlation matrix $\boldsymbol{R}^v$ is unknown. Liang and Zeger (1986) proposed to use a so-called working correlation matrix $\boldsymbol{R}_i(\boldsymbol{\xi})$, which depends on $\boldsymbol{\xi}$. Thus, as said previously, the following models will be developed based on this last approach. Next sections will discuss more about the GEE and their extensions for the purpose of this work.

4.2 Models

Let $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \ldots, \boldsymbol{Y}_n^\top)^\top$, with $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i^*})^\top$, $i = 1, \ldots, n$ be the response vector for all subjects in all time that they were collected, and considering the marginal distribution of Y_{it} as $Y_{it} \sim \text{beta}(\mu_{it}, \phi_{it}), \text{simplex}(\mu_{it}, \phi_{it})$ or $\text{BR}(\mu_{it}, \phi_{it}, \alpha)$, for $t = 1, \ldots, n_i^*$, $i = 1, \ldots, n$. The Semi-parametric Additive Model via GEE (SPAM-GEE) for any of these three distributions, Beta-SPAM-GEE, Simplex-SPAM-GEE and BR-SPAM-GEE, have the following systematic component:

$$\begin{aligned} \boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \boldsymbol{X\beta} + \sum_{j=1}^{k} h_j(\boldsymbol{z}_j), \\ \boldsymbol{\eta}_2 &= g_2(\boldsymbol{\phi}) = \boldsymbol{E\kappa} \end{aligned} \tag{4.6}$$

where $n^{**} = \sum_{i=1}^{n} n_i^* n$ is the sample size, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_n^\top)^\top$, with $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i^*})^\top$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_n^\top)^\top$, with $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{in_i^*})^\top$, $g_1(.)$ and $g_2(.)$ are strictly monotone and twice differentiable link functions, such that, $g_1 : (0, 1) \to \mathbb{R}$ and $g_2 : \mathbb{R}^+ \to \mathbb{R}$, $\boldsymbol{\eta}_1 = (g_1(\boldsymbol{\mu}_1)^\top, \ldots, g_1(\boldsymbol{\mu}_n)^\top)^\top$, $\boldsymbol{\eta}_2 = (g_2(\boldsymbol{\phi}_1)^\top, \ldots, g_2(\boldsymbol{\phi}_n)^\top)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$

and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)^\top \in \mathbb{R}^s$ are, respectively, a p-vector and a s-vector of unknown parameters, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top)^\top$, with $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i^*})^\top$ and $\boldsymbol{E} = (\boldsymbol{E}_1^\top, \ldots, \boldsymbol{E}_n^\top)^\top$, with $\boldsymbol{E}_i = (\boldsymbol{e}_{i1}, \ldots, \boldsymbol{e}_{in_i^*})^\top$ are design matrices of dimension, respectively, $n^{**} \times p$ and $n^{**} \times s$, in which $\boldsymbol{x}_i = (x_{it1}, \ldots, x_{itp})^\top$ and $\boldsymbol{e}_i = (e_{it1}, \ldots, e_{its})^\top$ for $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$, $\boldsymbol{z}_j = (\boldsymbol{z}_{1j}^\top, \ldots, \boldsymbol{z}_{nj}^\top)^\top, j = 1, \ldots, k$, where $\boldsymbol{z}_{ij} = (z_{i1j}, \ldots, z_{in_i^*j})$ is a specification vector associated with non-parametric components and $h_1, \ldots, h_k$ are an unknown smooth functions.

The $h_j(\boldsymbol{z}_j)$ can be rewritten as discussed before in the Section 1.2.1 as

$$h_j(\boldsymbol{z}_j) = \sum_{l=1}^{q_j} \gamma_{lj} b_{lj}(\boldsymbol{z}_j), \tag{4.7}$$

with γ_{lj} being the coefficients to be estimated, b_{lj}'s being l-th cubic B-spline and q_j is the number of cubic B-splines both related to the j-th P-spline at the vector $\boldsymbol{z}_j$, $j = 1, \ldots, k$ and $l = 1, \ldots, q_j$. As cubic B-splines are being adopted here, $b_{lj} = B_{l,d_l}$ with $d_l = 4$ as defined in Section 1.2.1, $k_j = d_l + q_j$ is the number of knots. It is possible to rewrite Equation (4.6) for the i-th observation, $i = 1, \ldots, n$, as follows

$$\boldsymbol{\eta}_{1i} = g_1(\boldsymbol{\mu}_i) = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \boldsymbol{b}_{i1}^\top \boldsymbol{\gamma}_1 + \cdots + \boldsymbol{b}_{ik}^\top \boldsymbol{\gamma}_k,$$

where $\boldsymbol{b}_{ij} = (b_{1j}(\boldsymbol{z}_{ij}), \ldots, b_{q_j j}(\boldsymbol{z}_{ij}))^\top$ and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{jq_j})^\top$, for $j = 1, \ldots, k$. Return to matrix form of $\boldsymbol{\eta}_1$, the predictors of the model can be rewritten as

$$\boldsymbol{\eta}_1 = g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j, \quad \boldsymbol{\eta}_2 = g_2(\boldsymbol{\phi}) = \boldsymbol{E}\boldsymbol{\kappa}, \tag{4.8}$$

where $\boldsymbol{B}_j$ is a matrix of dimension $n^{**} \times q_j$ composed by $b_{lj}(\boldsymbol{z}_{ij})$, for $i = 1, \ldots, n$, $j = 1, \ldots, k$ and $l = 1, \ldots, q_j$.

4.3 Penalized GEE

Based on the proposal of Liang and Zeger (1986) described in Section 4.1, it will be developed two distinct GEE for the mean and dispersion/precision parameters as made by X.-K. Song et al. (2004) and Freitas et al. (2021), generating the (models): Simplex-SPAM-GEE, Beta-SPAM-GEE and BR-SPAM-GEE. Indeed, given $\boldsymbol{\nu} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top)^\top$ and $\boldsymbol{\kappa}$, let us consider the vectors $\boldsymbol{d}_i^* = (\boldsymbol{u}_i^\top, \boldsymbol{q}_i^\top)^\top$, $\boldsymbol{u}_i = \boldsymbol{u}_i(\boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\kappa}) = (u_{i1}, \ldots, u_{in_i^*})^\top$ and $\boldsymbol{q}_i = \boldsymbol{q}_i(\boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\kappa}) = (q_{i1}, \ldots, q_{in_i^*})^\top$, $i = 1, \ldots, n$, where $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ are the zero mean vectors defined in Equation (4.3), related to μ_{it} and ϕ_{it}, respectively. They arel also mutually independent, satisfying the properties of regular estimating functions (Crowder, 1987).

A common choice for $\boldsymbol{u}_i$ is the score function, which will be done for the Simplex-SPAM-GEE. However, for the Beta-SPAM-GEE and BR-SPAM-GEE we adopt $\boldsymbol{u}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_i$. This choice implies a better interpretability for the correlation matrix of $\boldsymbol{u}_i$, since it becomes the correlation of $\boldsymbol{Y}$. Also, as we will show in the next sections, it implies in the insensitivity property (Jørgensen and Knudsen, 2004), which allows the updating the GEE of the mean and dispersion/precision parameters, in the estimation algorithm, separately.

4.3.1 Components for Beta-SPAM-GEE

Assuming that the marginal distributions follow the beta distribution, the log-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$, is given by:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \Bigg(\log(\Gamma(\boldsymbol{\phi}_i)) - \log(\Gamma(\boldsymbol{\mu}_i \boldsymbol{\phi}_i)) - \log(\Gamma([1-\boldsymbol{\mu}_i]\boldsymbol{\phi}_i)) + [\boldsymbol{\mu}_i\boldsymbol{\phi}_i - 1]\log(y_i) + ([1-\boldsymbol{\mu}_i]\boldsymbol{\phi}_i - 1)\log[1-\boldsymbol{y}_i] \Bigg),$$

where, $\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$, $\boldsymbol{\eta}_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{b}_{ij}^\top \boldsymbol{\gamma}_j$, $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$ and $\boldsymbol{\eta}_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$. Also, we have that $\boldsymbol{q}_i$ corresponds to the score function. Hence, $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ are given by,

$$\begin{aligned} \boldsymbol{u}_i &= \boldsymbol{y}_i - \boldsymbol{\mu}_i, \\ \boldsymbol{q}_i &= \boldsymbol{\mu}_i(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) + \log(1-\boldsymbol{y}_i) - \Psi((1-\boldsymbol{\mu}_i)\boldsymbol{\phi}_i) + \Psi(\boldsymbol{\phi}_i), \end{aligned} \tag{4.9}$$

where $\boldsymbol{\mu}^* = (\mu_{i1}^*, \ldots, \mu_{in_i^*}^*)^\top$, $\mu_{it}^* = \Psi(\mu_{it}\phi_{it}) - \Psi[(1-\mu_{it})\phi_{it}]$, $\boldsymbol{y}^* = (y_{i1}^*, \ldots, y_{in_i^*}^*)^\top$ and $y_{it}^* = \log[y_{it}/(1-y_{it})]$, for $i = 1, \ldots, n$ and $t = 1, \ldots, n_i^*$, $\Psi(.)$ denotes the digamma function, that is, $\Psi(z) = \partial \log \Gamma(z)/\partial z$, for $z > 0$. It can be noticed that $\mathbb{E}(u_{it}) = 0$ and, since $\boldsymbol{q}_i$ is the score function related to $\boldsymbol{\phi}_i$, $\mathbb{E}(q_{it}) = 0 \Leftrightarrow \mathbb{E}(\log(1-y_{it})) = \Psi((1-\mu_{it})\phi_{it}) + \Psi(\phi_{it})$. Thus, the use of $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ generate an optimal class of estimating functions and they present zero mean, are mutually independent and satisfying the properties of regular estimating functions, see (Crowder, 1987). For the i-th subject, some quantities, which will be used in the next sections are defined below,

$$\begin{aligned}
\boldsymbol{T}_{1i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^\top = \boldsymbol{G}_{1i}, \\
\boldsymbol{T}_{2i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{G}_{2i}\boldsymbol{C}_i, \\
\boldsymbol{T}_{21i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{0}, \\
\boldsymbol{T}_{12i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_{1i}}\right)^\top = \boldsymbol{G}_{1i}\boldsymbol{D}_i,
\end{aligned}$$

and $\boldsymbol{\Sigma}_i = \mathrm{Cov}(\boldsymbol{u}_i) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\xi})\boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \mathrm{diag}\{a_{i1}, \ldots, a_{in_i^*}\}$, $a_{it} = Var(Y_{it}) = [\mu_{it}(1-\mu_{it})]/(1+\phi_{it})$, $\boldsymbol{G}_{1i} = \mathrm{diag}\{(\partial\mu_{i1}/\partial\eta_{1i1}), \ldots, (\partial\mu_{in_i^*}/\partial\eta_{1in_i^*})\}$, $\boldsymbol{G}_{2i} = \mathrm{diag}\{(\partial\phi_{i1}/\partial\eta_{2i1}), \ldots, (\partial\phi_{in_i^*}/\partial\eta_{2in_i^*})\}$, $\boldsymbol{C}_i = \mathrm{diag}\{c_{i1}, \ldots, c_{in_i^*}\}$, $c_{it} = -\Psi'(\phi_{it}) + \mu_{it}^2\Psi'(\mu_{it}\phi_{it}) + (1-\mu_{it})^2\Psi'((1-$

$\mu_{it})\phi_{it})$, $\boldsymbol{D}_i = \text{diag}\{d_{i1}, \ldots, d_{in_i^*}\}$, and $d_{it} = \phi_{it}[\mu_{it}\Psi'(\mu_{it}\phi_{it}) - (1-\mu_{it})\Psi'((1-\mu_{it})\phi_{it})]$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. Furthermore, $\boldsymbol{R}(\boldsymbol{\xi})_{(n_i^* \times n_i^*)}$ is symmetric matrix satisfying the conditions for a working correlation matrix (Liang and Zeger, 1986), which depends on a correlation parameter vector $\boldsymbol{\xi}$ that is the same for all subjects.

4.3.2 Components for Simplex-SPAM-GEE

For the model based (marginal distributions) on the simplex distribution, the log-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$, is defined as:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(-\frac{1}{2}\left\{\log(2\pi\boldsymbol{\phi}_i) + 3\log[\boldsymbol{y}_i(1-\boldsymbol{y}_i)]\right\} - \frac{1}{2\boldsymbol{\phi}_i}d(\boldsymbol{y}_i;\boldsymbol{\mu}_i)\right),$$

where, $d(y;\mu)$ is defined in Equation (1.7), $\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$, $\boldsymbol{\eta}_{1i} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{k}\boldsymbol{b}_{ij}^\top\boldsymbol{\gamma}_j$, $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$ and $\boldsymbol{\eta}_{2i} = \boldsymbol{e}_i^\top\boldsymbol{\kappa}$. We have that $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ will be defined also as the score function. They are given by,

$$\boldsymbol{u}_i = -\frac{1}{2\boldsymbol{\phi}_i}d'(\boldsymbol{y}_i,\boldsymbol{\mu}_i), \quad \boldsymbol{q}_i = \frac{1}{2\boldsymbol{\phi}_i}\left[\frac{1}{\boldsymbol{\phi}_i}d(\boldsymbol{y}_i,\boldsymbol{\mu}_i) - 1\right], \tag{4.10}$$

where

$$d'(\boldsymbol{y}_i,\boldsymbol{\mu}_i) = \frac{\partial d(\boldsymbol{y}_i,\boldsymbol{\mu}_i)}{\partial \boldsymbol{\mu}_i} = -\frac{2(\boldsymbol{y}_i - \boldsymbol{\mu}_i)}{\boldsymbol{\mu}_i(1-\boldsymbol{\mu}_i)}\left[d(\boldsymbol{y}_i,\boldsymbol{\mu}_i) + \frac{1}{\boldsymbol{\mu}_i^2(1-\boldsymbol{\mu}_i)^2}\right],$$

$t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. It can be noticed that $\mathbb{E}(u_{it}) = 0 \Leftrightarrow \mathbb{E}(d'(y_{it},\mu_{it})) = 0$ and $\mathbb{E}(q_{it}) = 0 \Leftrightarrow \mathbb{E}(d(y_{it},\mu_{it})) = \phi_{it}$. Thus, the vectors $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ generate an optimal class of functions and they are vector with zero mean, mutually independent and satisfying the properties of regular estimating functions (Crowder, 1987). For the i-th subject, some quantities, which will be used in the next sections, are defined below

$$\begin{aligned}
\boldsymbol{T}_{1i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^\top = \boldsymbol{G}_{1i}\boldsymbol{A}_i \\
\boldsymbol{T}_{2i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{G}_{2i}\boldsymbol{C}_i, \\
\boldsymbol{T}_{21i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = \boldsymbol{T}_{12i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = \boldsymbol{0},
\end{aligned}$$

and $\boldsymbol{\Sigma}_i = \text{Cov}(\boldsymbol{u}_i) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\xi})\boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \text{diag}\{a_{i1}, \ldots, a_{in_i^*}\}$, $a_{it} = \phi_{it}^{-1}[(3\phi_{it})/(\mu_{it}(1-\mu_{it})) + 1/(\mu_{it}^3(1-\mu_{it})^3)]$, $\boldsymbol{G}_{1i} = \text{diag}\{(\partial\mu_{i1}/\partial\eta_{1i1}), \ldots, (\partial\mu_{in_i^*}/\partial\eta_{1in_i^*})\}$, $\boldsymbol{G}_{2i} = \text{diag}\{(\partial\phi_{i1}/\partial\eta_{2i1}), \ldots, (\partial\phi_{in_i^*}/\partial\eta_{2in_i^*})\}$, $\boldsymbol{C}_i = \text{diag}\{c_{i1}, \ldots, c_{in_i^*}\}$, $c_{ij} = (2\phi_{it}^2)^{-1}$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. Furthermore, $\boldsymbol{R}(\boldsymbol{\xi})$ is a $(n_i^* \times n_i^*)$ symmetric matrix satisfying the conditions for a working correlation matrix (Liang and Zeger, 1986), which depends on a correlation parameter vector $\boldsymbol{\xi}$ that is the same for all subjects.

4.3.3 Components for BR-SPAM-GEE

Assuming marginal BR distribution, the log-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$, is given by:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log[\boldsymbol{\varepsilon}_i + (1-\boldsymbol{\varepsilon}_i) b_{Y_i}(\boldsymbol{y}_i; \boldsymbol{\delta}_i, \boldsymbol{\phi}_i)], \tag{4.11}$$

where $\boldsymbol{\varepsilon}_i = 1 - \sqrt{1 - 4\alpha\boldsymbol{\mu}_i(1-\boldsymbol{\mu}_i)}$, $b_Y(y; ., .)$ is the beta density defined in Equation (1.1), $\boldsymbol{\delta}_i = \dfrac{\boldsymbol{\mu}_i - \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\alpha\boldsymbol{\mu}_i(1-\boldsymbol{\mu}_i)}}{\sqrt{1 - 4\alpha\boldsymbol{\mu}_i(1-\boldsymbol{\mu}_i)}}$, $\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$, $\boldsymbol{\eta}_{1i} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{b}_{ij}^\top \boldsymbol{\gamma}_j$, $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$ and $\boldsymbol{\eta}_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$.

Here, the log-likelihood is not easily tractable, since the necessary derivatives are not easy to obtain, see Section 2.2.1. Thus, instead of obtain $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ from the score function, we will consider the proposal of X.-K. Song et al. (2004). They constructed the zero mean vector for the estimating equation of the dispersion component of their distribution, based on a quantity whose expectation was a function of the dispersion parameter. Since we consider the precision parameter of the BR distribution, a useful approach is to consider $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{in_i^*})$, where $q_{it} = (y_{it} - \mu_{it})^2 - \mathrm{Var}(Y_{it})$, $\mathbb{E}[(Y_{it} - \mu_{it})^2] = \mathrm{Var}(Y_{it})$, and $\mathrm{Var}(Y_{it}) = g(\phi_{it})$ is defined in Equation (1.4), r $t = 1, \ldots, n_i^*$. Therefore, we have not only easier calculations, but also good properties of the related estimators. Hence, $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ are given by

$$\begin{aligned} \boldsymbol{u}_i &= \boldsymbol{y}_i - \boldsymbol{\mu}_i, \\ \boldsymbol{q}_i &= (\boldsymbol{y}_i - \boldsymbol{\mu}_i)^2 - \left\{ \frac{\boldsymbol{\delta}_i(1-\boldsymbol{\delta}_i)}{1+\boldsymbol{\phi}_i}(1-\boldsymbol{\varepsilon}_i)[1 - \boldsymbol{\varepsilon}_i(1+\boldsymbol{\phi}_i)] + \frac{\boldsymbol{\varepsilon}_i}{12}(4 - 3\boldsymbol{\varepsilon}_i) \right\}, \end{aligned} \tag{4.12}$$

$i = 1, \ldots, n$. It can be noticed that $\mathbb{E}(\boldsymbol{u}_i) = \boldsymbol{0}$ and $\mathbb{E}(\boldsymbol{q}_i) = \boldsymbol{0}$. Thus, the vectors $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ generate an optimal class of functions and they are vector with zero mean, mutually independent and satisfying the properties of regular estimating functions (Crowder, 1987). For the i-th subject, some quantities, which will be used in the next sections, are defined below

$$\begin{aligned} \boldsymbol{T}_{1i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^\top = \boldsymbol{G}_{1i} \\ \boldsymbol{T}_{2i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{G}_{2i}\boldsymbol{C}_i, \\ \boldsymbol{T}_{21i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{0}, \\ \boldsymbol{T}_{12i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_{1i}}\right)^\top = \boldsymbol{G}_{1i}D_i, \end{aligned}$$

where $\boldsymbol{\Sigma}_i = \text{Cov}(\boldsymbol{u}_i) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\xi})\boldsymbol{A}_i^{1/2}$, $\boldsymbol{A}_i = \text{diag}\{a_{i1}, \ldots, a_{in_i^*}\}$,

$$a_{it} = \frac{\boldsymbol{\delta}_i(1-\boldsymbol{\delta}_i)}{1+\boldsymbol{\phi}_i}(1-\boldsymbol{\varepsilon}_i)[1-\varepsilon(1+\boldsymbol{\phi}_i)] + \frac{\boldsymbol{\varepsilon}_i}{12}(4-3\boldsymbol{\varepsilon}_i),$$

$\boldsymbol{G}_{1i} = \text{diag}\{(\partial\mu_{i1}/\partial\eta_{1i1}), \ldots, (\partial\mu_{in_i^*}/\partial\eta_{1in_i^*})\}$, $\boldsymbol{G}_{2i} = \text{diag}\{(\partial\phi_{i1}/\partial\eta_{2i1}), \ldots, (\partial\phi_{in_i^*}/\partial\eta_{2in_i^*})\}$, $\boldsymbol{C}_i = \text{diag}\{c_{i1}, \ldots, c_{in_i^*}\}$, $c_{ij} = [(\delta_{it}(1-\delta_{it}))/(1+\phi_{it})^2](\varepsilon_{it}-1)$, $\boldsymbol{D}_i = \text{diag}\{d_{i1}, \ldots, d_{in_i^*}\}$,

$$d_{it} = (1-2\delta_{it})\frac{(1-\alpha)[1-\varepsilon_{it}(1+\phi_{it})]}{(1+\phi_{it})(1-\varepsilon_{it})^2} + \left[\frac{\delta_{it}(1-\delta_{it})}{(1+\phi_{it})}(-2-\phi_{it}+2\varepsilon_{it}(1+\phi_{it})) + \frac{2-3\varepsilon_{it}}{6}\right]\left(\frac{2\alpha(1-2\mu_{it})}{1-\varepsilon_{it}}\right),$$

$t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. Furthermore, $\boldsymbol{R}(\boldsymbol{\xi})$ is a $(n_i^* \times n_i^*)$ symmetric matrix satisfying the conditions for a working correlation matrix (Liang and Zeger, 1986), which depends on a correlation parameter vector $\boldsymbol{\xi}$ that is the same for all subjects. Finally, also following X.-K. Song et al. (2004) we adopt $\boldsymbol{\Upsilon}_i = \text{Cov}(\boldsymbol{q}_i) = \boldsymbol{J}_i = \text{diag}\{j_{i1}, \ldots, j_{in_i^*}\}$, where $j_{it} = \text{Var}[(Y_{it}-\mu_{it})^2]$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. For ease, the $\text{Var}[(Y_{it}-\mu_{it})^2]$, in the estimation process, will be obtained numerically.

4.3.4 Penalized GEE approach

As discussed in Section 2.2 and following Manghi et al. (2019), it is necessary adding a penalization in the GEE related with the non-parametric components. A penalty function based on differences between adjacent B-splines, as defined in Section 1.2.2, will be used here. Thus, considering the components related to each distribution as defined in the previous sections, the generalized estimating equations for $\boldsymbol{\nu} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \gamma_k^\top)^\top$ and $\boldsymbol{\kappa}$, respectively are given by:

$$\begin{aligned}\boldsymbol{\Psi}_{1p}(\boldsymbol{\nu}, \boldsymbol{\lambda}) &= \sum_{i=1}^{n} -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\nu}^\top}\right)^\top \text{Cov}(\boldsymbol{u}_i)^{-1}\boldsymbol{u}_i - \boldsymbol{\Lambda} = \boldsymbol{B}_i^{*\top}\boldsymbol{T}_{1i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{u}_i - \boldsymbol{\Lambda} \\ \boldsymbol{\Psi}_2(\boldsymbol{\kappa}) &= \sum_{i=1}^{n} -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\kappa}^\top}\right)^\top \text{Cov}(\boldsymbol{q}_i)^{-1}\boldsymbol{q}_i = \boldsymbol{E}_i^\top\boldsymbol{T}_{2i}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{q}_i,\end{aligned} \tag{4.13}$$

where $\boldsymbol{B}_i^* = (\boldsymbol{X}_i, \boldsymbol{B}_{ij}, \ldots, \boldsymbol{B}_{ik})$, $\boldsymbol{\Lambda} = (\boldsymbol{0}_p, \lambda_1\boldsymbol{\gamma}_1^\top\boldsymbol{\Lambda}_1^d, \ldots, \lambda_k\boldsymbol{\gamma}_k^\top\boldsymbol{\Lambda}_k^d)^\top$, $\boldsymbol{0}_{(p\times1)}$ is a vector of zeros and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^\top$ is the vector of smooth parameters and $\lambda_j > 0,\ \forall j, j = 1, \ldots, k$, $\boldsymbol{T}_{1i} = \text{diag}\{T_{1i1}, \ldots, T_{1in_i^*}\}$, $\boldsymbol{T}_{2i} = \text{diag}\{T_{2i1}, \ldots, T_{2in_i^*}\}$, $\boldsymbol{\Sigma}_i = \text{diag}\{\Sigma_{i1}, \ldots, \Sigma_{in_i^*}\}$ and $\boldsymbol{\Upsilon}_i = \text{diag}\{\Upsilon_{i1}, \ldots, \Upsilon_{in_i^*}\}$.

We can use a working correlation matrix instead of $\boldsymbol{\Upsilon}_i = \text{Cov}(\boldsymbol{q}_i)$. However this would lead to a large amount of parameters to be estimated. Then, following Freitas et al. (2021) and X.-K. Song et al. (2004), we assume, for the Simplex-SPAM-GEE and Beta-SPAM-GEE that $\boldsymbol{\Upsilon}_i = \boldsymbol{C}_i$ and for BR-SPAM-GEE, $\boldsymbol{\Upsilon}_i = \boldsymbol{J}_i$. As cited by Freitas et al. (2021), by assuming this ($\boldsymbol{\Upsilon}_i = \boldsymbol{C}_i$ or $\boldsymbol{J}_i$), even though we no longer have optimal linear estimation functions, we reduce the number of assumptions to make about the dependency structure of the data. We further assume that $\mathrm{Cov}(\boldsymbol{u}_i, \boldsymbol{q}_i) = 0$, which leads to a small number of parameters to be estimated, for more details see Prentice and Zhao (1991). Then, the penalized GEE for $\boldsymbol{\nu}$ and $\boldsymbol{\kappa}$ are given by $\boldsymbol{\Psi}_{1p}(\hat{\boldsymbol{\nu}}, \boldsymbol{\lambda}) = \boldsymbol{0}$ and $\boldsymbol{\Psi}_2(\hat{\boldsymbol{\kappa}}) = \boldsymbol{0}$, where $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{\kappa}}$ denotes the respective estimates.

For the next sections, unless for parameter estimation developments, it will be useful to rewrite the GEE's as $\boldsymbol{\Psi}^*(\boldsymbol{\nu}, \boldsymbol{\kappa}) = (\boldsymbol{\Psi}_{1p}(\boldsymbol{\nu}, \boldsymbol{\lambda})^\top, \boldsymbol{\Psi}_2(\boldsymbol{\kappa})^\top)^\top$, that is

$$\boldsymbol{\Psi}^*(\boldsymbol{\nu}, \boldsymbol{\kappa}) = \boldsymbol{Q}^\top \boldsymbol{W} \boldsymbol{T}^{-1} \boldsymbol{d}^* - \boldsymbol{\Lambda}_2, \tag{4.14}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{0}_p, \lambda_1 \boldsymbol{\gamma}_1^\top \boldsymbol{\Lambda}_1^d, \ldots, \lambda_k \boldsymbol{\gamma}_k^\top \boldsymbol{\Lambda}_k^d, \boldsymbol{0}_q)^\top$, $\boldsymbol{0}_{(q\times 1)}$ vector of zeros, $\boldsymbol{d}_i^* = (\boldsymbol{u}_i^\top, \boldsymbol{q}_i^\top)^\top$, $\boldsymbol{W} = \boldsymbol{T}_i \boldsymbol{P}_i^{-1} \boldsymbol{T}_i$,

$$\boldsymbol{Q}_i = \begin{pmatrix} \boldsymbol{B}_i^* & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{E}_i \end{pmatrix}, \; \boldsymbol{T}_i = \begin{pmatrix} \boldsymbol{T}_{11i} & \boldsymbol{T}_{12i} \\ \boldsymbol{T}_{21i} & \boldsymbol{T}_{22i} \end{pmatrix}, \boldsymbol{P}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Upsilon}_i \end{pmatrix}.$$

4.4 Parameter estimation

Before presenting the estimation process, it will be demonstrated that the GEE's for $\boldsymbol{\nu}$ and $\boldsymbol{\kappa}$ can be updated separately for all models by using the insensitivity property (Jørgensen and Knudsen, 2004). As defined in a generic way in Equation (4.2), the sensibility matrix can be write as

$$\boldsymbol{S}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{pmatrix},$$

where $\boldsymbol{S}_{11} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\nu}^\top} \boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right]$, $\boldsymbol{S}_{22} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{\Psi}_{2}(\boldsymbol{\kappa})\right]$, $\boldsymbol{S}_{12} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\nu}^\top} \boldsymbol{\Psi}_{2}(\boldsymbol{\kappa})\right]$ and $\boldsymbol{S}_{21} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right]$. For all models, the quantity $\boldsymbol{S}_{21}$ is zero, as demonstrated below.

- For Beta-SPAM-GEE and BR-SPAM-GEE:

$$\boldsymbol{S}_{21} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right] = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} (\boldsymbol{B}_i^{*\top} \boldsymbol{T}_{1i} \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y} - \boldsymbol{\mu}))\right] = \boldsymbol{0},$$

  because none of this components is function of $\boldsymbol{\phi}$ and as consequence of $\boldsymbol{\kappa}$, thus the derivative is zero;

- For Simplex-SPAM-GEE:

$$\begin{aligned} \boldsymbol{S}_{21} &= \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right] = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{E}_i^\top \boldsymbol{T}_{2i} \boldsymbol{\Upsilon}_i^{-1} \left(-\frac{1}{2\boldsymbol{\phi}_i} d'(\boldsymbol{y}_i, \boldsymbol{\mu}_i)\right)\right] \\ &= \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{E}_i^\top \boldsymbol{T}_{2i} \boldsymbol{\Upsilon}_i^{-1} \frac{1}{2\boldsymbol{\phi}^2} d'(\boldsymbol{y}_i, \boldsymbol{\mu}_i)\right] = \boldsymbol{0}; \end{aligned}$$

  when the expectation function is applied, all the matrix become zero because $\mathbb{E}(d'(\boldsymbol{y}_i, \boldsymbol{\mu}_i)) = \boldsymbol{0}$.

Thus, the sensibility matrix can be rewritten for all models as

$$\boldsymbol{S}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{0} & \boldsymbol{S}_{22} \end{pmatrix}, \tag{4.15}$$

and according with Jørgensen and Knudsen (2004), this configuration of the sensibility matrix implies in the insensitivity property, which allows us to separate $\boldsymbol{\nu}$ and $\boldsymbol{\kappa}$ onto two equations to be updated in each step of estimation algorithm, following Bonat et al. (2018) proposal. The other entries of this matrix are defined as $\boldsymbol{S}_{11} = -\boldsymbol{B}_i^{*\top}\boldsymbol{T}_{1i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{T}_{1i}\boldsymbol{B}_i^* - \boldsymbol{\Lambda}^*$, where $\boldsymbol{\Lambda}^* = \text{blockdiag}(\boldsymbol{0}_{pp}, \lambda_1\boldsymbol{\Lambda}_1^d, \ldots, \lambda_k\boldsymbol{\Lambda}_k^d)^\top$ and $\boldsymbol{0}_{pp}$ is a $p \times p$ matrix of zeros. Also, $\boldsymbol{S}_{22} = -\boldsymbol{E}_i^\top\boldsymbol{T}_{2i}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{T}_{2i}\boldsymbol{E}_i$ and $\boldsymbol{S}_{12} = -\boldsymbol{E}_i^\top\boldsymbol{T}_{12i}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{T}_{12i}\boldsymbol{X}_i$, for Beta-SPAM-GEE and BR-SPAM-GEE and $\boldsymbol{S}_{12} = \boldsymbol{0}$ for Simplex-SPAM-GEE.

Now, developing the estimation algorithm by applying a combination of Gauss-Seidel method for vector $\boldsymbol{\nu}$ and reweighed least squares for $\boldsymbol{\kappa}$, the $(u+1)$-th step of the iterative process for obtaining the maximum penalized likelihood estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k$ and $\boldsymbol{\kappa}$ by fixing λ, may be expressed as

$$\begin{aligned}
\boldsymbol{\beta}^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{W}_{1i}^{(u)} \boldsymbol{X}_i\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{W}_{1i}^{(u)} \left(\boldsymbol{z}_{1i}^{(u)} - \sum_{j=1}^k \boldsymbol{B}_{ij}\boldsymbol{\gamma}_j^{(u+1)}\right)\right], \\
\boldsymbol{\gamma}_j^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{B}_{ij}^\top \boldsymbol{W}_{1i}^{(u)} \boldsymbol{B}_{ij}\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{B}_{ij}^\top \boldsymbol{W}_{1i}^{(u)} \left(\boldsymbol{z}_{1i}^{(u)} - \boldsymbol{X}_i\boldsymbol{\beta}^{(u+1)} - \sum_{j^* \neq j} \boldsymbol{B}_{ij^*}\boldsymbol{\gamma}_{j^*}^{(u+1)}\right)\right], \\
\boldsymbol{\kappa}^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{E}_i^\top \boldsymbol{W}_{2i}^{(u)} \boldsymbol{E}_i\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{E}_i^\top \boldsymbol{W}_{2i}^{(u)} \boldsymbol{z}_{2i}^{(u)}\right],
\end{aligned}$$

for $j = 1, \ldots, k$ and $u = 0, 1, 2, \ldots$, where $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\boldsymbol{\gamma}}_j^{(0)}$ and $\hat{\boldsymbol{\kappa}}^{(0)}$ are the initial estimates, $\boldsymbol{z}_{1i} = \boldsymbol{\eta}_{1i} + \boldsymbol{T}_{1i}^{-1}\boldsymbol{u}_i$, $\boldsymbol{z}_{2i} = \boldsymbol{\eta}_{2i} + \boldsymbol{T}_{2i}^{-1}\boldsymbol{q}_i$, $\boldsymbol{W}_{1i} = \boldsymbol{T}_{1i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{T}_{1i}$ and $\boldsymbol{W}_{2i} = \boldsymbol{T}_{2i}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{T}_{2i}$.

For the model BR-SPAM-GEE, there is one extra parameter to be estimated, which is α. Thus, only for this model one more parameter will be updating in each iteration. The estimation process of it will be made through maximum likelihood, using an optimization algorithm. The one chosen for maximization process was the `optim` of the software `R` using the method L-BFGS-B (Byrd et al., 1995). The updating, in the step $(u+1)$ will happen using the current estimates of the other parameters, replacing them in the log-likelihood function, $l(\boldsymbol{\theta})$, defined in Equation (4.11). The optimization happens and the new estimate for α is used to updating the other parameters. It will occurs until the converge is reached.

4.5 Correlation parameters estimation

The results shown here are based on Liang and Zeger (1986). First, considering that $\text{Cov}(\boldsymbol{u}_i)$ can be $\sqrt{n}$-consistently estimated by $\sum_i \boldsymbol{u}_i\boldsymbol{u}_i^\top/n$ (Liang and Zeger, 1986).

Unstructured

In this case, we have $n^*(n^*-1)/2$, where $n^* = \max\{n_i^*\}_{i=1}^n$, parameters to be estimated. Let $\xi_{tt'}$ be the (t,t') element of $\boldsymbol{R}(\boldsymbol{\xi})$, for $t \neq t'$, which may be estimated by

$$\hat{\xi}_{tt'} = \frac{\sum_{i=1}^n u_{it}u_{it'}}{\sqrt{\sum_{i=1}^n u_{it}^2}\sqrt{\sum_{i=1}^n u_{it'}^2}}$$

Exchangeable

Here the diagonal elements of $\boldsymbol{R}(\boldsymbol{\xi})$ are 1 and the others are ξ, i.e., it is assumed that the correlation between any two observations of the same individual is always the same. Thus, ξ can be estimated by

$$\hat{\xi} = \frac{\sum_{i=1}^n \frac{2}{n_i^*(n_i^*-1)} \sum_{t>t'} u_{it}u_{it'}}{\sum_{i=1}^n \frac{1}{n_i^*} \sum_{t=1}^{n_i^*} u_{it}^2}.$$

First-order autoregressive (AR-1)

In this case, the diagonal elements of $\boldsymbol{R}(\boldsymbol{\xi})$ are 1 and for the t-th line and t'-th column are $\xi^{|t-t'|}$, for $t \neq t'$, that is, we admit that the correlation between two instants of time decays exponentially according to the distance of the observations. In this case, ξ can be estimated by

$$\hat{\xi} = \frac{\sum_{i=1}^n \sum_{t=1}^{n_i^*-1} u_{it}u_{i(t+1)}}{\sqrt{\sum_{i=1}^n \sum_{t=1}^{n_i^*-1} u_{it}^2 \sum_{i=1}^n \sum_{t=2}^{n_i^*} u_{it}^2}}.$$

4.6 Effective degrees of freedom

Following Manghi et al. (2019), the effective degrees of freedom will be derived considering the solution for the linear predictor, only related with $\boldsymbol{\nu}$, because it contains the non-parametric component. Hence, the solution for the linear predictor at the convergence of the iterative process, namely $\hat{\boldsymbol{\eta}}_{1i} = \boldsymbol{B}_i^*\hat{\boldsymbol{\nu}} = \hat{\boldsymbol{H}}(\boldsymbol{\lambda})\hat{\boldsymbol{z}}_{1i}$, where $\widehat{\boldsymbol{H}}(\boldsymbol{\lambda}) = \boldsymbol{B}_i^*(\boldsymbol{B}^{*\top}\widehat{\boldsymbol{W}}_{1i}\boldsymbol{B}^* + \boldsymbol{\Lambda})^{-1}\boldsymbol{B}_i^{*\top}\widehat{\boldsymbol{W}}_{1i}$ may be interpreted as a projection matrix or smoother, for $i = 1, \ldots, n$.

Therefore, based on Green and Silverman (1993), the effective degrees of freedom can be defined as $df(\boldsymbol{\lambda}) = \text{trace}\{\hat{\boldsymbol{H}}(\boldsymbol{\lambda})\}$. Notice that $df(\lambda_1)$ corresponds to the sum of the principal diagonal elements of the matrix $(\boldsymbol{B}^{*\top}\boldsymbol{W}_{1i}\boldsymbol{B}^* + \boldsymbol{\Lambda})^{-1}\boldsymbol{B}^{*\top}\boldsymbol{W}_{1i}\boldsymbol{B}^*$ from $(p+1)$-th position to the $(p+s_1)$-th position, and so on for $df(\lambda_j)$, for $j = 1, \ldots, k$.

4.7 Estimation of smoother parameters

Also, following Manghi et al. (2019), for the selection parameter they suggest to consider the generalized cross-validation method (Craven and Wahba, 1978; Wood, 2017), which consists in selecting the $\boldsymbol{\lambda}$, such that

$$\hat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}}\text{GCV}(\boldsymbol{\lambda}) = \text{argmin}_{\boldsymbol{\lambda}} \frac{\sum_{i=1}^{n} \boldsymbol{u}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{u}_i}{[1 - n^{**-1}\text{trace}\{\hat{\boldsymbol{H}}(\boldsymbol{\lambda})\}]^2},$$

where $n^{**} = \sum_{i=1}^{n} n_i^*$.

4.8 Model selection criteria

Among others, we can cite the following information criteria for correlated data models and some extensions: $\text{AIC}(\boldsymbol{\lambda}) = -2Q(\hat{\boldsymbol{\theta}}) + 2(df(\boldsymbol{\lambda}) + s^*)$, $\text{BIC}(\boldsymbol{\lambda}) = -2Q(\hat{\boldsymbol{\theta}}) + \log(n^{**})(df(\boldsymbol{\lambda}) + s^*)$, $\text{HQIC} = -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2[s^* + df(\boldsymbol{\lambda})]\log(\log(n))$, $\text{AICc} = \text{AIC} + (2k^{**}(k^{**}+1))/(n - k^{**} - 1)$, $\text{SABIC} = -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + [s^* + df(\boldsymbol{\lambda})]\log((n+2)/24)$ $\text{GAIC} = -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + k^*[s^* + df(\boldsymbol{\lambda})]$, where $k^* = 1, \ldots, \log(n)$, $k^{**} = [s^* + df(\boldsymbol{\lambda})]$, where $s^* = q + 1$ if the model is BR-SPAM-GEE and $s^* = q$ if the model is Beta-SPAM-GEE or Simplex-SPAM-GEE, and $Q(.,.)$ denotes the respective quasi-likelihood function.

4.9 Obtaining standard errors

To obtain the covariance matrix, $\text{Cov}(\hat{\boldsymbol{\theta}})$, and standard errors, $\text{SE}(\hat{\boldsymbol{\theta}})$, considering the sandwich covariance estimator $\hat{\boldsymbol{J}}^{-1}$ of $\hat{\boldsymbol{\theta}}$, which can be consistently estimated (Liang and Zeger, 1986) by:

$$\hat{\boldsymbol{J}}^{-1} = \left(\sum_{i=1}^{n} \boldsymbol{S}_i(\hat{\boldsymbol{\theta}})\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{Q}_i^\top \widehat{\boldsymbol{T}}_i \widehat{\boldsymbol{P}}_i \hat{\boldsymbol{d}}_i^* \hat{\boldsymbol{d}}_i^{*\top} \widehat{\boldsymbol{P}}_i^{-1} \widehat{\boldsymbol{T}}_i \boldsymbol{Q}_i\right) \left(\sum_{i=1}^{n} \boldsymbol{S}_i(\hat{\boldsymbol{\theta}})\right)^{-1}, \tag{4.16}$$

where $\boldsymbol{S}_i(\boldsymbol{\theta})$ is defined in Equation (4.15). Another option, following Manghi et al. (2017) proposal, is to obtain the so-called naive standard errors using the covariance matrix $\boldsymbol{J}^{*-1}$, given by

$$\hat{\boldsymbol{J}}^{*-1} = \left(\sum_{i=1}^{n} \boldsymbol{S}_i(\hat{\boldsymbol{\theta}})\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{Q}_i^\top \widehat{\boldsymbol{T}}_i \widehat{\boldsymbol{P}}_i \widehat{\boldsymbol{T}}_i \boldsymbol{Q}_i\right) \left(\sum_{i=1}^{n} \boldsymbol{S}_i(\hat{\boldsymbol{\theta}})\right)^{-1}.$$

Thereby, $\text{Cov}(\hat{\boldsymbol{\theta}})$ is equal to $\hat{\boldsymbol{J}}^{-1}$ or $\hat{\boldsymbol{J}}^{*-1}$ and $\text{SE}(\hat{\boldsymbol{\theta}})$, to $\hat{\boldsymbol{J}}^{-1/2}$ or $\hat{\boldsymbol{J}}^{*-1/2}$.

4.10 Hypothesis testing

To test the significance of linear statistical hypothesis $\boldsymbol{C}\varsigma = \boldsymbol{d}$, where $\text{rank}(\boldsymbol{C}) = l$, $l \geq p$ or $l \geq q$, it can be used the Wald-type statistic (see Hardin (2005), for example), that is,

$$\xi_W = (\boldsymbol{C}\hat{\varsigma} - \boldsymbol{d})^\top (\boldsymbol{C}\boldsymbol{V}_\varsigma \boldsymbol{C}^\top)^{-1} (\boldsymbol{C}\hat{\varsigma} - \boldsymbol{d}),$$

where under H_0, $\xi_W \xrightarrow[n\to\infty]{D} \chi_l^2$ and, furthermore, $\boldsymbol{\varsigma}$ can be either $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, $\boldsymbol{V}_\varsigma$ is the matrix of variance-covariance regards to $\boldsymbol{\varsigma}$ extract from $\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}})$.

4.11 Model fit diagnostic tools

As said in the Sections 2.11 and 3.11, performing diagnostics analysis are important in order to check the goodness-of-fit of the estimated model and to evaluate the model assumptions. Thereby, it will be introduced for the SPAM-GEE models graphical tools for detecting departures from the postulated model, influential observations and local influential analysis.

4.11.1 Residual analysis

Some proposals of residuals could be using in this cases to analysis to goodness of the fit, as for example, the the ordinary residuals based on the proposal of Venezuela (2008), and later modified by Freitas et al. (2021) to deal with a full parametric model, in which the mean and dispersion/precision parameter had linear predictors related with each of them. For the graphical analysis of this residuals, QQ plot with envelope can be done by building appropriate a simulated normal plot, via gaussian copulas to simulate the random variables. However, it could entail a computational cost depending on the size of the sample under study.

To circumvent this disadvantage, another option of residuals is the quantile, used by us in the models for independent data. We use in the same development presented in Section 2.11.2, which is, for the i-th subject in the t-th repeated measure, the quantile residuals are defined as $r_{it} = \Phi^{-1}(F(Y_{it}|\hat{\boldsymbol{\theta}}))$.

Some tests were conducted for us to evaluate if this proposal could be used in the case of correlated data. Then, 100 replicas were generated from models and we analyzed the mean, variance, kurtosis and skewness for the 100 residuals sample. In average the quantile residuals presented a mean and a variance close to 0 and 1, respectively, and kurtosis equals to 2.9 and a skeness close to 0. Furthermore, from the histogram of each samples, the normality supposition seems to be reasonable for these residuals, thus pointwise envelopes could be constructed based on the standard normal distribution, as made in the independent models.

Also, for the proposal residuals, usual plots as the index plot of residuals can be done to evaluate goodness of the fit.

4.11.2 Local Influence

As made in the previous Sections 2.11.4 and 3.11.1, it will be developed local influence tools for our approach via GEE. Cadigan and Farrell (2002) proposed the generalized the local influence instead of likelihood displacement, which is given by $\text{FD}(\boldsymbol{\omega}) = 2[\mathcal{F}(\hat{\boldsymbol{\theta}}) - \mathcal{F}(\hat{\boldsymbol{\theta}}_{\omega})]$, where $\mathcal{F}$ is a fit function, assumed doubly differentiable for

$\boldsymbol{\theta}$, whose estimate $\hat{\boldsymbol{\theta}}$ is the solution of $(\partial\mathcal{F}(\boldsymbol{\theta})/\boldsymbol{\theta}) = \mathbf{0}$. The aim is to study the local behaviour of FD($\boldsymbol{\omega}$) for any value of $\boldsymbol{\omega}$ in a neighborhood of $\boldsymbol{\omega}_0$, which represented the null perturbation vector, such that $\mathcal{F}(\hat{\boldsymbol{\theta}}) = \mathcal{F}(\hat{\boldsymbol{\theta}}_{\omega_0}) \Rightarrow \text{FD}(\boldsymbol{\omega}_0) = \mathbf{0}$. Venezuela et al. (2011) proposed under the Cadigan and Farrell (2002) methodology, replace the likelihood equations by the estimating equations and find the eigenvector $\boldsymbol{l}_{max}$ corresponding to the largest eigenvalue of the matrix

$$\hat{\boldsymbol{B}}_G = -\boldsymbol{\Delta}\boldsymbol{S}(\boldsymbol{\theta})^{-1}\boldsymbol{\Delta}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}},$$

where $\boldsymbol{\Delta} = (\partial\boldsymbol{\Psi}_{1p}(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}^\top, \partial\boldsymbol{\Psi}_2(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}^\top)^\top$ and $\boldsymbol{S}(\boldsymbol{\theta})$ is the sensibility matrix. If the interest is in a local influence relative to a partition of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, for example $\boldsymbol{\theta}_1$, the $\boldsymbol{l}_{max}$ corresponding to the largest eigenvalue of the matrix $-\boldsymbol{\Delta}(\boldsymbol{S}^{-1}(\boldsymbol{\theta}) - \boldsymbol{S}(\boldsymbol{\theta})_c^{-1})\boldsymbol{\Delta}$ can be used as a measure for this purpose, where

$$\boldsymbol{S}(\boldsymbol{\theta})_c = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{22} \end{pmatrix}, \quad \text{where} \quad \boldsymbol{S}_{22} = \frac{\partial\boldsymbol{\Psi}_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2^\top}.$$

As Freitas et al. (2021) said, this concept is important when we want to build measures of local influence for $\boldsymbol{\nu}$ and $\boldsymbol{\kappa}$ individually. Three perturbation schemes will be developed here for the models: case-weight perturbation, response perturbation and working correlation matrix perturbation.

Case-weight perturbation

Let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$, where $\boldsymbol{\omega}_i = (\omega_{i1,\ldots,\omega_{in_i^*}})^\top$, $i = 1, \ldots, n$, be a perturbed vector. The GEE's under this perturbation scheme are given by

$$\boldsymbol{\Psi}^*(\boldsymbol{\nu}, \boldsymbol{\kappa}|\boldsymbol{\omega}) = \boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{T}^{-1}\text{diag}\{\boldsymbol{\omega}\}\boldsymbol{d}^* - \boldsymbol{\Lambda}_2.$$

The non perturbation vector, $\boldsymbol{\omega}_0$, assumes $\omega_{it} = 1, i = 1, \ldots, n,\ t = 1, \ldots, n_i^*$. For this scheme, the matrix $\hat{\boldsymbol{B}}_G$ is given by

$$\text{diag}\{\boldsymbol{d}^*\}\boldsymbol{T}^{-1}\boldsymbol{W}^\top\boldsymbol{Q}(\boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{T}^{-1}\text{diag}\{\boldsymbol{d}^*\}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Response perturbation

Consider the additive perturbation scheme in y_{it} given by $y_{\omega it} = y_{it} + \omega_{it}$, where non perturbation vector assumes $\omega_{it} = 0$, i.e., $\boldsymbol{\omega}_0 = \mathbf{0}$. Analyzing the the concatenation of GEE's defined in Equation (4.14), the only quantity that depends on $\boldsymbol{y}$ is the $\boldsymbol{d}^*$. Therefore, considering $\boldsymbol{d}_\omega^*$ the vector $\boldsymbol{d}^*$ perturbed in the response variable. The perturbed estimating equation is given by:

$$\boldsymbol{\Psi}^*(\boldsymbol{\nu}, \boldsymbol{\kappa}|\boldsymbol{\omega}) = \boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{T}^{-1}\text{diag}\{\boldsymbol{\omega}\}\boldsymbol{d}_\omega^* - \boldsymbol{\Lambda}_2,$$

where $\boldsymbol{d}_\omega = (\boldsymbol{d}_{\omega 1}^{*\top}, \ldots, \boldsymbol{d}_{\omega n}^{*\top})^\top$, $\boldsymbol{d}_{\omega_i}^* = (d_{\omega i1}^*, \ldots, d_{\omega i n_i^*}^*)^\top$, $i = 1, \ldots, n$. In this case, the matrix $\boldsymbol{\Delta}$ is given by $\boldsymbol{Q}^\top \boldsymbol{W} \boldsymbol{T}^{-1} \boldsymbol{\mathcal{B}}$, where $\boldsymbol{\mathcal{B}} = \partial \boldsymbol{d}_\omega^* / \partial \boldsymbol{\omega}^\top = \operatorname{diag}\{\partial d_{\omega 11}/\partial \omega_{11}, \ldots, \partial d_{\omega n n_i^*}/\partial \omega_{n n_i^*}\}$. The matrix $\widehat{\boldsymbol{B}}_G$ under response perturbation is given by

$$\boldsymbol{\mathcal{B}} \boldsymbol{T}^{-1} \boldsymbol{W}^\top \boldsymbol{Q} (\boldsymbol{Q}^\top \boldsymbol{W} \boldsymbol{Q})^{-1} \boldsymbol{Q}^\top \boldsymbol{W} \boldsymbol{T}^{-1} \boldsymbol{\mathcal{B}} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \hat{\omega}_0}.$$

For the Beta-SPAM-GEE, we have that

$$\begin{aligned}
\frac{\partial u_{\omega it}}{\partial \omega_{it}} &= 1, \\
\frac{\partial q_{\omega it}}{\partial \omega_{it}} &= \mu_{it} \left[\frac{1}{y_{it} + \omega_{it}} + \frac{1}{1 - y_{it} - \omega_{it}} \right] - \frac{1}{1 - y_{it} - \omega_{it}},
\end{aligned}$$

for the BR-SPAM-GEE,

$$\frac{\partial u_{\omega it}}{\partial \omega_{it}} = 1, \quad \frac{\partial q_{\omega it}}{\partial \omega_{it}} = 2(y_{it} - \mu_{it}),$$

and for the Simplex-SPAM-GEE,

$$\begin{aligned}
\frac{\partial u_{\omega it}}{\partial \omega_{it}} &= \frac{1}{\phi_{it} \mu_{it}^3 (1 - \mu_{it})^3} \left[\frac{3(y_{\omega it} - \mu_{it})^2}{y_{\omega it}(1 - y_{\omega it})} - \frac{(y_{\omega it} - \mu_{it})^3 (1 - 2 y_{\omega it})}{y_{\omega it}^2 (1 - y_{\omega it})^2} + 1 \right], \\
\frac{\partial q_{\omega it}}{\partial \omega_{it}} &= \mu_{it} \left[\frac{1}{y_{it} + \omega_{it}} + \frac{1}{1 - y_{it} - \omega_{it}} \right] - \frac{1}{1 - y_{it} - \omega_{it}},
\end{aligned}$$

therefore $\boldsymbol{\mathcal{B}} = (\partial \boldsymbol{u}_\omega / \partial \boldsymbol{\omega}, \partial \boldsymbol{q}_\omega / \partial \boldsymbol{\omega})^\top$.

Working correlation matrix perturbation

Let $\boldsymbol{R}(\boldsymbol{\xi})$ be a working correlation matrix characterized by a $\binom{n_i^*}{2}$-vector $\boldsymbol{\xi} = \left(\xi_{12}, \ldots, \xi_{\binom{n_i^*}{2}}\right)^\top$. As each unit has its own working correlation matrix, a possible perturbation scheme in the correlation vector $\boldsymbol{\xi}$ can be given by (Venezuela et al., 2011), $\xi_{\omega i(jl)} = \dfrac{\xi_{jl}}{\omega_{i(jl)}}$, where $i = 1, \ldots, n$, $j < l$ and $j, l = 1, \ldots, n_i^*$. For this perturbation scheme, $\boldsymbol{\omega} = (\omega_{1(12)}, \ldots, \omega_{1((n_i^*-1)n_i^*)}, \ldots, \omega_{n(12)}, \ldots, \omega_{n((n_i^*-1)n_i^*)})^\top$ is a perturbation vector and $\boldsymbol{\omega}_0$ is a vector with ones. The perturbation estimating equation is given by:

$$\boldsymbol{\Psi}_1(\boldsymbol{\beta}, \boldsymbol{\omega}) = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{\Sigma}_\omega^{-1} \boldsymbol{u} - \boldsymbol{\Lambda}.$$

Then, each column of matrix $\boldsymbol{\Delta}$ can be expressed by:

$$\frac{\partial \boldsymbol{\Psi}_{1p}(\boldsymbol{\nu}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}_{(jl)}^\top} = \boldsymbol{X}^\top \boldsymbol{T}_1 \frac{\partial \boldsymbol{\Sigma}_\omega^{-1}}{\partial \boldsymbol{\omega}_{(jl)}^\top} \boldsymbol{u},$$

where the derivative of $\boldsymbol{\Sigma}_\omega^{-1}$ with respect to $\boldsymbol{\omega}_{(j,l)}^\top$ is given by $\partial\boldsymbol{\Sigma}_\omega/\partial\boldsymbol{\omega}^\top = \Big(\partial\boldsymbol{\Sigma}_{\omega 1}/\partial\boldsymbol{\omega}_1^\top, \ldots, \partial\boldsymbol{\Sigma}_{\omega n}/\partial\boldsymbol{\omega}_n^\top\Big)$. The i-th diagonal block of $\boldsymbol{\Sigma}_\omega$ is $\boldsymbol{\Sigma}_{\omega i} = \sqrt{\mathrm{Var}(\boldsymbol{u}_i)}\boldsymbol{R}(\boldsymbol{\xi}_{\omega i})\sqrt{\mathrm{Var}(\boldsymbol{u}_i)}$, with $\boldsymbol{\xi}_{\omega i} = \Big(\xi_{\omega i(12)}, \ldots, \xi_{\omega i((n_i^*-1)n_i^*)}\Big)$ and $i = 1, \ldots, n$. We also have that:

$$\frac{\partial\boldsymbol{\Sigma}_{\omega i}}{\partial\omega_{i(jl)}} = \sqrt{\mathrm{Var}(\boldsymbol{u}_i)}\frac{\partial\boldsymbol{R}(\boldsymbol{\xi}_{\omega i})}{\partial\omega_{i(jl)}}\sqrt{\mathrm{Var}(\boldsymbol{u}_i)},$$

where $\partial\boldsymbol{R}(\boldsymbol{\xi}_{\omega i})/\partial\omega_{i(jl)}$ is a $(n_i^* \times n_i^*)$ symmetric matrix with null diagonal and jl and lj elements are equal to $-\xi_{jl}$, with $i = 1, \ldots, n$ and $j, l = 1, \ldots, n_i^*$.

4.12 Simulation Studies

In this section, the results of four simulation studies. Section 4.12.1 is related to the analysis of parameter recovery. In Section 4.12.2, we. In Section 4.12.3 is present a link function misspecification study. The results for the first study is in Appendix, whereas the other studies will be available at `https://github.com/aureaflg/Simulations-study.git`, where we made reproducible codes for one can view the graphs and tables. The correlated data has been generate for each scenario is using t-copulas based on Student-t distribution considering a degree of freedom equal to 3.

4.12.1 Study 1: Parameters recovery

Here we considered some scenarios of interest defined by the combination of the levels of some factors of interest. The factors (with the respective levels within parenthesis) are number of subjects and their repetition number (10 subjects and 5 times, 10 subjects and 10 times, 50 subjects and 10 times) which implies in a sample (n) (50,100,500), regression model (Beta, Simplex, BR), link function (probit, logit, cauchit, cloglog, loglog), modeled parameters (mean and dispersion) and considering the correlation matrix as exchangeable. For each scenario, $R = 100$ replicates of Monte Carlo were generated. Each scenario are described below.

Furthermore, the considered linear predictors and link functions for related scenarios presented in Table 16, are given by $g(\mu_{it}) = \beta_1 X_{it} + cos(Z_{it})$, $\log(\phi_{it}) = \kappa_0 + \kappa_1 E_{it}$, for $t = 1, \ldots, n_i^*$ $i = 1, \ldots, n^{**}$, where $X_{it} \overset{iid}{\sim} \text{uniform}(0,1)$, $E_{it} \overset{iid}{\sim} \text{uniform}(0,1)$ and $Z_{it} \overset{iid}{\sim} \text{uniform}(0,3\pi)$, $F_{it} \overset{iid}{\sim} \text{uniform}(0,1)$, $M_{it} \overset{iid}{\sim} \text{uniform}(0,1)$, is either the probit, logit, cauchit, cloglog. The actual parameter values considered loglog link were $\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$, whereas for the other link were: $\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$. This change to the loglog link function is justified by the fact that, depending on the parameters chosen, generating from the models using this link function can generate values very close to one

Table 16 – Scenarios of simulation study 1

Regression model (modelling mean and dispersion)	n	Link functions
betabeta(μ_{it}, ϕ_{it})	50 100 500	probit, logit, cauchit, cloglog and loglog.
simplex(μ_{it}, ϕ_{it})	50 100 500	probit, logit, cauchit, cloglog and loglog.
BR($\mu_{it}, \phi_{it}, \alpha$)	50 100 500	probit, logit, cauchit, cloglog and loglog.

or zero, which could compromise the fit of the models to the simulated data even for the BR-SPAM-GEE model.

For all models, when $n = 50$, 40 knots were used and the smoothing parameter was fixed at $\lambda = 500$, whereas, when $n = 100$ or $n = 500$, 50 knots were used and the smoothing parameter was fixed at $\lambda = 500$. These values lead to a smooth fitted curve for non-parametric component. As the correlation matrix considered was the exchangeable, the only correlation parameter ξ was fixed as 0.8.

The related results are presented in Appendix B.3, where Figures 64 to 68 are related to Beta-SPAM-GEE, Figures 69 to 73 are related to Simplex-SPAM, and, Figures 74 to 78 concern to BR-SPAM.

In a general way, for all models the regression parameters were properly recovered, as the sample size increases, the estimates become more accurate. The estimated curves are for all sample sizes representing the behavior of the actual curve, in general well. Specifically for the BR-SPAM-GEE the parameter α under probit the recovery was not good enough. In these cases, the sample size increases as well as the bias of estimates under probit and under logit, the estimates are better for sample size 50 than sample size 100. Furthermore, for all models the correlation parameter ξ was under estimated. A possible reason for this is happening is that the copulas used to generate a dependency structure, which is an approximation of the true correlation matrix, can generate bias in the adjustment.

4.12.2 Study 2: Correlation matrix misspecification

In this section we considered only the scenarios varying the correlation matrix (exchangeable,AR-1,unstructured), the regression models (Beta,Simplex,BR) and the parameter related with correlation ξ (0.8,0.3). For generating model, we setting the sample size in 100 (10 subjects and 10 repeated measure), choose the link function as cloglog with R=1 replica, we setting the exchangeable as the structure of correlation matrix and one value for ξ (0.8 and 0.3) and then we fit the simulated data using the others correlation structures. The same model defined in study 1 (see Section 4.12.1) were considered here, except by the parameter ξ, which is either 0.8 and 0.3 and $g_1(.)$ is setting is cloglog.

Although the estimates were, in general, properly recovered under the three correlation structure, it was possible to notice that fitted parameters with BR-SAM-GEE under the AR-1 structure were closer to the actual value of parameters. In other cases, the AR-1 and unstructured had the same or slightly worse performance of exchangeable. It can be concluded that the recovery is reasonable even under misspecification.

4.12.3 Study 3: Link function misspecification

Only the scenarios varying the regression models (beta, simplex, BR) and the link functions (probit, logit, cauchit, cloglog,loglog) are considered. We set the sample size in 500 (50 subjects and 10 repeated measure), modeling the mean and the dispersion. We generate only one replica from each model, setting a distribution and a link function for it, and fit the simulated data using a model with the same distribution but varying the link function, using all the different ones from the one used to generate the simulated data. In Table 17 we present the 60 scenarios of interest.

Table 17 – Scenarios of simulation study 2

Distribution that will generate and fit the simulated data	Actual link function	Link function considered in fitted model
beta(μ_{it}, ϕ_{it}), simplex(μ_{it}, ϕ_{it}) or BR$(\mu_{it}, \phi_{it}, \alpha)$	logit	probit, cauchit, cloglog and loglog
	probit	logit, cauchit, cloglog and loglog
	cauchit	logit, probit, cloglog and loglog
	cloglog	logit, probit, cauchit and loglog
	loglog	logit, probit, cloglog and cauchit

The considered linear predictors and link functions for related scenarios presented in Table 17, are given by $g(\mu_{it}) = \beta_1 X_{it} + cos(Z_{it})$, $\log(\phi_{it}) = \kappa_0 + \kappa_1 E_{it}$, for $i = 1, \ldots, n$, $t = 1, \ldots, n_i^*$, where $X_{it} \overset{iid}{\sim} \text{uniform}(0, 1)$, $E_{it} \overset{iid}{\sim} \text{uniform}(0, 1)$ and $Z_{it} \overset{iid}{\sim} \text{uniform}(0, 3\pi)$, is either the probit, logit, cauchit, cloglog. The actual parameter values considered loglog link were $\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$, whereas for the other cases were $\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$. This change for actual values of parameters under loglog link is justified for the same reason discussed in Section 4.12.1. We considered both the comparison of the observed and predicted values as well as the parameter recovery. The correlation structure used were exchangeable, with $\xi = 0.8$.

The Beta-SPAM-GEE and Simplex-SPAM-GEE presented better recovery of real mean when the link function that generating was symmetric and the link of fitted model was also symmetric. In general the parameters, in these cases, were properly recovered, even when the fitting model do not match with the one the fitted model. For the BR-SPAM-GEE, in general, the well-recovery depends on if the generating and fitting link function are both symmetric, when one of them are not, the parameter is bad recovery.

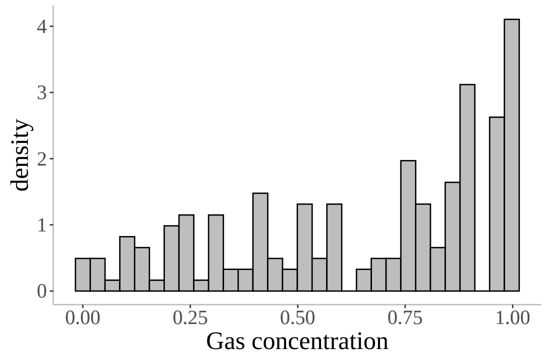
4.13 Real data analysis

The data set that will be analyzed in this section is related to an ophthalmology study on the use of intraocular gas in retinal repair surgeries Meyers et al. (1992). The response is the percent of gas left in the eye recorded as proportion (a percent), the covariates are the Gas level concentration with levels equal to (-1, 0 and 1), corresponding to the concentration levels 15%, 20% and 25%, respectively and the Time (days) after the gas injection. This data was previously analyzed by Song and Tan (2000) and X.-K. Song et al. (2004) considering the Simplex regression model under constant and varying dispersion, respectively. Their models suggest the use of the logarithm of time and the squared logarithm of time terms.

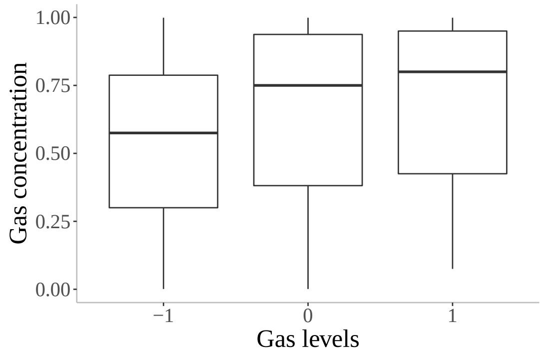
The number of subjects was 31, and they were collected in different amounts of time. The subject 31, were the only one collected 15 times, the others were collected at most 11 times, because it is an unbalanced study. Then, to be possible estimate the correlation, the last four observations of this subject were removed from fitting. Furthermore, Figure 23c indicates a non-linear relation between Gas concentration and Logarithm of time. Also, Figure 23b indicates that the response variability may be related Gas levels covariate as well. Figure 23f indicates that the correlation structure decay over lags.

Under our approach, the model Simplex-SPAM-GEE did not converge using all available link functions and correlation matrix. One reason could be the vector scores (u_i and q_i) chosen here, which were different from those chosen by X.-K. Song et al. (2004). Another difference from their approach is a generalized estimating equation defined by them to estimate the correlations parameters. Their approach available in the `R` package, `simplexreg`, only works for just one combination of link function and correlation matrix, also using exactly the same set of covariates, which may indicate a high sensitivity of models based on Simplex for some cases.

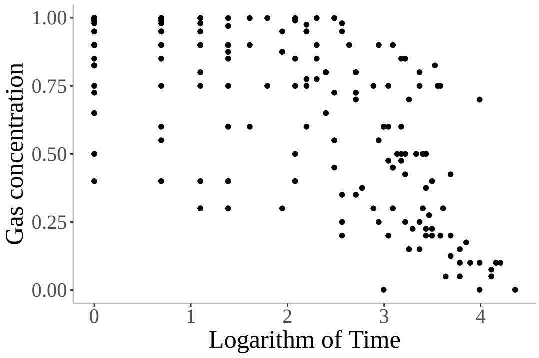
Therefore, for the fit under our approach will considered only Beta-SPAM-GEE and BR-SPAM-GEE using the combination of the link function for the mean and the dispersion parameter and the correlation matrix that made the models converged, which are presented in Table 18. For choose the correlation matrix, the decay saw in the explanatory analysis indicates the AR-1 as the best choice, but we decided also compare with the Exchangeable structure.



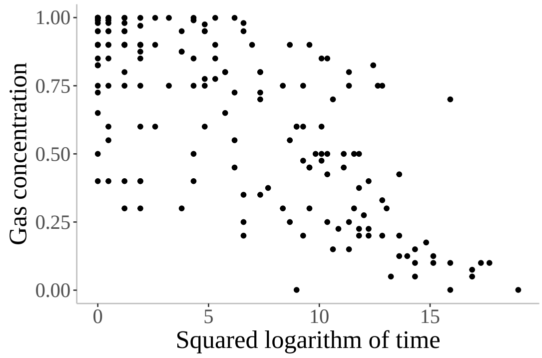
(a) Histogram of Gas concentration.



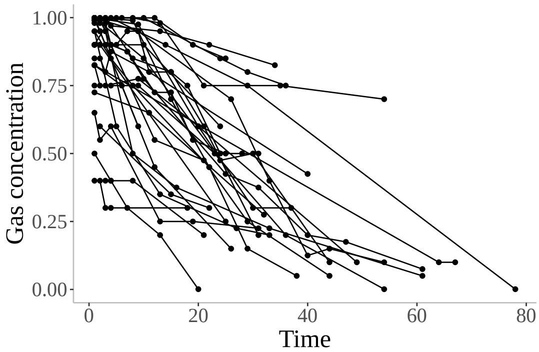
(b) Boxplots of Gas concentration by categories of Levels concentration.



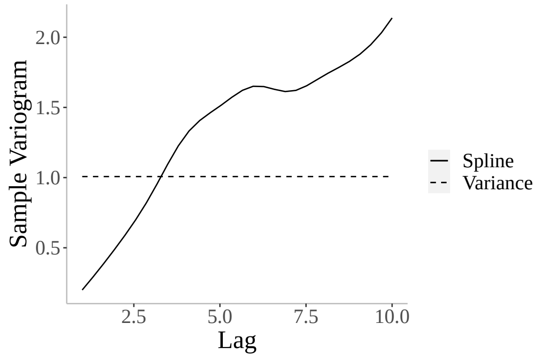
(c) Scatter plot of Gas concentration versus Logarithm of time.



(d) Scatter plot of Gas concentration versus Squared logarithm of time.



(e) Individual profiles.



(f) Variogram.

Figure 23 – Explanatory analysis plots of risk perception data set.

Table 18 – Structure for each model.

Models	Link function for μ	Link function for ϕ	Correlation matrix
Beta-SPAM-GEE	cloglog, cauchit, probit,logit	sqrt	Exchangeable and AR-1
	loglog	log	
BR-SPAM-GEE	cloglog, loglog, cauchit,probit,logit	sqrt	Exchangeable and AR-1

Initially, we fitted a general model for all combination in the Table 18:

$$
\begin{aligned}
Y_{itl} &\sim \text{beta}(\mu_{itl}, \phi_{itl}) \text{ or BR}(\mu_{itl}, \phi_{itl}, \alpha),\\
g_1(\mu_{itl}) &= \beta_0 + (\beta_1)_l + h_1(z^*_{ti}) + h_2(z^{**}_{ti}),\\
g_2(\phi_{itl}) &= \kappa_0 + (\kappa_1)_l + \kappa_2(z^*_{it}),\ i = 1, \dots, 31,\ t = 1, \dots, n^*_i,
\end{aligned}
$$

where, $l = -1, 0, 1$, and $(\beta_1)_{-1} = (\kappa_1)_{-1} = 0$. The parameters (β_1) and (κ_1) are related with Gas levels and κ_2 are associated with the variable $\boldsymbol{z}^*$, which is the logarithm of Time. Also, $\boldsymbol{z}^{**}$ is the Squared logarithm of time. Here the logarithm and the squared logarithm function is being used following the proposal of Song and Tan (2000) and X.-K. Song et al. (2004).

And then, for each model, we try some set of covariates for the fit. However, the addition of some covariates made the estimation process does not converge. Therefore, we considered only the covariates significant to 5% level and which that made possible the estimation process convergence. Therefore, for Beta-SPAM-GEE, the final model was (for all link functions for the mean and the dispersion parameter the selected covariates were the same):

$$
\begin{aligned}
Y_{itl} &\sim \text{beta}(\mu_{itl}, \phi_{itl}),\\
g_1(\mu_{it}) &= \beta_0 + h_1(z^*_i),\\
g_2(\phi_{itl}) &= \kappa_0 + (\kappa_1)_l,\ i = 1, \dots, 31,\ t = 1, \dots, n^*_i,
\end{aligned} \tag{4.17}
$$

$l = -1, 0, 1$, and $(\kappa_1)_{-1} = 0$. Then, the covariate Gas levels remains in ϕ predictor in the levels 20% and 25% concentration (levels 0 and 1) and other (level -1).

And, for BR-SPAM-GEE, the final model was (for all link functions for the mean and the dispersion parameter the selected covariates were the same):

$$
\begin{aligned}
Y_{itl} &\sim \text{BR}(\mu_{itl}, \phi_{itl}, \alpha),\\
g_1(\mu_{it}) &= \beta_0 + h_2(z^{**}_i),\\
g_2(\phi_{itl}) &= \kappa_0 + (\kappa_1)_l,\ i = 1, \dots, 31,\ t = 1, \dots, n^*_i
\end{aligned}
$$

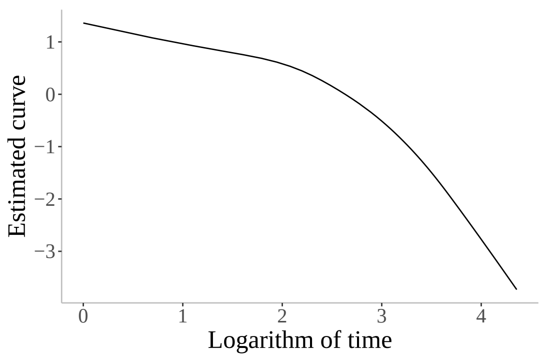
$l = -1, 0, 1$, and $(\kappa_1)_{-1} = (\kappa_1)_0 = 0$. The covariate Gas levels remains in ϕ predictor only in the levels 25% concentration (level=1).

Then, based on the information criteria (IC), we chosen the final model, which is the one selected as the best model among those that were proposed in this work for this data. All IC, except SABIC, see Tables 20, indicate, the Exchangeable as the best correlation structure for this data and the Beta-SPAM-GEE under logit (mean) and sqrt (dispersion) link functions as the better choice. Therefore, the final model is the one presented in Equation 4.17 where $g_1(.)$ is logit, $g_2(.)$ is the sqrt and the correlation structure is Exchangeable.

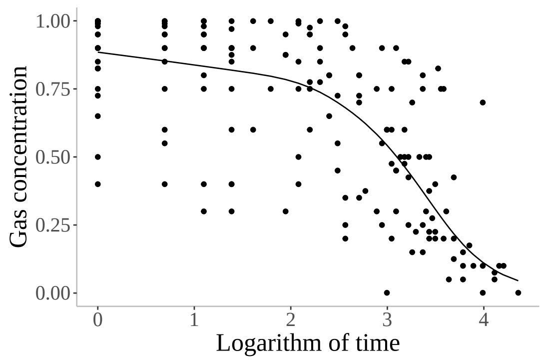
Now, we will check the goodness of fit of the selected model at the global and local analysis. Figure 24b shows the fitted curve for non-parametric part and Figure 24a, the adjusted mean for model Beta-SPAM-GEE. The effective degrees of freedom were 4.20 using 10 knots and a smooth parameter $\lambda = 10$, which was chosen by minimizing the AIC. Some adjustment results can be seen in Figure 25. All the residuals are well within the the confidence bands of the QQ plot (Figure 25a), even though there is some no random pattern. The other residuals plots show no tendency (Figure 25b) as well the normality seems to be reasonable according to the histogram (Figure 25d).

Table 19 – Inferential results for the Beta-SPAM-GEE.

Parameters	Estimate	SE	p-value	$IC_{95\%}$
β_0	0.67	0.31	0.0294	[0.07;1.28]
κ_0	1.69	0.15	<0.0001	[1.40;1.97]
$(\kappa_1)_0$	-3.26	0.16	<0.0001	[-3.58;-2.95]
$(\kappa_1)_1$	-3.66	0.19	<0.0001	[-4.03;-3.29]
Correlation parameter ξ	0.60	-	-	-



(a) Estimated non-parametric curve.



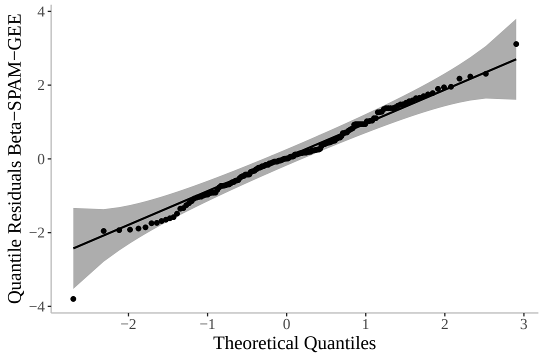
(b) Observed and predicted response mean as a function of Logarithm of time.

Figure 24 – Estimated curves for Beta-SPAM-GEE.

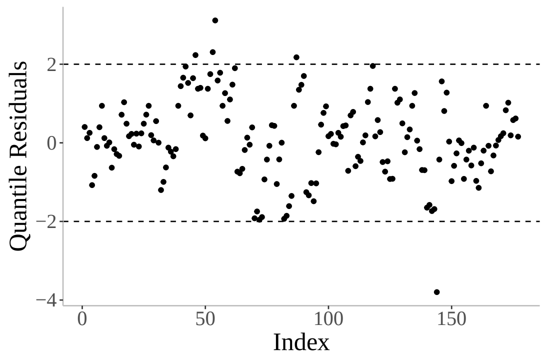
Moreover, Figure 26 presents the the local influence analysis under case-weight perturbation related to $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$ (Figures 26a and 26b, respectively), response perturbation scheme related to $\boldsymbol{\mu}$ and $\boldsymbol{\phi}$ (Figures 26c and 26d, respectively) and the correlation perturbation scheme (Figure 26e).

From Figure 26, we see that the observations (with the respective scheme and parameter) #72,#73,#82,#83 (Case-weight for $\boldsymbol{\mu}$),#144 (Case-weight for $\boldsymbol{\phi}$), #149 (Response for $\boldsymbol{\mu}$) and #144 (Response for $\boldsymbol{\phi}$) were flagged, whereas for the for correlation perturbation scheme no observations were flagged.

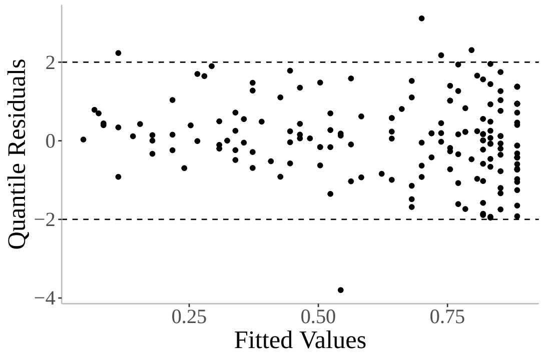
After removing each flagged observation one at a time, the significance of the parameters are not modified. Also, the behavior of the residuals and QQ plot remained



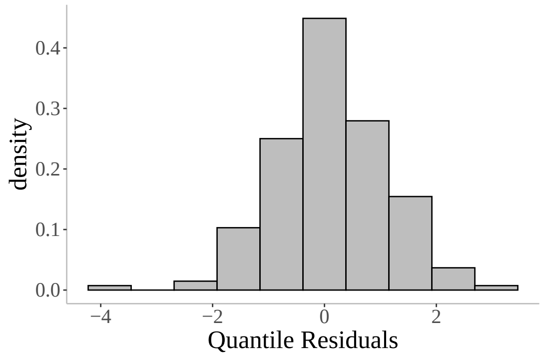
(a) Quantile-quantile plot with 99% envelopes.



(b) Index plot of quantile residuals.



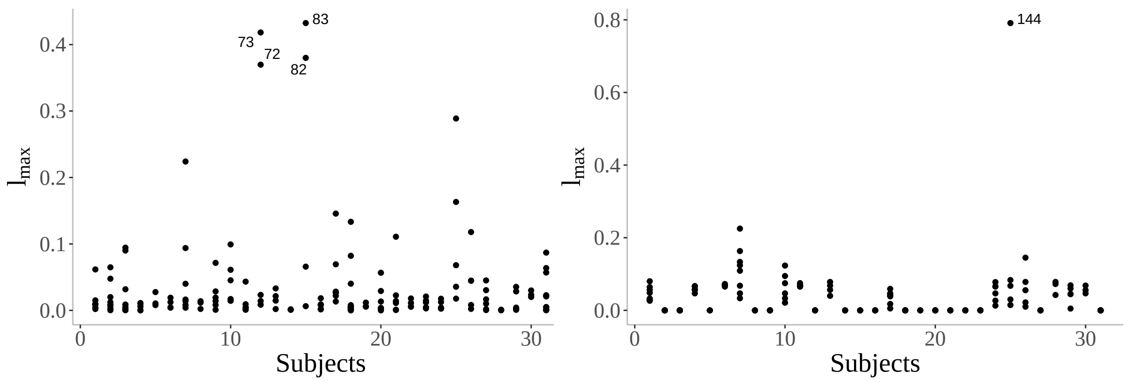
(c) Fitted values versus quantile residuals.



(d) Histogram of quantile residuals.

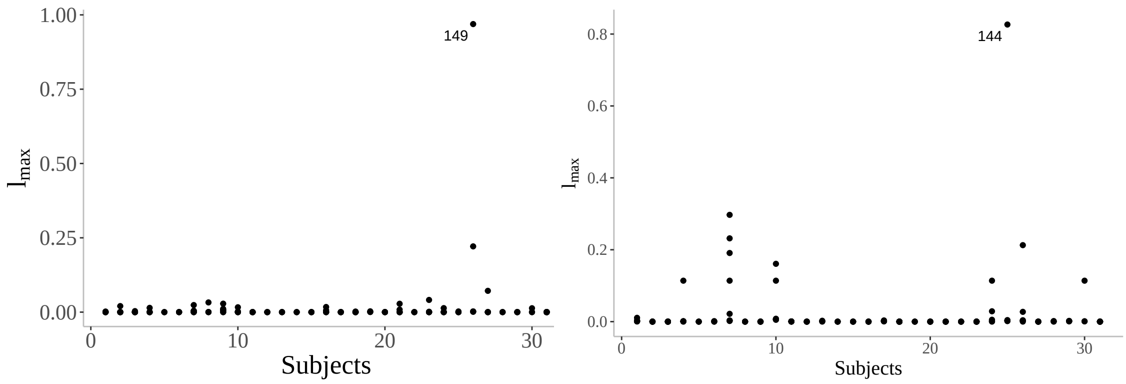
Figure 25 – Results of fitted Beta-SPAM-GEE.

the same. When #72,#73,#82,#83 were removed, some new observations were flagged in the local influence analysis under the case-weight perturbation, which may indicate that these observations could be covering up the behaviour of these others. In conclusion, once that the exclusion of the flagged observations did not lead to significant changes on the overall model fit, no further changes are necessary. From the final model, it possible to notice that, Gas levels variable is significant factor attributed to the heterogeneous dispersion. Also, the estimated non-parametric curve indicates that the Gas concentration slowly until the logarithm of the time 2, and, after this point, its decay rate increases.



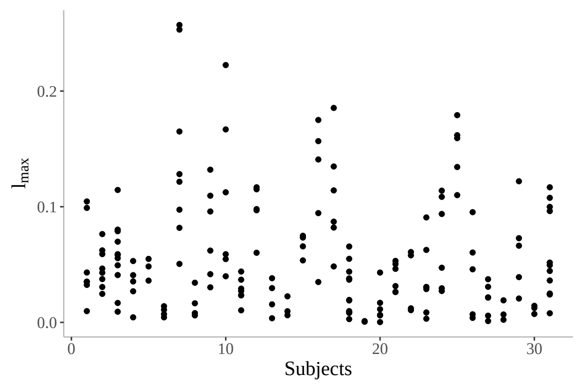
(a) Case-weight perturbation scheme related to $\boldsymbol{\mu}$.

(b) Case-weight perturbation scheme related to $\boldsymbol{\phi}$.



(c) Response perturbation scheme related to $\boldsymbol{\mu}$.

(d) Response perturbation scheme related to $\boldsymbol{\phi}$.



(e) Correlation perturbation scheme.

Figure 26 – Index plot of $\boldsymbol{l}_{max}$ under perturbation schemes for Beta-SPAM-GEE.

Table 20 – Information Criteria for the two models.

		Exchangeable					AR-1				
Model	Link functions μ and ϕ	AIC	AICC	SABIC	HQIC	BIC	AIC	AICC	SABIC	HQIC	BIC
Beta-SPAM-GEE	cloglog and sqrt	192.04	193.05	178.27	201.98	216.55	192.15	193.41	176.51	203.44	220.00
	loglog and log	**190.56**	191.51	**177.30**	200.14	214.17	**190.10**	**191.27**	**175.11**	200.92	216.78
	probit and sqrt	191.59	192.63	177.62	201.68	216.47	191.32	192.62	175.48	202.77	219.54
	logit and sqrt	190.67	**191.50**	178.42	**199.51**	**212.48**	190.47	191.50	176.55	**200.52**	**215.24**
BR-SPAM-GEE	cauchit and sqrt	327.74	328.67	314.68	337.17	350.99	366.23	367.47	350.72	377.43	393.84
	logit and sqrt	**212.55**	**213.64**	**198.16**	**222.94**	**238.16**	**220.49**	**221.95**	**203.60**	**232.69**	**250.57**
	probit and sqrt	217.49	218.93	200.68	229.62	247.41	226.69	228.69	206.50	241.27	262.64
	cloglog and sqrt	227.94	229.28	211.80	239.60	256.69	239.66	241.52	220.25	253.67	274.21
	loglog and sqrt	252.78	254.10	236.75	264.35	281.31	270.47	272.27	251.47	284.19	304.30

5 Semi-parametric additive models for dependent and limited augmented data

In this Chapter we present extensions of the models developed in Chapter 4, allowing the response belongs to the [0,1] interval (as well as the respective particular cases). The regression structures and probability distributions are those presented before. Therefore, the model class developed in the previous Chapter is a particular case of this developed in this Chapter. The proposal models will be the Semi-parametric Additive ZOAB (ZOAB-SPAM-GEE), the Semi-parametric Additive ZOAS (ZOAS-SPAM-GEE), and the Semi-parametric Additive ZOABR (ZOABR-SPAM-GEE) models via GEE.

5.1 Models

Let $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \ldots, \boldsymbol{Y}_n^\top)^\top$, with $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i^*})^\top$, $i = 1, \ldots, n$ be the response vector, and considering the marginal distribution of Y_{it} as $Y_{it} \sim \text{ZOAB}(\mu_{it}, \phi_{it}, p_{0it}, p_{1it})$, $\text{ZOAS}(\mu_{it}, \phi_{it}, p_{0it}, p_{1it})$ or $\text{ZOABR}(\mu_{it}, \phi_{it}, p_{0it}, p_{1it}, \alpha)$, for $t = 1, \ldots, n_i^*$, $i = 1, \ldots, n$. the Semi-parametric Additive Model via GEE (SPAM-GEE) for any of the three distributions, ZOAB-SPAM-GEE, ZOAS-SPAM-GEE and ZOABR-SPAM-GEE, have the following systematic component:

$$
\begin{aligned}
\boldsymbol{\eta}_1 &= g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} h_j(\boldsymbol{z}_j), \\
\boldsymbol{\eta}_2 &= g_2(\boldsymbol{\phi}) = \boldsymbol{E}\boldsymbol{\kappa}, \\
(\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1) &= H(\boldsymbol{p}_0, \boldsymbol{p}_1) = (h_0(\boldsymbol{p}_0, \boldsymbol{p}_1), h_1(\boldsymbol{p}_0, \boldsymbol{p}_1))
\end{aligned} \tag{5.1}
$$

where $n^{**} = \sum_{i=1}^{n} n_i^*$ is the sample size, $\boldsymbol{\mu}_i = \mathbb{E}(\boldsymbol{Y}_i | \boldsymbol{Y}_i \in (0,1))$, $\boldsymbol{p}_{0i} = \mathbb{P}(\boldsymbol{Y}_i = \boldsymbol{0})$, $\boldsymbol{p}_{1i} = \mathbb{P}(\boldsymbol{Y}_i = \boldsymbol{1})$ and $1 - \boldsymbol{p}_{0i} - \boldsymbol{p}_{1i} = \mathbb{P}(\boldsymbol{Y}_i \in (0,1))$. The functions $\boldsymbol{\eta}_{1i} = \boldsymbol{X}_i^\top \boldsymbol{\beta}$, $\boldsymbol{\eta}_{2i} = \boldsymbol{E}_i^\top \boldsymbol{\kappa}$, $\boldsymbol{\zeta}_{0i} = \boldsymbol{F}_i^\top \boldsymbol{\rho}$, $\boldsymbol{\zeta}_{1i} = \boldsymbol{M}_i^\top \boldsymbol{\tau}$ are linear predictors, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_s)^\top \in \mathbb{R}^s$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{s_0})^\top \in \mathbb{R}^{s_0}$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{s_1})^\top \in \mathbb{R}^{s_1}$ are vectors of unknown parameters.

Also, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top)^\top$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i^*})^\top$, $\boldsymbol{E} = (\boldsymbol{E}_1^\top, \ldots, \boldsymbol{E}_n^\top)^\top$, $\boldsymbol{E}_i = (\boldsymbol{e}_{i1}, \ldots, \boldsymbol{e}_{in_i^*})^\top$, $\boldsymbol{F} = (\boldsymbol{F}_1^\top, \ldots, \boldsymbol{F}_n^\top)^\top$, $\boldsymbol{F}_i = (\boldsymbol{f}_{i1}, \ldots, \boldsymbol{f}_{in_i^*})^\top$, $\boldsymbol{M} = (\boldsymbol{M}_1^\top, \ldots, \boldsymbol{M}_n^\top)^\top$, $\boldsymbol{M}_i = (\boldsymbol{m}_{i1}, \ldots, \boldsymbol{m}_{in_i^*})^\top$, are design matrices of dimension, respectively, $n^{**} \times p$, $n^{**} \times s$, $n^{**} \times s_0$ and $n^{**} \times s_1$, in which $\boldsymbol{x}_i = (x_{it1}, \ldots, x_{itp})^\top$, $\boldsymbol{e}_i = (e_{it1}, \ldots, e_{its})^\top$, $\boldsymbol{f}_i = (f_{it1}, \ldots, f_{its_0})^\top$, $\boldsymbol{m}_i = (m_{it1}, \ldots, m_{its_1})^\top$, for $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$, $\boldsymbol{z}_j = (\boldsymbol{z}_{1j}^\top, \ldots, \boldsymbol{z}_{nj}^\top)^\top, j = 1, \ldots, k$, $\boldsymbol{z}_{ij} = (z_{i1j}, \ldots, z_{in_i^*j})$ is a specification vector associated with non-parametric components and $h_1, \ldots, h_k$ are an unknown smooth functions.

In addition, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_n^\top)^\top$, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i^*})^\top$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^\top, \ldots, \boldsymbol{\phi}_n^\top)^\top$, $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{in_i^*})^\top$, $\boldsymbol{p}_0 = (\boldsymbol{p}_{01}^\top, \ldots, \boldsymbol{p}_{0n}^\top)^\top$, $\boldsymbol{p}_{0i} = (p_{0i1}, \ldots, p_{0in_i^*})^\top$, $\boldsymbol{p}_1 = (\boldsymbol{p}_{11}^\top, \ldots, \boldsymbol{p}_{1n}^\top)^\top$, $\boldsymbol{p}_{1i} = (p_{1i1}, \ldots, p_{1in_i^*})^\top$, $g_1(.)$ and $g_2(.)$ are strictly monotonic and twice differentiable link functions, such that, $g_1 : (0,1) \to \mathbb{R}$ and $g_2 : \mathbb{R}^+ \to \mathbb{R}$, while H is admitted being a bijective transformation of the set $\mathbb{C} = \{(p_{0i}, p_{1i}) : 0 < p_{0i} < 1, 0 < p_{1i} < 1 - p_{0i}\}$ to $\mathbb{R}^2$, twice differentiable. As discussed in Section 3.1, it can define H as,

$$H(\boldsymbol{p}_{0i}, \boldsymbol{p}_{1i}) = (h_0(\boldsymbol{p}_{0i}, \boldsymbol{p}_{1i}), h_1(\boldsymbol{p}_{0i}, \boldsymbol{p}_{1i})) = \left(h\left(\frac{\boldsymbol{p}_{0i}}{1 - \boldsymbol{p}_{0i} - \boldsymbol{p}_{1i}}\right), h\left(\frac{\boldsymbol{p}_{1i}}{1 - \boldsymbol{p}_{0i} - \boldsymbol{p}_{1i}}\right)\right),$$

and following Ospina (2008), the h chosen was the logarithm function, it results that

$$\frac{\boldsymbol{p}_{0i}}{1 - \boldsymbol{p}_{0i} - \boldsymbol{p}_{1i}} = \exp(\boldsymbol{\zeta}_{0i}), \quad \frac{\boldsymbol{p}_{1i}}{1 - \boldsymbol{p}_{0i} - \boldsymbol{p}_{1i}} = \exp(\boldsymbol{\zeta}_{1i}),$$

and,

$$\begin{gathered} \boldsymbol{p}_{0i} = \frac{e^{\boldsymbol{\zeta}_{0i}}}{1 + e^{\boldsymbol{\zeta}_{0i}} + e^{\boldsymbol{\zeta}_{1i}}}, \quad \boldsymbol{p}_{1i} = \frac{e^{\boldsymbol{\zeta}_{1i}}}{1 + e^{\boldsymbol{\zeta}_{0i}} + e^{\boldsymbol{\zeta}_{1i}}}, \\ 1 - \boldsymbol{p}_{0i} - \boldsymbol{p}_{1i} = \frac{1}{1 + e^{\boldsymbol{\zeta}_{0i}} + e^{\boldsymbol{\zeta}_{1i}}}. \end{gathered} \tag{5.2}$$

The $h_j(\boldsymbol{z}_j)$ can be rewritten as discussed before in the Section 1.2.1 as

$$h_j(\boldsymbol{z}_j) = \sum_{l=1}^{q_j} \gamma_{lj} b_{lj}(\boldsymbol{z}_j), \tag{5.3}$$

with γ_{lj} being the coefficients to be estimated, b_{lj}'s being l-th cubic B-spline that compounds the j-th P-spline at the vector $\boldsymbol{z}_j$, $j = 1, \ldots, k$ and $l = 1, \ldots, q_j$. As cubic B-splines are being adopted here, $b_{lj} = B_{l,d_l}$ with $d_l = 4$ as defined in Section 1.2.1, q_j is the number of cubic B-splines that compounds the j-th P-spline, $k_j = d_l + q_j$ is the number of knots. Thus, it is possible to rewrite Equation (4.6) as follow for the i-th observation, $i = 1, \ldots, n$

$$\boldsymbol{\eta}_{1i} = g_1(\boldsymbol{\mu}_i) = \boldsymbol{X}_i^\top \boldsymbol{\beta} + \boldsymbol{b}_{i1}^\top \gamma_1 + \cdots + \boldsymbol{b}_{ik}^\top \gamma_k,$$

where $\boldsymbol{b}_{ij} = (b_{1j}(\boldsymbol{z}_{ij}), \ldots, b_{q_j j}(\boldsymbol{z}_{ij}))^\top$ and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{jq_j})^\top$, for $j = 1, \ldots, k$. Return to matrix form of $\boldsymbol{\eta}_1$, this predictor of the model can be rewritten as

$$\boldsymbol{\eta}_1 = g_1(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{B}_j \boldsymbol{\gamma}_j, \tag{5.4}$$

where $\boldsymbol{B}_j$ is a matrix of dimension $n^{**} \times q_j$ composed by $b_{lj}(\boldsymbol{z}_{ij})$, for $i = 1, \ldots, n$, $j = 1, \ldots, k$ and $l = 1, \ldots, q_j$.

5.2 Penalized GEE

Based on the proposal of Liang and Zeger (1986), described in Section 4.1, it will be developed four distinct GEE's for the mean dispersion/precision, probability of occurrence of zero and one parameters, based on Section 4.3, for ZOAS-SPAM-GEE, ZOAB-SPAM-GEE and ZOABR-SPAM-GEE.

To construct the GEE's, given $\boldsymbol{\nu} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \gamma_k^\top)^\top$, $\boldsymbol{\kappa}$, $\boldsymbol{\sigma} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, lets consider the vectors $\boldsymbol{d}_i^* = (\boldsymbol{u}_i^\top, \boldsymbol{q}_i^\top, \boldsymbol{c}_{0i}, \boldsymbol{c}_{1i})^\top$, $\boldsymbol{u}_i = \boldsymbol{u}_i(\boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\kappa}) = (u_{i1}, \ldots, u_{in_i^*})^\top$, $\boldsymbol{q}_i = \boldsymbol{q}_i(\boldsymbol{y}_i, \boldsymbol{\nu}, \boldsymbol{\kappa}) = (q_{i1}, \ldots, q_{in_i^*})^\top$, $\boldsymbol{c}_{0i} = \boldsymbol{c}_{0i}(\boldsymbol{y}_i, \boldsymbol{\rho}, \boldsymbol{\tau}) = (\boldsymbol{c}_{0i1}, \ldots, \boldsymbol{c}_{0in_i^*})^\top$, $\boldsymbol{c}_{1i} = \boldsymbol{c}_{1i}(\boldsymbol{y}_i, \boldsymbol{\rho}, \boldsymbol{\tau}) = (\boldsymbol{c}_{1i1}, \ldots, \boldsymbol{c}_{1in_i^*})^\top$, $i = 1, \ldots, n$, where $\boldsymbol{u}_i$, $\boldsymbol{q}_i$, $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$ are the zero mean vectors as defined in Equation (4.3) related to the μ_{it}, ϕ_{it}, p_{0it} and p_{1it}, respectively, mutually independent and satisfying the properties of regular estimating functions (Crowder, 1987).

The choices for $\boldsymbol{u}_i$, $\boldsymbol{q}_i$, $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$ for all models will be made as the same way of Section 4.3. The choices will be made to bring advantages in the algorithm, as the the insensitivity property (Jørgensen and Knudsen, 2004), which allows a way of update the GEE's for mean and dispersion/precision parameters, in the estimation algorithm, separately.

Let us to define a quantity important for next subsections. Given that the distributions considering here, ZOAB, ZOAS and ZOABR, are a mixing of two other distributions (beta/simplex/beta rectangular with a Bernoulli), it is useful to define for each Y_{it}, the variable Z_i^* as

$$z_i^* = \begin{cases} 0, & if\ y_i \in (0,1) \\ 1, & if\ y_i \in \{0,1\} \end{cases} \tag{5.5}$$

It implies that when $Z_i^* = 0$, Y_i follows the beta(μ_{it}, ϕ_{it})/simplex(μ_{it}, ϕ_{it})/BR $(\mu_{it}, \phi_{it}, \alpha)$ distributions, while, when $Z_i^* = 1$, Y_i follows the Bernoulli(ϱ) distribution, where $\varrho = p_{1it}/(p_{0it} + p_{1it})$. It is also known that the probability of $Z_i^* = 1$, $\mathbb{P}(Y_i = 0 \text{ or } Y_i = 1) = \mathbb{P}(Z_i^* = 1)$ is equal to $p_{0i} + p_{1i}$, and, as consequence $\mathbb{P}(Z_i^* = 0) = 1 - p_{0i} - p_{1i}$. Thus, $Z_i^* \sim$ Bernoulli$(p_{0i} + p_{1i})$. Therefore, more details of the components that will compound the GEE's for each model, are described in the next subsections.

5.2.1 Components for ZOAB-SPAM-GEE

For the model based on the ZOAB distribution, the log-likelihood function $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \gamma_1^\top, \ldots, \gamma_k^\top, \boldsymbol{\kappa}^\top, \boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, considering $\boldsymbol{\varphi} = (\boldsymbol{\beta}^\top, \gamma_1^\top, \ldots, \gamma_k^\top, \boldsymbol{\kappa}^\top)^\top$ and $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, is defined as $l(\boldsymbol{\theta}) = l_1(\boldsymbol{\vartheta}) + l_2(\boldsymbol{\varphi})$, where,

$$\begin{aligned} l_1(\boldsymbol{\vartheta}) &= \sum_{i=1}^{n} z_i^*[(1-y_i)\log(p_{0i}) + y_i \log(p_{1i})] + (1-z_i^*)\log(1-p_{0i}-p_{1i}), \\ l_2(\boldsymbol{\varphi}) &= \sum_{i=1}^{n}(1-z_i^*)\{\log(\Gamma(\phi_i)) - \log(\Gamma(\mu_i\phi_i)) - \log(\Gamma([1-\mu_i]\phi_i)) \\ &\quad + [\mu_i\phi_i - 1]\log(y_i) + ([1-\mu_i]\phi_i - 1)\log[1-y_i]\}; \end{aligned}$$

$\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$, $\boldsymbol{\eta}_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{b}_{ij}^\top \boldsymbol{\gamma}_j$, $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$, $\boldsymbol{\eta}_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$, p_{0i} and p_{1i} defined in the Equation (5.2).

For this model, the quantity $\boldsymbol{q}_i$, $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$ will be defined as the score functions. Hence, $\boldsymbol{u}_i$, $\boldsymbol{q}_i$, $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$ are given by,

$$\begin{aligned} \boldsymbol{u}_i &= (1-\boldsymbol{z}_i^*)(\boldsymbol{y}_i - \boldsymbol{\mu}_i), \\ \boldsymbol{q}_i &= (1-\boldsymbol{z}_i^*)(\boldsymbol{\mu}_i(\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) + \log(1-\boldsymbol{y}_i) - \Psi((1-\boldsymbol{\mu}_i)\boldsymbol{\phi}_i) + \Psi(\boldsymbol{\phi}_i)), \\ \boldsymbol{c}_{0i} &= \boldsymbol{z}_i^*(1-\boldsymbol{y}_i) - \boldsymbol{p}_{0i}, \\ \boldsymbol{c}_{1i} &= \boldsymbol{z}_i^* \boldsymbol{y}_i - \boldsymbol{p}_{1i} \end{aligned} \tag{5.6}$$

where $\boldsymbol{\mu}^* = (\mu_{i1}^*, \ldots, \mu_{in_i^*}^*)^\top$, $\mu_{it}^* = \Psi(\mu_{it}\phi_{it}) - \Psi[(1-\mu_{it})\phi_{it}]$, $\boldsymbol{y}^* = (y_{i1}^*, \ldots, y_{in_i^*}^*)^\top$ and $y_{it}^* = \log[y_{it}/(1-y_{it})]$, for $i = 1, \ldots, n$ and $t = 1, \ldots, n_i^*$, with $\Psi(.)$ denoting digamma function, which is, $\Psi(z) = \partial \log \Gamma(z)/\partial z$, for $z > 0$. It can be noticed that $\mathbb{E}(u_{it}) = 0$ and, as $\boldsymbol{q}_i$, $\boldsymbol{c}_{i1}$ and $\boldsymbol{c}_{i1}$ are the score functions related to $\boldsymbol{\phi}_i$, $\boldsymbol{p}_{0i}$ and $\boldsymbol{p}_{1i}$, respectively, which implies to $\mathbb{E}(q_{it}) = 0 \Leftrightarrow \mathbb{E}(\log(1-y_{it})) = \Psi((1-\mu_{it})\phi_{it}) + \Psi(\phi_{it})$, $\mathbb{E}(c_{1it}) = 0 \Leftrightarrow \mathbb{E}(z_{it}^*(1-y_{it})) = p_{0it}$ and $\mathbb{E}(c_{2it}) = 0 \Leftrightarrow \mathbb{E}(z_{it}^* y_{it}) = p_{1it}$, respectively. Thus, the vectors $\boldsymbol{u}_i$, $\boldsymbol{q}_i$, $\boldsymbol{c}_{1i}$ and $\boldsymbol{c}_{2i}$ generate an optimal class of functions and they are vector with zero mean, mutually independent and satisfying the properties of regular estimating functions (Crowder, 1987). For the i-th subject, some quantities, which will be used in the next sections, are defined below

$$\begin{aligned} \boldsymbol{T}_{1i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^\top = \boldsymbol{G}_{1i}, \\ \boldsymbol{T}_{2i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{G}_{2i}\boldsymbol{C}_i, \\ \boldsymbol{T}_{21i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{0}, \\ \boldsymbol{T}_{12i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_{1i}}\right)^\top = \boldsymbol{G}_{1i}\boldsymbol{D}_i, \\ \boldsymbol{T}_{3i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\zeta}_{0i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{p}_{0i}}\frac{\partial \boldsymbol{p}_{0i}}{\partial \boldsymbol{\zeta}_{0i}}\right)^\top = \boldsymbol{G}_{3i} \\ \boldsymbol{T}_{4i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\zeta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{p}_{1i}}\frac{\partial \boldsymbol{p}_{1i}}{\partial \boldsymbol{\zeta}_{1i}}\right)^\top = \boldsymbol{G}_{4i} \end{aligned}$$

$$\boldsymbol{T}_{34i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\zeta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{p}_{0i}}\frac{\partial \boldsymbol{p}_{0i}}{\partial \boldsymbol{\zeta}_{1i}}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\zeta}_{0i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{p}_{1i}}\frac{\partial \boldsymbol{p}_{1i}}{\partial \boldsymbol{\zeta}_{0i}}\right)^\top = \boldsymbol{G}_{5i}$$

where, $\boldsymbol{\Sigma}_i = \mathrm{Cov}(\boldsymbol{u}_i) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\xi})\boldsymbol{A}_i^{1/2}$, $\boldsymbol{A}_i = \mathrm{diag}\{a_{i1}, \ldots, a_{in_i^*}\}$, $a_{it} = Var(Y_{it}) = [\mu_{it}(1-\mu_{it})]/(1+\phi_{it})$, $\boldsymbol{G}_{1i} = \mathrm{diag}\{(\partial\mu_{i1}/\partial\eta_{1i1}), \ldots, (\partial\mu_{in_i^*}/\partial\eta_{1in_i^*})\}$, $\boldsymbol{G}_{2i} = \mathrm{diag}\{(\partial\phi_{i1}/\partial\eta_{2i1}), \ldots, (\partial\phi_{in_i^*}/\partial\eta_{2in_i^*})\}$, $\boldsymbol{C}_i = \mathrm{diag}\{c_{i1}, \ldots, c_{in_i^*}\}$, $c_{it} = (1-z_{it}^*)(-\Psi'(\phi_{it}) + \mu_{it}^2\Psi'(\mu_{it}\phi_{it}) + (1-\mu_{it})^2\Psi'((1-\mu_{it})\phi_{it}))$, $\boldsymbol{D}_i = \mathrm{diag}\{d_{i1}, \ldots, d_{in_i^*}\}$, $d_{it} = (1-z_{it}^*)\phi_{it}[\mu_{it}\Psi'(\mu_{it}\phi_{it}) - (1-$

$\mu_{it}\Psi'(1-\mu_{it}))]$, $\boldsymbol{G}_{3i} = \text{diag}\{g_{3i1}, \ldots, g_{3in_i^*}\}$, $g_{3it} = p_{0it}(1-p_{0it})$, $\boldsymbol{G}_{4i} = \text{diag}\{g_{4i1}, \ldots, g_{4in_i^*}\}$, $g_{4it} = p_{1it}(1-p_{1it})$, $\boldsymbol{G}_{5i} = \text{diag}\{g_{5i1}, \ldots, g_{5in_i^*}\}$, $g_{5it} = -p_{0it}p_{1it}$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. Also, $\boldsymbol{R}(\boldsymbol{\xi})$ is a $(n_i^* \times n_i^*)$ symmetric matrix satisfying the conditions for a working correlation matrix (Liang and Zeger, 1986), which depends on a correlation parameter vector $\boldsymbol{\xi}$ that is the same for all subjects.

5.2.2 Components for ZOAS-SPAM-GEE

For the model based on the ZOAS distribution, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, considering $\boldsymbol{\varphi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ and $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, is defined as $l(\boldsymbol{\theta}) = l_1(\boldsymbol{\vartheta}) + l_2(\boldsymbol{\varphi})$, where,

$$
\begin{aligned}
l_1(\boldsymbol{\vartheta}) &= \sum_{i=1}^{n} z_i^*[(1-y_i)\log(p_{0i}) + y_i \log(p_{1i})] + (1-z_i^*)\log(1-p_{0i}-p_{1i}),\\
l_2(\boldsymbol{\varphi}) &= \sum_{i=1}^{n}(1-z_i^*)\left\{-\frac{1}{2}\left\{\log(2\pi\phi_i) + 3\log[y_i(1-y_i)]\right\} - \frac{1}{2\phi_i} d(y_i;\mu_i)\right\},
\end{aligned}
$$

where, $d(y;\mu)$ is defined Equation (1.7), $\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$, $\boldsymbol{\eta}_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{b}_{ij}^\top \boldsymbol{\gamma}_j$, $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$ and $\boldsymbol{\eta}_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$.

For this model, $\boldsymbol{u}_i$, $\boldsymbol{q}_i$, $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$ will be defined as the score functions. They are given by,

$$
\begin{aligned}
&\boldsymbol{u}_i = -(1-\boldsymbol{z}_i^*)\frac{1}{2\boldsymbol{\phi}_i} d'(\boldsymbol{y}_i, \boldsymbol{\mu}_i), \quad \boldsymbol{q}_i = (1-\boldsymbol{z}_i^*)\frac{1}{2\boldsymbol{\phi}_i}\left[\frac{1}{\boldsymbol{\phi}_i} d(\boldsymbol{y}_i, \boldsymbol{\mu}_i) - 1\right],\\
&\boldsymbol{c}_{0i} = \boldsymbol{z}_i^*(1-\boldsymbol{y}_i) - \boldsymbol{p}_{0i}, \quad \boldsymbol{c}_{1i} = \boldsymbol{z}_i^* \boldsymbol{y}_i - \boldsymbol{p}_{1i},
\end{aligned}
\tag{5.7}
$$

where

$$
d'(\boldsymbol{y}_i, \boldsymbol{\mu}_i) = \frac{\partial d(\boldsymbol{y}_i, \boldsymbol{\mu}_i)}{\partial \boldsymbol{\mu}_i} = -\frac{2(\boldsymbol{y}_i - \boldsymbol{\mu}_i)}{\boldsymbol{\mu}_i(1-\boldsymbol{\mu}_i)}\left[d(\boldsymbol{y}_i, \boldsymbol{\mu}_i) + \frac{1}{\boldsymbol{\mu}_i^2(1-\boldsymbol{\mu}_i)^2}\right],
$$

$t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$.

It can be noticed that $\mathbb{E}(u_{it}) = 0 \Leftrightarrow \mathbb{E}(d'(y_{it}, \mu_{it})) = 0$ and $\mathbb{E}(q_{it}) = 0 \Leftrightarrow \mathbb{E}(d(y_{it}, \mu_{it})) = \phi_{it}$, which is guaranteed since this quantities are the score functions. Thus, the vectors $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ generate an optimal class of functions and they are vector with zero mean, mutually independent and satisfying the properties of regular estimating functions (Crowder, 1987). For the i-th subject, some quantities, which will be used in the next sections, are defined below

$$\boldsymbol{T}_{1i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^\top = \boldsymbol{G}_{1i}\boldsymbol{A}_i$$

$$\boldsymbol{T}_{2i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{G}_{2i}\boldsymbol{C}_i,$$

$$\boldsymbol{T}_{21i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = \boldsymbol{T}_{12i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = \boldsymbol{0},$$

$$\boldsymbol{T}_{3i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\zeta}_{0i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{p}_{0i}}\frac{\partial \boldsymbol{p}_{0i}}{\partial \boldsymbol{\zeta}_{0i}}\right)^\top = \boldsymbol{G}_{3i}$$

$$\boldsymbol{T}_{4i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\zeta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{p}_{1i}}\frac{\partial \boldsymbol{p}_{1i}}{\partial \boldsymbol{\zeta}_{1i}}\right)^\top = \boldsymbol{G}_{4i}$$

$$\boldsymbol{T}_{34i} = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\zeta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{p}_{0i}}\frac{\partial \boldsymbol{p}_{0i}}{\partial \boldsymbol{\zeta}_{1i}}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\zeta}_{0i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{p}_{1i}}\frac{\partial \boldsymbol{p}_{1i}}{\partial \boldsymbol{\zeta}_{0i}}\right)^\top = \boldsymbol{G}_{5i},$$

and $\boldsymbol{\Sigma}_i = \text{Cov}(\boldsymbol{u}_i) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\xi})\boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \text{diag}\{a_{i1}, \ldots, a_{in_i^*}\}$, with

$$a_{it} = (1 - z_{it}^*)\frac{1}{\phi_{it}}\left[\frac{3\phi_{it}}{\mu_{it}(1-\mu_{it})} + \frac{1}{\mu_{it}^3(1-\mu_{it})^3}\right],$$

$\boldsymbol{C}_i = \text{diag}\{c_{i1}, \ldots, c_{in_i^*}\}$, where $c_{ij} = (1 - z_{it}^*)\dfrac{1}{2\phi_{it}^2}$, $\boldsymbol{G}_{1i} = \text{diag}\{(\partial\mu_{i1}/\partial\eta_{1i1}), \ldots, (\partial\mu_{in_i^*}/\partial\eta_{1in_i^*})\}$, $\boldsymbol{G}_{2i} = \text{diag}\{(\partial\phi_{i1}/\partial\eta_{2i1}), \ldots, (\partial\phi_{in_i^*}/\partial\eta_{2in_i^*})\}$, $\boldsymbol{G}_{3i} = \text{diag}\{g_{3i1}, \ldots, g_{3in_i^*}\}$, $g_{3it} = p_{0it}(1 - p_{0it})$, $\boldsymbol{G}_{4i} = \text{diag}\{g_{4i1}, \ldots, g_{4in_i^*}\}$, $g_{4it} = p_{1it}(1 - p_{1it})$, $\boldsymbol{G}_{5i} = \text{diag}\{g_{5i1}, \ldots, g_{5in_i^*}\}$, $g_{5it} = -p_{0it}p_{1it}$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. Also, $\boldsymbol{R}(\boldsymbol{\xi})$ is a $(n_i^* \times n_i^*)$ symmetric matrix satisfying the conditions for a working correlation matrix (Liang and Zeger, 1986), which depends on a correlation parameter vector $\boldsymbol{\xi}$ that is the same for all subjects.

5.2.3 Components for ZOABR-SPAM-GEE

For the model based on the ZOABR distribution, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha, \boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, considering $\boldsymbol{\varphi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$ and $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, is defined as $l(\boldsymbol{\theta}) = l_1(\boldsymbol{\vartheta}) + l_2(\boldsymbol{\varphi})$, where,

$$l_1(\boldsymbol{\vartheta}) = \sum_{i=1}^n z_i^*[(1 - y_i)\log(p_{0i}) + y_i\log(p_{1i})] + (1 - z_i^*)\log(1 - p_{0i} - p_{1i}), \tag{5.8}$$

$$l_2(\boldsymbol{\varphi}) = \sum_{i=1}^n (1 - z_i^*)\log[\boldsymbol{\varepsilon}_i + (1 - \boldsymbol{\varepsilon}_i)b_{Y_i}(\boldsymbol{y}_i; \boldsymbol{\delta}_i, \boldsymbol{\phi}_i)], \tag{5.9}$$

with $\boldsymbol{\varepsilon}_i = 1 - \sqrt{1 - 4\alpha\boldsymbol{\mu}_i(1 - \boldsymbol{\mu}_i)}$, $b_Y(y; ., .)$ is the beta density defined in Equation (1.1), $\boldsymbol{\delta}_i = \dfrac{\boldsymbol{\mu}_i - \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4\alpha\boldsymbol{\mu}_i(1 - \boldsymbol{\mu}_i)}}{\sqrt{1 - 4\alpha\boldsymbol{\mu}_i(1 - \boldsymbol{\mu}_i)}}$, where, $\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$, $\boldsymbol{\eta}_{1i} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^k \boldsymbol{b}_{ij}^\top\boldsymbol{\gamma}_j$, $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$ and $\boldsymbol{\eta}_{2i} = \boldsymbol{e}_i^\top\boldsymbol{\kappa}$.

For this model, the log-likelihood related with the continuous part is not easily tractable to obtain certain quantities derived from it as commented in the Section 2.2.1. Thus, the vectors $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ will not be defined based on quantities as score functions, for example. An alternative, in this case, can be the proposal of X.-K. Song et al. (2004), who constructed the zero mean vector for the estimating equation of the dispersion component of their distribution, based on a quantity whose expectation was a function of the dispersion parameter. In our case, we are working with the precision parameter of the beta rectangular distribution, thus a useful option is to define $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{in_i^*})$, with $q_{it} = (y_{it} - \mu_{it})^2 - \text{Var}(Y_{it})$, where $\mathbb{E}[(Y_{it} - \mu_{it})^2] = \text{Var}(Y_{it})$, and $\text{Var}(Y_{it})$, defined in Equation (1.4), is a function of ϕ_{it}, for $t = 1, \ldots, n_i^*$. This proposal vector will make calculations easier and still having good properties. For the $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$, it will be remained the use of score functions, since the log-likelihood related with discrete part is tractable. Hence, $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ for ZOABR-SPAM-GEE are defined below

$$
\begin{aligned}
\boldsymbol{u}_i &= (1 - \boldsymbol{z}_i^*)(\boldsymbol{y}_i - \boldsymbol{\mu}_i), \\
\boldsymbol{q}_i &= (1 - \boldsymbol{z}_i^*)\left[(\boldsymbol{y}_i - \boldsymbol{\mu}_i)^2 - \left\{\frac{\boldsymbol{\delta}_i(1 - \boldsymbol{\delta}_i)}{1 + \boldsymbol{\phi}_i}(1 - \boldsymbol{\varepsilon}_i)[1 - \boldsymbol{\varepsilon}_i(1 + \boldsymbol{\phi}_i)] + \frac{\boldsymbol{\varepsilon}_i}{12}(4 - 3\boldsymbol{\varepsilon}_i)\right\}\right], \quad (5.10) \\
\boldsymbol{c}_{0i} &= \boldsymbol{z}_i^*(1 - \boldsymbol{y}_i) - \boldsymbol{p}_{0i}, \quad \boldsymbol{c}_{1i} = \boldsymbol{z}_i^* \boldsymbol{y}_i - \boldsymbol{p}_{1i},
\end{aligned}
$$

$i = 1, \ldots, n$. It can be noticed that $\mathbb{E}(\boldsymbol{u}_i) = \boldsymbol{0}$ and $\mathbb{E}(\boldsymbol{q}_i) = \boldsymbol{0}$. Thus, the vectors $\boldsymbol{u}_i$ and $\boldsymbol{q}_i$ generate an optimal class of functions and they are vector with zero mean, mutually independent and satisfying the properties of regular estimating functions (Crowder, 1987). For the i-th subject, some quantities, which will be used in the next sections, are defined below

$$
\begin{aligned}
\boldsymbol{T}_{1i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^\top = \boldsymbol{G}_{1i} \\
\boldsymbol{T}_{2i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{G}_{2i}\boldsymbol{C}_i, \\
\boldsymbol{T}_{21i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\eta}_{2i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\phi}_i}\frac{\partial \boldsymbol{\phi}_i}{\partial \boldsymbol{\eta}_{2i}}\right)^\top = \boldsymbol{0}, \\
\boldsymbol{T}_{12i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\eta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\mu}_i}\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_{1i}}\right)^\top = \boldsymbol{G}_{1i}D_i, \\
\boldsymbol{T}_{3i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\zeta}_{0i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{p}_{0i}}\frac{\partial \boldsymbol{p}_{0i}}{\partial \boldsymbol{\zeta}_{0i}}\right)^\top = \boldsymbol{G}_{3i} \\
\boldsymbol{T}_{4i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\zeta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{p}_{1i}}\frac{\partial \boldsymbol{p}_{1i}}{\partial \boldsymbol{\zeta}_{1i}}\right)^\top = \boldsymbol{G}_{4i} \\
\boldsymbol{T}_{34i} &= -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\zeta}_{1i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{p}_{0i}}\frac{\partial \boldsymbol{p}_{0i}}{\partial \boldsymbol{\zeta}_{1i}}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\zeta}_{0i}^\top}\right)^\top = -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{p}_{1i}}\frac{\partial \boldsymbol{p}_{1i}}{\partial \boldsymbol{\zeta}_{0i}}\right)^\top = \boldsymbol{G}_{5i},
\end{aligned}
$$

and $\boldsymbol{\Sigma}_i = \text{Cov}(\boldsymbol{u}_i) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\xi})\boldsymbol{A}_i^{1/2}$, where $\boldsymbol{A}_i = \text{diag}\{a_{i1}, \ldots, a_{in_i^*}\}$,

$$a_{it} = (1 - z_{it}^*)\left\{\frac{\boldsymbol{\delta}_i(1-\boldsymbol{\delta}_i)}{1+\boldsymbol{\phi}_i}(1-\boldsymbol{\varepsilon}_i)[1-\varepsilon(1+\boldsymbol{\phi}_i)] + \frac{\boldsymbol{\varepsilon}_i}{12}(4-3\boldsymbol{\varepsilon}_i)\right\},$$

$\boldsymbol{G}_{1i} = \text{diag}\{(\partial\mu_{i1}/\partial\eta_{1i1}), \ldots, (\partial\mu_{in_i^*}/\partial\eta_{1in_i^*})\}$, $\boldsymbol{C}_i = \text{diag}\{c_{i1}, \ldots, c_{in_i^*}\}$,

$$c_{ij} = (1 - z_{it}^*)\frac{\delta_{it}(1-\delta_{it})}{(1+\phi_{it})^2}(\varepsilon_{it} - 1),$$

$\boldsymbol{G}_{2i} = \text{diag}\{(\partial\phi_{i1}/\partial\eta_{2i1}), \ldots, (\partial\phi_{in_i^*}/\partial\eta_{2in_i^*})\}$, $\boldsymbol{D}_i = \text{diag}\{d_{i1}, \ldots, d_{in_i^*}\}$, where

$$d_{it} = (1 - z_{it}^*)\Bigg\{(1-2\delta_{it})\frac{(1-\alpha)[1-\varepsilon_{it}(1+\phi_{it})]}{(1+\phi_{it})(1-\varepsilon_{it})^2} + \left[\frac{\delta_{it}(1-\delta_{it})}{(1+\phi_{it})}(-2-\phi_{it}+2\varepsilon_{it}(1+\phi_{it})) + \frac{2-3\varepsilon_{it}}{6}\right]\left(\frac{2\alpha(1-2\mu_{it})}{1-\varepsilon_{it}}\right)\Bigg\},$$

$\boldsymbol{G}_{3i} = \text{diag}\{g_{3i1}, \ldots, g_{3in_i^*}\}$, $g_{3it} = p_{0it}(1-p_{0it})$, $\boldsymbol{G}_{4i} = \text{diag}\{g_{4i1}, \ldots, g_{4in_i^*}\}$, $g_{4it} = p_{1it}(1-p_{1it})$, $\boldsymbol{G}_{5i} = \text{diag}\{g_{5i1}, \ldots, g_{5in_i^*}\}$, $g_{5it} = -p_{0it}p_{1it}$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. Also, $\boldsymbol{R}(\boldsymbol{\xi})$ is a $(n_i^* \times n_i^*)$ symmetric matrix satisfying the conditions for a working correlation matrix (Liang and Zeger, 1986), which depends on a correlation parameter vector $\boldsymbol{\xi}$ that is the same for all subjects. A last necessary quantity is the $\boldsymbol{\Upsilon}_i = \text{Cov}(\boldsymbol{q}_i)$. Also, following X.-K. Song et al. (2004) who suggest adopting $\text{Cov}(\boldsymbol{q}_i)$ as the $\boldsymbol{J}_i = \text{diag}\{j_{i1}, \ldots, j_{in_i^*}\}$, with $j_{it} = \text{Var}[(Y_{it} - \mu_{it})^2]$, $t = 1, \ldots, n_i^*$ and $i = 1, \ldots, n$. For ease, this quantity, in the estimation process, will be obtained numerically.

5.2.4 Penalized GEE for all models

As discussed in Section 4.3.4, a penalty function based on differences will be added on GEE related with the non-parametric components as defined in the Section 1.2.2. Moreover, Kong et al. (2015) discussed that an approach using only one GEE to estimating the parameters related with the discrete and continuous part may not be identifiable because those parts are typically confounded (e.g., share information). Hence, they recommended to estimate the parameters related with those parts in separate equations by introducing the variables Z_{it}^* $(i = 1, \ldots, n; t = 1, \ldots, n_i^*)$, which indicates whether the random variable Y_{it} is from a Bernoulli distribution or beta/simplex/BR distribution.

Thus, considering the components related to each distribution, the generalized estimating equations for repeated measures for $\boldsymbol{\nu} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \gamma_k^\top)^\top$, $\boldsymbol{\kappa}$, $\boldsymbol{\sigma} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, respectively, are given by:

$$\Psi_{1p}(\boldsymbol{\nu},\boldsymbol{\lambda}) = \sum_{i=1}^{n} -\mathbb{E}\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\nu}^\top}\right)^\top \mathrm{Cov}(\boldsymbol{u}_i)^{-1}\boldsymbol{Z}^{**}\boldsymbol{u}_i - \boldsymbol{\Lambda} = \boldsymbol{B}_i^{*\top}\boldsymbol{T}_{1i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{Z}^{**}\boldsymbol{u}_i - \boldsymbol{\Lambda}$$

$$\Psi_2(\boldsymbol{\kappa}) = \sum_{i=1}^{n} -\mathbb{E}\left(\frac{\partial \boldsymbol{q}_i}{\partial \boldsymbol{\kappa}^\top}\right)^\top \mathrm{Cov}(\boldsymbol{q}_i)^{-1}\boldsymbol{Z}^{**}\boldsymbol{q}_i = \boldsymbol{E}_i^\top\boldsymbol{T}_{2i}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{Z}^{**}\boldsymbol{q}_i,$$

$$\begin{aligned}
\Psi_3(\boldsymbol{\sigma}) &= \sum_{i=1}^{n}\begin{pmatrix} -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\rho}^\top}\right)^\top & -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\rho}^\top}\right)^\top \\ -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{0i}}{\partial \boldsymbol{\tau}^\top}\right)^\top & -\mathbb{E}\left(\frac{\partial \boldsymbol{c}_{1i}}{\partial \boldsymbol{\tau}^\top}\right)^\top \end{pmatrix}\mathrm{Cov}(\boldsymbol{c}_i)\boldsymbol{c}_i \\
&= \sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{F}_i^\top\boldsymbol{T}_{3i} & \boldsymbol{F}_i^\top\boldsymbol{T}_{34i} \\ \boldsymbol{M}_i^\top\boldsymbol{T}_{34i} & \boldsymbol{M}_i^\top\boldsymbol{T}_{4i}\end{pmatrix}\begin{pmatrix}\mathrm{Cov}(\boldsymbol{c}_{0i}) & \mathrm{Cov}(\boldsymbol{c}_{0i},\boldsymbol{c}_{1i}) \\ \mathrm{Cov}(\boldsymbol{c}_{0i},\boldsymbol{c}_{1i}) & \mathrm{Cov}(\boldsymbol{c}_{1i})\end{pmatrix}^{-1}\boldsymbol{c}_i \\
&= \sum_{i=1}^{n}\boldsymbol{Q}_2^\top\boldsymbol{T}_i^*\boldsymbol{\Phi}^{-1}\boldsymbol{c}_i = \sum_{i=1}^{n}\boldsymbol{Q}_2^\top\boldsymbol{W}_{3i}\boldsymbol{T}_i^{*-1}\boldsymbol{c}_i
\end{aligned}$$

where $\boldsymbol{c}_i = (\boldsymbol{c}_{0i}^\top, \boldsymbol{c}_{1i}^\top)^\top$, $\boldsymbol{B}_i^* = (\boldsymbol{X}_i, \boldsymbol{B}_{ij}, \ldots, \boldsymbol{B}_{ik})$, $\boldsymbol{\Lambda} = (\boldsymbol{0}_p, \lambda_1\boldsymbol{\gamma}_1^\top\boldsymbol{\Lambda}_1^d, \ldots, \lambda_k\boldsymbol{\gamma}_k^\top\boldsymbol{\Lambda}_k^d)^\top$, with $\boldsymbol{0}_p$ being a $p\times 1$ vector of zeros and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$ being the vector of smooth parameters and $\lambda_j > 0,\ \forall j, j = 1, \ldots, k$, $\boldsymbol{T}_{1i} = \mathrm{diag}\{T_{1i1}, \ldots, T_{1in_i^*}\}$, $\boldsymbol{T}_{2i} = \mathrm{diag}\{T_{2i1}, \ldots, T_{2in_i^*}\}$, $\boldsymbol{\Sigma}_i = \mathrm{diag}\{\Sigma_{i1}, \ldots, \Sigma_{in_i^*}\}$ and $\boldsymbol{\Upsilon}_i = \mathrm{diag}\{\Upsilon_{i1}, \ldots, \Upsilon_{in_i^*}\}$. Also, $\boldsymbol{T}_{3i} = \mathrm{diag}\{T_{3i1}, \ldots, T_{3in_i^*}\}$, $\boldsymbol{T}_{4i} = \mathrm{diag}\{T_{4i1}, \ldots, T_{4in_i^*}\}$, $\boldsymbol{T}_{34i} = \mathrm{diag}\{T_{34i1}, \ldots, T_{34in_i^*}\}$,

$$\boldsymbol{Q}_{2i} = \begin{pmatrix}\boldsymbol{F}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M}_i\end{pmatrix}, \quad \boldsymbol{W}_{3i} = \boldsymbol{T}_i^*\boldsymbol{\Delta}_i^{-1}\boldsymbol{T}_i^*, \quad \boldsymbol{T}_i^* = \begin{pmatrix}\boldsymbol{F}_i^\top\boldsymbol{T}_{3i} & \boldsymbol{F}_i^\top\boldsymbol{T}_{34i} \\ \boldsymbol{M}_i^\top\boldsymbol{T}_{34i} & \boldsymbol{M}_i^\top\boldsymbol{T}_{4i}\end{pmatrix}$$

$$\boldsymbol{\Phi} = \begin{pmatrix}\mathrm{Cov}(\boldsymbol{c}_{0i}) & \mathrm{Cov}(\boldsymbol{c}_{0i},\boldsymbol{c}_{1i}) \\ \mathrm{Cov}(\boldsymbol{c}_{0i},\boldsymbol{c}_{1i}) & \mathrm{Cov}(\boldsymbol{c}_{1i})\end{pmatrix}.$$

In addition, $\boldsymbol{Z}_i^{**} = \mathrm{diag}\{(1-z_{i1}^*), \ldots, (1-z_{in_i^*}^*)\}$, which is present in the GEE's related to the parameters of continuous part to indicates what observations come from beta/simplex or BR distributions. On the other hand, the vectors $\boldsymbol{c}_{0i}$ and $\boldsymbol{c}_{1i}$, contains a component $\boldsymbol{Z}_i^*$, which indicates observation that are from discrete part.

The matrix $\boldsymbol{\Upsilon}_i = \mathrm{Cov}(\boldsymbol{q}_i)$, $\mathrm{Cov}(\boldsymbol{c}_{0i})$ and $\mathrm{Cov}(\boldsymbol{c}_{1i})$, can be replace by a work correlation matrices but, it would imply a high amount of disturbance parameters to be estimated. In the sense, it will be assume, as made by Freitas et al. (2021) and X.-K. Song et al. (2004), that for ZOAS-SPAM-GEE and ZOAB-SPAM-GEE that $\boldsymbol{\Upsilon}_i = \boldsymbol{C}_i$ and for ZOABR-SPAM-GEE, $\boldsymbol{\Upsilon}_i = \boldsymbol{J}_i$. Also, that $\mathrm{Cov}(\boldsymbol{u}_i, \boldsymbol{q}_i) = \mathrm{Cov}(\boldsymbol{c}_{0i}, \boldsymbol{c}_{1i}) = 0$ (Prentice and Zhao, 1991). As Freitas et al. (2021) cited, by assuming this, we move away from an optimal linear estimation function but reduce the amount of assumptions to be made about the dependence structure of the data. Then, the penalized GEE for $\boldsymbol{\nu}$, $\boldsymbol{\kappa}$, $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$ are given by $\Psi_{1p}(\hat{\boldsymbol{\nu}}, \boldsymbol{\lambda}) = \boldsymbol{0}$, $\Psi_2(\hat{\boldsymbol{\kappa}}) = \boldsymbol{0}$ and $\Psi_3(\hat{\boldsymbol{\sigma}}) = \boldsymbol{0}$, where $\hat{\boldsymbol{\nu}}$, $\hat{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\sigma}}$ denotes the respective estimates.

For next sections, except for the parameter estimation, it is useful defined the GEE's concatenated, i.e. $\Psi^*(\boldsymbol{\nu},\boldsymbol{\kappa}) = (\Psi_{1p}(\boldsymbol{\nu},\boldsymbol{\lambda})^\top, \Psi_2(\boldsymbol{\kappa})^\top, \Psi_3(\boldsymbol{\sigma})^\top)^\top$, which can be rewritten as,

$$\boldsymbol{\Psi}^*(\boldsymbol{\nu},\boldsymbol{\kappa}) = \sum_{i=1}^{n} \boldsymbol{Q}_i^\top \boldsymbol{W}_i \boldsymbol{T}_i^{-1} \boldsymbol{d}_i^* - \boldsymbol{\Lambda}_2, \tag{5.11}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{0}_p, \lambda_1\boldsymbol{\gamma}_1^\top\boldsymbol{\Lambda}_1^d, \ldots, \lambda_k\boldsymbol{\gamma}_k^\top\boldsymbol{\Lambda}_k^d, \boldsymbol{0}_q, \boldsymbol{0}_{s_0+s_1})^\top$, $\boldsymbol{0}_q$ being a $q \times 1$ vector of zeros, $\boldsymbol{0}_{s_0+s_1}$ being a $(s_0+s_1)\times 1$ vector of zeros, $\boldsymbol{d}_i^* = (\boldsymbol{u}_i^\top, \boldsymbol{q}_i^\top, \boldsymbol{c}_i^\top)^\top$, $\boldsymbol{W} = \boldsymbol{T}_i\boldsymbol{P}_i^{-1}\boldsymbol{Z}_i^{***}\boldsymbol{T}_i$, $\boldsymbol{Z}_i^{***} =$ blockdiag$\{\boldsymbol{Z}_i^{**}, \boldsymbol{Z}_i^{**}, \boldsymbol{I}_{s^*}\}$, $\boldsymbol{I}_{s^*}$, being a $(s_0+s_1)\times(s_0+s_1)$ identity matrix,

$$\boldsymbol{Q}_i = \begin{pmatrix} \boldsymbol{B}_i^* & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{E}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Q}_{2i} \end{pmatrix}, \ \boldsymbol{T}_i = \begin{pmatrix} \boldsymbol{T}_{11i} & \boldsymbol{T}_{12i} & \boldsymbol{0} \\ \boldsymbol{T}_{21i} & \boldsymbol{T}_{22i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{T}_i^* \end{pmatrix}, \boldsymbol{P}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Upsilon}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Phi}_i \end{pmatrix}.$$

5.3 Parameter estimation

Before presenting the estimation process, it will be demonstrated that the GEE's for $\boldsymbol{\nu}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\sigma}$ can be updated separately for all models by using the insensitivity property (Jørgensen and Knudsen, 2004). As defined in a generic way in Equation (4.2), the sensibility matrix can be write as

$$\boldsymbol{S}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \boldsymbol{S}_{13} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} & \boldsymbol{S}_{23} \\ \boldsymbol{S}_{31} & \boldsymbol{S}_{32} & \boldsymbol{S}_{33} \end{pmatrix},$$

where $\boldsymbol{S}_{11} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\nu}^\top}\boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right]$, $\boldsymbol{S}_{22} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{\Psi}_2(\boldsymbol{\kappa})\right]$, $\boldsymbol{S}_{12} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\nu}^\top}\boldsymbol{\Psi}_2(\boldsymbol{\kappa})\right]$, $\boldsymbol{S}_{21} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right]$, $\boldsymbol{S}_{33} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\sigma}^\top}\boldsymbol{\Psi}_3(\boldsymbol{\sigma})\right]$, $\boldsymbol{S}_{31} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\sigma}^\top}\boldsymbol{\Psi}_{1p}(\boldsymbol{\nu})\right]$, $\boldsymbol{S}_{32} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\sigma}^\top}\boldsymbol{\Psi}_2(\boldsymbol{\kappa})\right]$, $\boldsymbol{S}_{13} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\beta}^\top}\boldsymbol{\Psi}_3(\boldsymbol{\sigma})\right]$ and $\boldsymbol{S}_{23} = \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{\Psi}_3(\boldsymbol{\sigma})\right]$

For prove the insensitivity, it is enough to show that the quantities $\boldsymbol{S}_{21}$, $\boldsymbol{S}_{31}$ and $\boldsymbol{S}_{32}$ are zero. For $\boldsymbol{S}_{13}$ and $\boldsymbol{S}_{23}$ it a natural imply given that the discrete part and continuous part are estimate without have component in common. The demonstration for $\boldsymbol{S}_{21}$ can be seen in Section 4.4. Thus, the sensibility matrix can be rewritten for all models as

$$\boldsymbol{S}_i(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} & \boldsymbol{S}_{13} \\ \boldsymbol{0} & \boldsymbol{S}_{22} & \boldsymbol{S}_{23} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{S}_{33} \end{pmatrix}, \tag{5.12}$$

and according with Jørgensen and Knudsen (2004), this configuration of the sensibility matrix implies in the insensitivity property, which allows us to separate $\boldsymbol{\nu}$, $\boldsymbol{\kappa}$, $\boldsymbol{\sigma}$ onto two equations to be updated in each step of estimation algorithm, following Bonat et al. (2018) proposal. The other entries of this matrix are defined as $\boldsymbol{S}_{11} = -\boldsymbol{B}_i^{*\top}\boldsymbol{T}_{1i}\boldsymbol{Z}_i^{**}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{T}_{1i}\boldsymbol{B}_i^* - \boldsymbol{\Lambda}^*$, with $\boldsymbol{\Lambda}^* = (\mathbf{0}_p, \lambda_1\boldsymbol{\Lambda}_1^d, \ldots, \lambda_k\boldsymbol{\Lambda}_k^d)^\top$, $\boldsymbol{S}_{22} = -\boldsymbol{E}_i^\top \boldsymbol{T}_{2i}\boldsymbol{Z}_i^{**}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{T}_{2i}\boldsymbol{E}_i$ and $\boldsymbol{S}_{12} = -\boldsymbol{E}_i^\top \boldsymbol{T}_{12i}\boldsymbol{Z}_i^{**}\boldsymbol{\Upsilon}_i^{-1}$ $\boldsymbol{T}_{12i}\boldsymbol{X}_i$, for ZOAB-SPAM-GEE and ZOABR-SPAM-GEE and $\boldsymbol{S}_{12} = \mathbf{0}$ for ZOAS-SPAM-GEE. Also, $\boldsymbol{S}_{13} = \boldsymbol{S}_{23} = \mathbf{0}$ and $\boldsymbol{S}_{33} = -\boldsymbol{Q}_{2i}^\top \boldsymbol{T}_i^* \boldsymbol{\Phi}_i^{-1}\boldsymbol{T}_i^*\boldsymbol{Q}_{2i}$

Now, developing the estimation algorithm by applying a combination of Gauss-Seidel method for the vector $\boldsymbol{\nu}$ and reweighed least squares for, $\boldsymbol{\kappa}$ and $\boldsymbol{\sigma}$, the $(u+1)$-th step of the iterative process for obtaining the maximum penalized likelihood estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_k$, $\boldsymbol{\kappa}$ and $\boldsymbol{\sigma}$ by fixing λ, may be expressed as

$$\begin{aligned}
\boldsymbol{\beta}^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{W}_{1i}^{(u)}\boldsymbol{X}_i\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{X}_i^\top \boldsymbol{W}_{1i}^{(u)}\left(\boldsymbol{z}_{1i}^{(u)} - \sum_{j=1}^k \boldsymbol{B}_{ij}\boldsymbol{\gamma}_j^{(u+1)}\right)\right], \\
\boldsymbol{\gamma}_j^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{B}_{ij}^\top \boldsymbol{W}_{1i}^{(u)}\boldsymbol{B}_{ij}\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{B}_{ij}^\top \boldsymbol{W}_{1i}^{(u)}\left(\boldsymbol{z}_{1i}^{(u)} - \boldsymbol{X}_i\boldsymbol{\beta}^{(u+1)} - \sum_{j^*\neq j} \boldsymbol{B}_{ij^*}\boldsymbol{\gamma}_{j^*}^{(u+1)}\right)\right], \\
\boldsymbol{\kappa}^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{E}_i^\top \boldsymbol{W}_{2i}^{(u)}\boldsymbol{E}_i\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{E}_i^\top \boldsymbol{W}_{2i}^{(u)}\boldsymbol{z}_{2i}^{(u)}\right], \\
\boldsymbol{\sigma}^{(u+1)} &= \left(\sum_{i=1}^n \boldsymbol{Q}_{2i}^\top \boldsymbol{W}_{3i}^{(u)}\boldsymbol{Q}_{2i}\right)^{-1} \left[\sum_{i=1}^n \boldsymbol{Q}_{2i}^\top \boldsymbol{W}_{3i}^{(u)}\boldsymbol{z}_{3i}^{(u)}\right]
\end{aligned}$$

for $j = 1, \ldots, k$ and $u = 0, 1, 2, \ldots$, where $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\boldsymbol{\gamma}}_j^{(0)}$ and $\hat{\boldsymbol{\kappa}}^{(0)}$ are the initial estimates, $\boldsymbol{z}_{1i} = \boldsymbol{\eta}_{1i} + \boldsymbol{T}_{1i}^{-1}\boldsymbol{u}_i$, $\boldsymbol{z}_{2i} = \boldsymbol{\eta}_{2i} + \boldsymbol{T}_{2i}^{-1}\boldsymbol{q}_i$, $\boldsymbol{z}_{3i} = \boldsymbol{\zeta}_i + \boldsymbol{T}_i^{*-1}\boldsymbol{c}_i$, $\boldsymbol{\zeta}_i = (\zeta_{0i}, \zeta_{1i})$, $\boldsymbol{W}_{1i} = \boldsymbol{T}_{1i}\boldsymbol{Z}_i^{**}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{T}_{1i}$, $\boldsymbol{W}_{2i} = \boldsymbol{T}_{2i}\boldsymbol{Z}_i^{**}\boldsymbol{\Upsilon}_i^{-1}\boldsymbol{T}_{2i}$ and $\boldsymbol{W}_{3i} = \boldsymbol{T}_i^*\boldsymbol{\Phi}_i^{-1}\boldsymbol{T}_i^*$.

For the model ZOABR-SPAM-GEE, there is one extra parameter to be estimated, which is α. Thus, only for this model one more parameter will be updating in each iteration. The estimation process of it will be made through maximum likelihood, using an optimization algorithm. The one chosen for maximization process was the `optim` of the software `R` using the method L-BFGS-B (Byrd et al., 1995). The updating, in the step $(u+1)$ will happen using the current estimates of the other parameters, replacing them in the log-likelihood function, $l(\boldsymbol{\theta})$, defined in Equation (5.8). The optimization happens and the new estimate for α is used to updating the other parameters. It will occurs until the converge is reached.

The parameters related to the correlation matrix are estimated as described in Section 4.5 depend on the dependence structure assumed (AR-1,Exchangeable and Unstructured).

5.4 Effective degrees of freedom

Following Manghi et al. (2019), the effective degrees of freedom will be derived considering the solution for the linear predictor, only related with $\boldsymbol{\nu}$, because it contains the non-parametric component. Hence, the solution for the linear predictor at the convergence of the iterative process, namely $\hat{\boldsymbol{\eta}}_{1i} = \boldsymbol{B}_i^* \hat{\boldsymbol{\nu}} = \hat{\boldsymbol{H}}(\boldsymbol{\lambda}) \hat{\boldsymbol{z}}_{1i}$, where $\widehat{\boldsymbol{H}}(\boldsymbol{\lambda}) = \boldsymbol{B}_i^* (\boldsymbol{B}^{*\top} \widehat{\boldsymbol{W}}_{1i} \boldsymbol{B}^* + \boldsymbol{\Lambda})^{-1} \boldsymbol{B}_i^{*\top} \widehat{\boldsymbol{W}}_{1i}$ may be interpreted as a projection matrix or smoother, for $i = 1, \ldots, n$.

Therefore, based on Green and Silverman (1993), the effective degrees of freedom can be defined as $df(\boldsymbol{\lambda}) = \text{trace}\{\hat{\boldsymbol{H}}(\boldsymbol{\lambda})\}$. Notice that $df(\lambda_1)$ corresponds to the sum of the principal diagonal elements of the matrix $(\boldsymbol{B}^{*\top} \boldsymbol{W}_{1i} \boldsymbol{B}^* + \boldsymbol{\Lambda})^{-1} \boldsymbol{B}^{*\top} \boldsymbol{W}_{1i} \boldsymbol{B}^*$ from $(p+1)$-th position to the $(p+s_1)$-th position, and so on for $df(\lambda_j)$, for $j = 1, \ldots, k$.

5.5 Estimation of smoother parameters

Also, following Manghi et al. (2019), for the selection parameter they suggest to consider the generalized cross-validation method (Craven and Wahba, 1978; Wood, 2017), which consists in selecting the $\boldsymbol{\lambda}$, such that

$$\widehat{\boldsymbol{\lambda}} = \text{argmin}_{\boldsymbol{\lambda}} \text{GCV}(\boldsymbol{\lambda}) = \text{argmin}_{\boldsymbol{\lambda}} \frac{\sum_{i=1}^n \boldsymbol{u}_i^\top \hat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{u}_i}{[1 - n^{**-1} \text{trace}\{\hat{\boldsymbol{H}}(\boldsymbol{\lambda})\}]^2},$$

where $n^{**} = \sum_{i=1}^n n_i^*$.

5.6 Obtaining standard errors

To obtain the covariance matrix, $\text{Cov}(\hat{\boldsymbol{\theta}})$, and standard errors, $\text{SE}(\hat{\boldsymbol{\theta}})$, considering the sandwich covariance estimator $\hat{\boldsymbol{J}}^{-1}$ of $\hat{\boldsymbol{\theta}}$, which can be consistently estimated (Liang and Zeger, 1986) by:

$$\hat{\boldsymbol{J}}^{-1} = \left(\sum_{i=1}^n \boldsymbol{S}_i(\hat{\boldsymbol{\theta}}) \right)^{-1} \left(\sum_{i=1}^n \boldsymbol{Q}_i^\top \hat{\boldsymbol{T}}_i \widehat{\boldsymbol{P}}_i^{-1} \hat{\boldsymbol{d}}_i^* \hat{\boldsymbol{d}}_i^{*\top} \widehat{\boldsymbol{P}}_i^{-1} \hat{\boldsymbol{T}}_i \boldsymbol{Q}_i \right) \left(\sum_{i=1}^n \boldsymbol{S}_i(\hat{\boldsymbol{\theta}}) \right)^{-1}, \tag{5.13}$$

where $\boldsymbol{S}_i(\boldsymbol{\theta})$ is defined in Equation (5.12). Another option, following Manghi et al. (2017) proposal, is to obtain the so-called naive standard errors using the covariance matrix $\boldsymbol{J}^{*-1}$, given by

$$\hat{\boldsymbol{J}}^{*-1} = \left(\sum_{i=1}^n \boldsymbol{S}_i(\hat{\boldsymbol{\theta}}) \right)^{-1} \left(\sum_{i=1}^n \boldsymbol{Q}_i^\top \hat{\boldsymbol{T}}_i \widehat{\boldsymbol{P}}_i^{-1} \hat{\boldsymbol{T}}_i \boldsymbol{Q}_i \right) \left(\sum_{i=1}^n \boldsymbol{S}_i(\hat{\boldsymbol{\theta}}) \right)^{-1}.$$

Thereby, $\text{Cov}(\hat{\boldsymbol{\theta}})$ is equal to $\hat{\boldsymbol{J}}^{-1}$ or $\hat{\boldsymbol{J}}^{*-1}$ and $\text{SE}(\hat{\boldsymbol{\theta}})$, to $\hat{\boldsymbol{J}}^{-1/2}$ or $\hat{\boldsymbol{J}}^{*-1/2}$. Then, $\text{Cov}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{J}}$ and $\text{SE}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{J}}^{-1/2}$.

5.7 Hypothesis testing

To test the significance of linear statistical hypothesis $\boldsymbol{C}\varsigma = \boldsymbol{d}$, where $\text{rank}(\boldsymbol{C}) = l$, $l \geq p$ or $l \geq q$, it can be used the Wald-type statistic (see Hardin (2005), for example), that is,

$$\xi_W = (\boldsymbol{C}\hat{\boldsymbol{\varsigma}} - \boldsymbol{d})^\top (\boldsymbol{C}\boldsymbol{V}_\varsigma \boldsymbol{C}^\top)^{-1} (\boldsymbol{C}\hat{\boldsymbol{\varsigma}} - \boldsymbol{d}),$$

where under $\xi_W \xrightarrow[n\to\infty]{D} \chi^2_l$ and, furthermore, $\boldsymbol{\varsigma}$ can be either $\boldsymbol{\beta}$, $\boldsymbol{\kappa}$, $\boldsymbol{V}_\varsigma$ is the matrix of variance-covariance regards to $\boldsymbol{\varsigma}$ extract from $\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}})$.

5.8 Model fit diagnostic tools

As said in the Sections 2.11 and 3.11, performing diagnostics analysis are important in order to check the goodness-of-fit of the estimated model and to evaluate the model assumptions. Thereby, it will be introduced for the augmented SPAM-GEE, graphical tools for detecting departures from the postulated model, influential observations and local influential analysis.

5.8.1 Residual analysis

A option of residual, used in the models for augmented independent data, is based on Silva et al. (2020) proposal, the randomized quantile residuals. Then, the residuals are defined as $r_i = \Phi(G_i), i = 1, \ldots, n$, where $\Phi(.)$ denotes the cumulative distribution function of a standard Normal distribution, G_i is a uniform random variable on the interval $(a_i, b_i]$, with $a_i = \lim_{y \uparrow y_i} F(y; \hat{\upsilon}, \hat{\varrho}, \hat{\mu}, \hat{\phi})$, $b_i = F(y_i; \hat{\upsilon}, \hat{\varrho}, \hat{\mu}, \hat{\phi})$, where $\hat{\varrho} = \hat{p}_{1i}/\hat{\upsilon}$ and $\hat{\upsilon} = \hat{p}_{0i} + \hat{p}_{1i}$.

Some tests were conducted for us to evaluate if this proposal could be used in the case of correlated data. Then, 100 replicas were generated from models and we analyzed the mean, variance, kurtosis and skewness for the 100 residuals sample. In average the quantile residuals presented a mean and a variance close to 0 and 1, respectively, and kurtosis equals to 3 and a skeness close to 0. Furthermore, from the histogram of each samples, the normality supposition seems to be reasonable for these residuals, thus pointwise envelopes could be constructed based on the standard normal distribution, as made in the independent models.

5.8.2 Local Influence

As made in the previous Section 4.11.2 it will be developed local influence tools for our approach via GEE based on the Cadigan and Farrell (2002) proposal, which is assesses the possible influential observations using the generalized the local influence instead of likelihood displacement, which is given by $\text{FD}(\boldsymbol{\omega}) = 2[\mathcal{F}(\hat{\boldsymbol{\theta}}) - \mathcal{F}(\hat{\boldsymbol{\theta}}_\omega)]$, where $\mathcal{F}$ is a fit function, assumed doubly differentiable for $\boldsymbol{\theta}$, whose estimate $\hat{\boldsymbol{\theta}}$ is the solution of $(\partial \mathcal{F}(\boldsymbol{\theta})/\boldsymbol{\theta}) = \mathbf{0}$. Venezuela et al. (2011) proposed under the Cadigan and Farrell (2002) methodology, replace the likelihood equations by the estimating equations, to study the local behaviour of FD($\boldsymbol{\omega}$) for any value of $\boldsymbol{\omega}$ in a neighborhood of $\boldsymbol{\omega}_0$, which represented the null perturbation vector, and find the eigenvector $\boldsymbol{l}_{max}$ corresponding to the largest eigenvalue of the matrix

$$\hat{\boldsymbol{B}}_G = -\boldsymbol{\Delta}\boldsymbol{S}(\boldsymbol{\theta})^{-1}\boldsymbol{\Delta}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\hat{\boldsymbol{\omega}}},$$

where $\boldsymbol{\Delta} = (\partial\boldsymbol{\Psi}_{1p}(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}^\top, \partial\boldsymbol{\Psi}_2(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}^\top)^\top$ and $\boldsymbol{S}(\boldsymbol{\theta})$ is the sensibility matrix. If the interest is in a local influence relative to a partition of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, for example $\boldsymbol{\theta}_1$, the $\boldsymbol{l}_{max}$ corresponding to the largest eigenvalue of the matrix $-\boldsymbol{\Delta}(\boldsymbol{S}^{-1}(\boldsymbol{\theta}) - \boldsymbol{S}(\boldsymbol{\theta})_c^{-1})\boldsymbol{\Delta}$ can be used as a measure for this purpose, where

$$\boldsymbol{S}(\boldsymbol{\theta})_c = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{22} \end{pmatrix}, \quad \text{where} \quad \boldsymbol{S}_{22} = \frac{\partial\boldsymbol{\Psi}_1(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_2^\top}.$$

This concept is important, because here it will be built measures of local influence for $\boldsymbol{\nu}$ $\boldsymbol{\kappa}$ and $\boldsymbol{\sigma}$ individually.

Two perturbation schemes will be developed for the models: case-weight perturbation and working correlation matrix perturbation.

Case-weight perturbation

Let $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$, where $\boldsymbol{\omega}_i = (\omega_{i1,\ldots,\omega_{in_i^*}})^\top$, $i = 1, \ldots, n$, be a perturbed vector. The GEE's under this perturbation scheme are given by

$$\boldsymbol{\Psi}^*(\boldsymbol{\nu}, \boldsymbol{\kappa}|\boldsymbol{\omega}) = \boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{T}^{-1}\text{diag}\{\boldsymbol{\omega}\}\boldsymbol{d}^* - \boldsymbol{\Lambda}_2.$$

The non perturbation vector, $\boldsymbol{\omega}_0$, assumes $\omega_{it} = 1, i = 1, \ldots, n,\ t = 1, \ldots, n_i^*$. For this scheme, the matrix $\hat{\boldsymbol{B}}_G$ is given by

$$\text{diag}\{\boldsymbol{d}^*\}\boldsymbol{T}^{-1}\boldsymbol{W}^\top\boldsymbol{Q}(\boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{Q})^{-1}\boldsymbol{Q}^\top\boldsymbol{W}\boldsymbol{T}^{-1}\text{diag}\{\boldsymbol{d}^*\}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Working correlation matrix perturbation

Let $\boldsymbol{R}(\boldsymbol{\xi})$ be a working correlation matrix characterized by a $\binom{n_i^*}{2}$-vector $\boldsymbol{\xi} = \left(\xi_{12}, \ldots, \xi_{\binom{n_i^*}{2}}\right)^\top$. As each unit has its own working correlation matrix, a possible perturbation scheme in the correlation vector $\boldsymbol{\xi}$ can be given by (Venezuela et al., 2011), $\xi_{\omega i(jl)} = \dfrac{\xi_{jl}}{\omega_{i(jl)}}$, where $i = 1, \ldots, n$, $j < l$ and $j, l = 1, \ldots, n_i^*$. For this perturbation scheme, $\boldsymbol{\omega} = (\omega_{1(12)}, \ldots, \omega_{1((n_i^*-1)n_i^*)}, \ \ldots, \omega_{n(12)}, \ldots, \omega_{n((n_i^*-1)n_i^*)})^\top$ is a perturbation vector and $\boldsymbol{\omega}_0$ is a vector with ones. The perturbation estimating equation is given by:

$$\boldsymbol{\Psi}_1(\boldsymbol{\beta}, \boldsymbol{\omega}) = \boldsymbol{X}^\top\boldsymbol{T}_1\boldsymbol{\Sigma}_\omega^{-1}\boldsymbol{u} - \boldsymbol{\Lambda}.$$

Then, each column of matrix $\boldsymbol{\Delta}$ can be expressed by:

$$\frac{\partial \boldsymbol{\Psi}_{1p}(\boldsymbol{\nu}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega}_{(jl)}^\top} = \boldsymbol{X}^\top \boldsymbol{T}_1 \frac{\partial \boldsymbol{\Sigma}_\omega^{-1}}{\partial \boldsymbol{\omega}_{(jl)}^\top} \boldsymbol{u},$$

where the derivative of $\boldsymbol{\Sigma}_\omega^{-1}$ with respect to $\boldsymbol{\omega}_{(j,l)}^\top$ is given by $\partial \boldsymbol{\Sigma}_\omega / \partial \boldsymbol{\omega}^\top = \Big(\partial \boldsymbol{\Sigma}_{\omega 1} / \partial \boldsymbol{\omega}_1^\top, \ldots, \partial \boldsymbol{\Sigma}_{\omega n} / \partial \boldsymbol{\omega}_n^\top \Big)$. The i-th diagonal block of $\boldsymbol{\Sigma}_\omega$ is $\boldsymbol{\Sigma}_{\omega i} = \sqrt{\mathrm{Var}(\boldsymbol{u}_i)} \boldsymbol{R}(\boldsymbol{\xi}_{\omega i}) \sqrt{\mathrm{Var}(\boldsymbol{u}_i)}$, with $\boldsymbol{\xi}_{\omega i} = \left(\xi_{\omega i(12)}, \ldots, \xi_{\omega i((n_i^*-1)n_i^*)} \right)$ and $i = 1, \ldots, n$. We also have that:

$$\frac{\partial \boldsymbol{\Sigma}_{\omega i}}{\partial \omega_{i(jl)}} = \sqrt{\mathrm{Var}(\boldsymbol{u}_i)} \frac{\partial \boldsymbol{R}(\boldsymbol{\xi}_{\omega i})}{\partial \omega_{i(jl)}} \sqrt{\mathrm{Var}(\boldsymbol{u}_i)},$$

where $\partial \boldsymbol{R}(\boldsymbol{\xi}_{\omega i}) / \partial \omega_{i(jl)}$ is a $(n_i^* \times n_i^*)$ symmetric matrix with null diagonal and jl and lj elements are equal to $-\xi_{jl}$, with $i = 1, \ldots, n$ and $j, l = 1, \ldots, n_i^*$.

5.9 Simulation Studies

In this section, the results of four simulation studies. Section 5.9.1 is related to the analysis of parameter recovery. In Section 5.9.2, we perform a correlation matrix misspecification study. In Section 5.9.3 is present a link function misspecification study. The results for the first study is in Appendix, whereas the other studies will be available at https://github.com/aureaflg/Simulations-study.git, where we made reproducible codes for one can view the graphs and tables.

Some observations about the estimation process for the next subsection are necessary to be done. First, the way that data has been generate for each scenario is using t-copulas based on Student-t distribution considering a degree of freedom equal to 3. This generator process only approximates the correlation structure, therefore it is possible to obtain bias estimates for parameters related to the correlation matrix. Also, the way that the data is being generated considers a correlation between all subjects, hence the response vector whether it comes from (0,1) interval or comes from $\{0, 1\}$, will have a dependence structure. However, our proposal considering only a correlation matrix for data coming from the $(0, 1)$ interval. Then, to estimate the correlation parameters using simulated data, we considered the vector $\boldsymbol{u}_i^* = (1 - \boldsymbol{z}_i^*)^{-1} \boldsymbol{u}_i$ for estimator presented in the Section 4.5, which means, to estimate the correlation parameters using all the observations regardless of where it was generated.

Furthermore, for the ZOABR-SPAM-GEE, the estimation of parameter α presented some problems, then we put a restriction in the interval were the estimates could be obtained based on the initial values α_0. The interval where it could be estimate was setting as $[max(\alpha_0 - 0.3, 0.01); max(\alpha_0 + 0.3, 0.99)]$. The value 0.3 was chosen arbitrarily and tests was done so that the algorithm could run.

5.9.1 Study 1: Parameters recovery

Here we considered some scenarios of interest defined by the combination of the levels of some factors of interest. The factors (with the respective levels within parenthesis) are number of subjects and their repetition number (10 subjects and 5 times, 10 subjects and 10 times, 50 subjects and 10 times) which implies in a sample (n) (50,100,500), regression model (ZOAB, ZOAS, ZOABR), link function (probit, logit, cauchit, cloglog, loglog), modeled parameters (mean and dispersion) and considering the correlation matrix as exchangeable. For each scenario, $R = 100$ replicates of Monte Carlo were generated. Each scenario are described below.

Table 21 – Scenarios of simulation study 1

Regression model (modelling mean and dispersion)	n	Link functions
ZOAB$(\mu_{it}, \phi_{it}, p_{0it}, p_{1it})$	50 100 500	probit, logit, cauchit, cloglog and loglog.
ZOAS$(\mu_{it}, \phi_{it}, p_{0it}, p_{1it})$	50 100 500	probit, logit, cauchit, cloglog and loglog.
ZOABR$(\mu_{it}, \phi_{it}, p_{0it}, p_{1it}, \alpha)$	50 100 500	probit, logit, cauchit, cloglog and loglog.

Furthermore, the considered models for related scenarios presented, are given by $\log(p_{0it}/(1 - p_{0it} - p_{1it})) = \rho_0 + \rho_1 F_{it}$, $\log(p_{1it}/(1 - p_{0it} - p_{1it})) = \tau_0 + \tau_1 M_{it}$, $g(\mu_{it}) = \beta_1 X_{it} + cos(Z_{it})$, $\log(\phi_{it}) = \kappa_0 + \kappa_1 E_{it}$, for $i = 1, \ldots, n, t = 1, \ldots, n_i^*$ $i = 1, \ldots, n^{**}$, where $X_{it} \overset{iid}{\sim} \text{uniform}(0, 1)$, $E_{it} \overset{iid}{\sim} \text{uniform}(0, 1)$, $F_{it} \overset{iid}{\sim} \text{uniform}(0, 1)$, $M_{it} \overset{iid}{\sim} \text{uniform}(0, 1)$ and $Z_{it} \overset{iid}{\sim} \text{uniform}(0, 3\pi)$, $g(.)$ is either the probit, logit, cauchit, cloglog.

The actual parameter values considered loglog link were $\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$, whereas for the other link were: $\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$. This change to the loglog link function is justified by the fact that, depending on the parameters chosen, generating from the models using this link function can generate values very close to one or zero, which could compromise the fit of the models to the simulated data even for the BR-SPAM-GEE model.

For all models, when $n = 50$, 40 knots were used and the smoothing parameter was fixed at $\lambda = 500$, whereas, when $n = 100$ or $n = 500$, 50 knots were used and the smoothing parameter was fixed at $\lambda = 500$. These values lead to a smooth fitted curve for non-parametric component. As the correlation matrix considered was the exchangeable, the only correlation parameter ξ was fixed as 0.8.

The related results are presented in Appendix B.3, where Figures 79 to 83 are related to ZOAB-SPAM-GEE, Figures 84 to 88 are related to ZOAS-SPAM, and, Figures 89 to 93 concern to ZOABR-SPAM.

In a general way, for all models the regression parameters were properly recovered, as the sample size increases, the estimates become more accurate. The estimated curves are for sample size $n = 50$ representing the behavior of the curves, in general well, but it presents great variability, in contrast to $n = 100$ and $n = 500$ where there is less variability around the real curve. Specifically for the ZOABR-SPAM-GEE the parameter α under all link functions the recovery was not good enough, the estimates were biased. Furthermore, for all models the correlation parameter ξ was under estimated and for ZOAS-SPAM-GEE present the most biased recovery. A possible reason, as said before, for this is happening is that the copulas used generate a dependency structure that are an approximation of the true correlation matrix, which can generate bias in the adjustment.

5.9.2 Study 2: Correlation matrix misspecification

In this section we considered only the scenarios varying the correlation matrix (exchangeable,AR-1,unstructured), the regression models (ZOAB,ZOAS,ZOABR) and the parameter related with correlation ξ (0.8,0.3). We setting the sample size in 100 and chose the link function as cloglog with R=1 replica. The same model defined in were considered, except by the parameter ξ, which is either 0.8 and 0.3.

Although the estimates were, in general, properly recovered under the three correlation structure, it was possible to notice slightly differences among fitted parameters, and in some cases AR-1 fitted better then the model under exchangeable. The unstructured had the same or slightly worse performance of exchangeable. It can be concluded that the recovery is reasonable even under misspecification.

5.9.3 Study 3: Link function misspecification

In this section, we considered only the scenarios varying the regression models (ZOAB, ZOAS, ZOABR) and the link functions (probit, logit, cauchit, cloglog,loglog). We set the sample size in 500 (50 subjects and 10 repeated measure), modeling the mean and the dispersion. We generate only one replica from each model, setting a distribution and a link function for it, and fit the simulated data using a model with the same distribution but varying the link function, using all the different ones from the one used to generate the simulated data. In Table 22 we present the 60 scenarios of interest.

The considered linear predictors and the link functions for related scenarios presented, are given by $\log(p_{0it}/(1-p_{0it}-p_{1it})) = \rho_0 + \rho_1 F_{it}$, $\log(p_{1it}/(1-p_{0it}-p_{1it})) = \tau_0 + \tau_1 M_{it}$, $g(\mu_{it}) = \beta_1 X_{it} + cos(Z_{it})$, $\log(\phi_{it}) = \kappa_0 + \kappa_1 E_{it}$, for $i = 1, \ldots, n,\ t = 1, \ldots, n_i^*$, where $X_{it} \overset{iid}{\sim}$ uniform$(0, 1)$, $E_{it} \overset{iid}{\sim}$ uniform$(0, 1)$, $Z_{it} \overset{iid}{\sim}$ uniform$(0, 3\pi)$, $F_{it} \overset{iid}{\sim}$ uniform$(0, 1)$

Table 22 – Scenarios of simulation study 2

Distribution that will generate and fit the simulated data	Actual link function	Link function considered in fitted model
ZOAB$(\mu_{it}, \phi_{it}, p_{0it}, p_{1it})$, ZOAS$(\mu_{it}, \phi_{it}, p_{0it}, p_{1it})$ or ZOABR$(\mu_{it}, \phi_{it}, p_{0it}, p_{1it}, \alpha)$	logit	probit, cauchit, cloglog and loglog
	probit	logit, cauchit, cloglog and loglog
	cauchit	logit, probit, cloglog and loglog
	cloglog	logit, probit, cauchit and loglog
	loglog	logit, probit, cloglog and cauchit

and $M_{it} \stackrel{iid}{\sim}$ uniform$(0, 1)$, $g(.)$ is either the probit, logit, cauchit, cloglog.

The actual parameter values considered loglog link were $\beta_1 = 1$, $\kappa_0 = 1$, $\kappa_1 = 2$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$, whereas for the other cases were $\beta_1 = -1$, $\kappa_0 = 2$, $\kappa_1 = -3$, $\alpha = 0.7$, $\rho_0 = \tau_0 = -1.8$ and $\rho_1 = \tau_1 = 1.5$. This change for actual values of parameters under loglog link is justified for the same reason discussed in Section 5.9.1. We considered both the comparison of the observed and predicted values as well as the parameter recovery. The correlation structure used were exchangeable, with $\xi = 0.8$.

The all models presented bad recovery of real mean when the link function that generating was symmetric and the adjuster model used loglog and cloglog as link function, (both are asymmetric link). The recovery of regression parameters related to the mean and to the dispersion/precision, they were all well estimated, if there is a bias, it is small.

5.10 Real data analysis

The data used to illustrate the augmented and correlated models is the same used for Chapter 3 (Section 3.14), a psychometric study of risk perception obtained by Carlstrom et al. (2000). As described there, we will use to analysis only the risk perception, which comes from the subjective questionnaire.

In this section, we selected 3 of 22 activities, which results in a total of 588 subjects and each of them are evaluated in different frequencies generating repeated measurements, which will generated a large sample size. Then, to circumvent this problem, we selected a random sample of 200 subjects out of a total 588, which gave us a sample size (subjects $\times$ their repetitions) of 577. As it is an unbalanced study, i.e., the number of repeated measure, says $n_i^*, i = 1, \ldots, 200$, for each subject may be different. The n_i^* for this data only assumes 1, 2 or 3.

The risk perception of each subject is defined in the interval [0,100], such that the higher the value, the higher the risk. Hence, the response will be the risk perception (Risk) about the 3 risk related with health (tested for gene that predisposes to breast cancer, (HRT), tested for gene that predisposes to heart disease, (BRCA), ride bicycle one mile each day in an urban area, (BIKE)) transformed to the interval [0,1]. Furthermore, as the covariates we consider: the gender (Gender), ethnic background (Ethnic), age (Age) and the worldview obtained from a related questionnaire (Wvcat). The levels of each factor is presented in Table 23.

Table 23 – Levels of factors

Variables	Levels
Ethnic	Caucasian (1); African-American (2) Mexican-American (3); Taiwanese-American (4)
Wvcat	unclassifiable (0); individualist (1) hierarchicalist (2); egalitarian (3)
Gender	female (0); male (1)

Figure 27 indicates that the covariates Wvcat and Ethnic could influence the responses variability, whereas Gender does not seem to influence variability. Also, this Figure shows the frequency plot of Risk and plots related with covariate Age, which will be treated by the non-parametric term. Considering, now Figure 18 and the Tables 24, 25 and 26 related with the discrete part, i.e. Risk = 0 and Risk = 1, it is possible to note that the levels of variable Wvcat is the one that seems more different for Risk = 0, which may indicate that this variable is more significant for the response, whereas for Risk = 1 the covariates do not seem different within the categories. Figure 28 related to Age, presents different median and dispersion for Risk=0 and Risk=1, while for the Risk equals to zero it is seen a positive skewness, for equals 1 it is seen a negative skewness, which may indicates that this variable may influence the response.

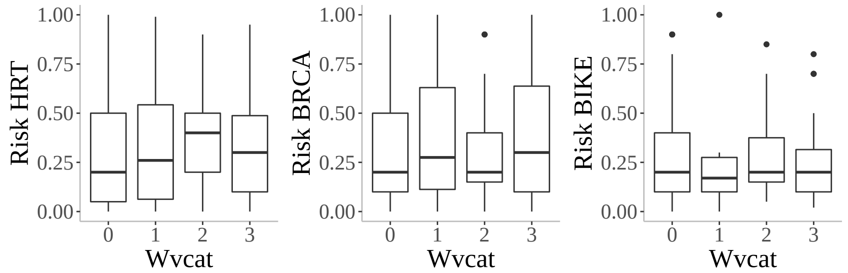
The choice of link function and the correlation structure, was made by comparison of diagnostic tools proposed. The better choices were: for the three models the logit for mean and the sqrt for dispersion/precision, also the exchangeable seems to be reasonable given that it is a multivariate study and the responses are not directly related. The initial model fitted for ZOAB-SPAM-GEE, ZOAS-SPAM-GEE and ZOABR-SPAM-GEE was defined including all covariates in all linear predictors, both in the discrete part and in the

Table 24 – Number of observation by the levels of Ethnic variable

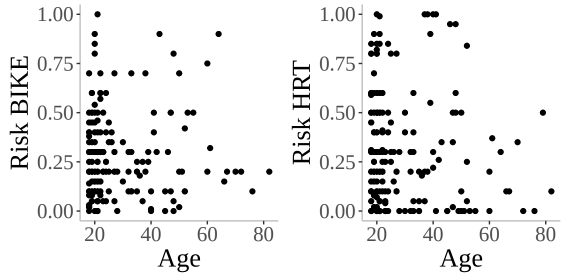
	Risk		
Ethnic	0	1	Total
1	12	0	12
2	25	9	34
3	8	4	12
4	21	0	21
Total	66	13	79

Table 25 – Number of observation by the levels of Gender variable

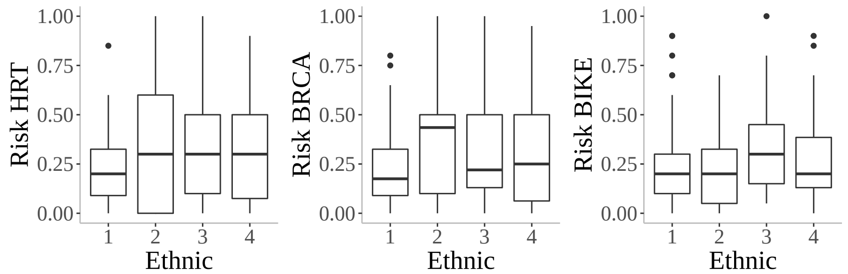
	Risk		
Gender	0	1	Total
0	30	7	37
1	36	6	42
Total	66	13	79



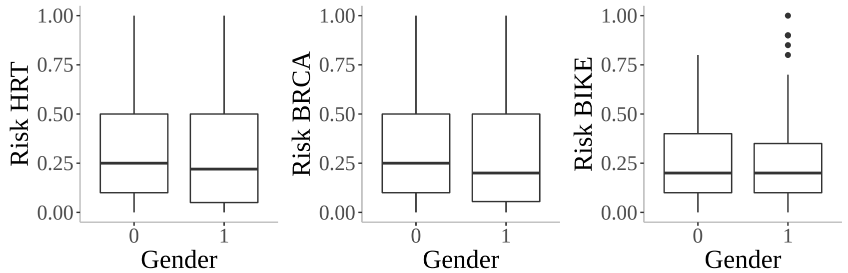
(a) Boxplot of Risk for each response by categories of Wvcat covariate.



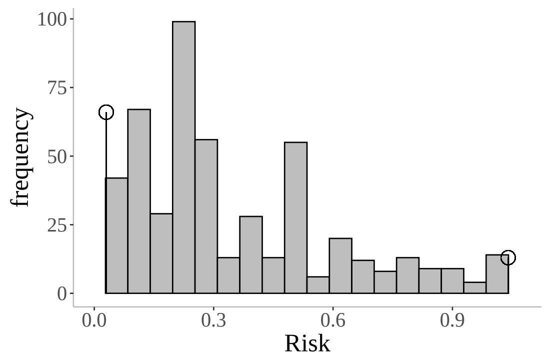
(b) Scatter plot of Risk versus Age.



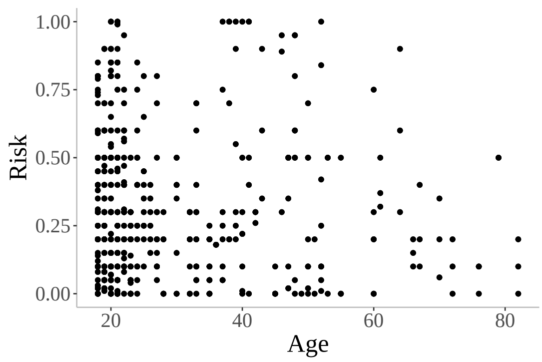
(c) Boxplot of Risk for each response by categories of Ethnic covariate.



(d) Boxplot of Risk for each response by categories of Gender covariate.



(e) Histogram of Risk by the frequency.



(f) Risk for all response variable together.

Figure 27 – Explanatory analysis plots of risk perception data set.

Table 26 – Number of observation by the levels of Wvcat variable

	Risk		
Wvcat	0	1	Total
0	51	10	61
1	3	2	5
2	6	0	6
3	6	1	7
Total	66	13	79

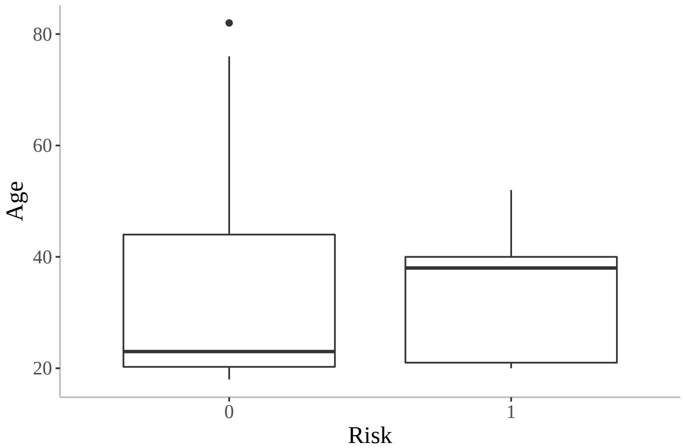


Figure 28 – Boxplots for Age by Risk=0 and Risk=1.

continuous part, as shown below:

$$\begin{aligned}
Y_{itljk} &\sim \text{ZOAB}(\mu_{itljk}, \phi_{itljk}, p_{0itljk}, p_{1itljk}) \\
&\quad \text{or } \ \text{ZOABR}(\mu_{itljk}, \phi_{itljk}, p_{0itljk}, p_{1itljk}, \alpha) \\
&\quad \text{or ZOAS}(\mu_{itljk}, \phi_{itljk}, p_{0itljk}, p_{1itljk}), \\
g_1(\mu_{iljk}) &= \beta_0 + (\beta_1)_l + (\beta_2)_j + (\beta_3)_k + h_1(z_{it}), \\
g_2(\phi_{iljk}) &= \kappa_0 + (\kappa_1)_l + (\kappa_2)_j + (\kappa_3)_k + \kappa_4 z_{it}, \\
\log(p_{0iljk}/(1 - p_{0iljk} - p_{1iljk})) &= \rho_0 + (\rho_1)_l + (\rho_2)_j + (\rho_3)_k + \rho_4 z_{it} \\
\log(p_{1iljk}/(1 - p_{0iljk} - p_{1iljk})) &= \tau_0 + (\tau_1)_l + (\tau_2)_j + (\tau_3)_k + \tau_4 z_{it},
\end{aligned}$$

where, $i = 1, \ldots, 577$, $t = 1, \ldots, n_i^*$, $l = 1, 2, 3, 4$, $j = 0, 1, 2, 3$, $k = 0, 1$, $(\beta_1)_1 = (\beta_2)_0 = (\beta_3)_0 = 0$, $(\kappa_1)_1 = (\kappa_2)_0 = (\kappa_3)_0 = 0$, $(\rho_1)_1 = (\rho_2)_0 = (\rho_3)_0 = 0$, $(\tau_1)_1 = (\tau_2)_0 = (\tau_3)_0 = 0$. The parameters (β_1), (κ_1), (ρ_1) and (τ_1) are related with Ethnic. The parameters (β_2), (κ_2), (ρ_2) and (τ_2) are related with Wvcat. The (β_3), (κ_3), (ρ_3) and (τ_3) are related with Gender. At last, the parameters κ_4, ρ_4 and τ_4 are associated with the variable $\boldsymbol{z}$ Age standard.

After we fitted the models, it was selected only the covariates that were under a significance level of 10% for each model and excluded all non significant covariates, combining the equivalent levels within each significant covariate (according to the non-significance of the respective parameters). The selected variables for each model are presented below. The first model proposed is for ZOAB-SPAM-GEE:

$$\begin{aligned} Y_{itl} &\sim \text{ZOAB}(\mu_{itl}, \phi_{itl}, p_{0itj}, p_1), \\ g_1(\mu_{itl}) &= \beta_0 + (\beta_1)_l + h_1(z_{it}), \\ g_2(\phi_{itlj}) &= \kappa_0 + (\kappa_1)_l + (\kappa_2)_j, \\ \log(p_{0itj}/(1 - p_{0itj} - p_1)) &= \rho_0 + (\rho_2)_j + \rho_4 z_{it}, \end{aligned}$$

where $l = 1, 2, 3, 4$, $j = 0, 1, 2, 3$, $(\rho_2)_0 = (\rho_2)_1 = (\rho_2)_2 = 0$ and $(\beta_1)_1 = 0$ and $(\kappa_1)_1 = (\kappa_2)_0 = (\kappa_2)_2 = 0$.

The second model proposed is for ZOAS-SPAM-GEE:

$$\begin{aligned} Y_{iljk} &\sim \text{ZOAS}(\mu_{itljk}, \phi_{itjk}, p_{0it}, p_1), \\ g_1(\mu_{itljk}) &= \beta_0 + (\beta_1)_l + (\beta_2)_j + (\beta_3)_k + h_1(z_{it}), \\ g_2(\phi_{itjk}) &= \kappa_0 + (\kappa_2)_j + (\kappa_3)_k + \kappa_4 z_i, \\ \log(p_{0it}/(1 - p_{0it} - p_1)) &= \rho_0 + \rho_4 z_i \end{aligned}$$

where $l = 1, 2, 3, 4$, $j = 0, 1, 2, 3$, $(\beta_1)_1 = (\beta_2)_0 = (\beta_3)_0 = 0$ and $(\kappa_2)_0 = (\kappa_3)_0 = 0$.

At last, the third proposed model is for ZOABR-SPAM-GEE as follows:

$$\begin{aligned} Y_{itlk} &\sim \text{ZOABR}(\mu_{itl}, \phi_{itk}, p_{0it}, p_1, \alpha), \\ g_1(\mu_{itl}) &= \beta_0 + (\beta_1)_l + h_1(z_{it}), \\ g_2(\phi_{itl}) &= \kappa_0 + (\kappa_3)_k, \\ \log(p_{0it}/(1 - p_{0it} - p_1)) &= \rho_0 + \rho_4 z_{it} \end{aligned}$$

where $l = 1, 2, 3, 4$, $k = 1, 2$, $(\beta_1)_1 = (\kappa_3)_1 = 0$.

The QQ plot with envelope for the quantile residuals related to the three models, indicates a not good model fit (see Figure 29), for ZOAS-SPAM-GEE neither for ZOABR-SPAM-GEE. The best fit was the ZOAB-SPAM-GEE, hence was the model chosen to do the fit, because seems the one more appropriate for this the data set. Figure 30 shows the fitted curve for non-parametric part. The effective degrees of freedom were 4.2198 for ZOAB-SPAM-GEE, using 50 knots and for a smooth parameter $\lambda = 500$, which was chosen by minimizing the AIC. From Figure 31, it is possible to see that the residuals are well within the the confidence bands of the QQ plot, even though there is some no random pattern. The other residuals plots show no tendency as well the normality seems to be reasonable according to the histogram. And from Figure 32, we can see the QQ plot for each repeated risk (HRT,BRCA and BIKE).

Furthermore, Figure 33 presents the the local influence analysis under case-weight perturbation related to $\boldsymbol{\mu}$, $\boldsymbol{\phi}$ and $\boldsymbol{p}_0$ (Figures 33a, 33b and 33c, respectively) and

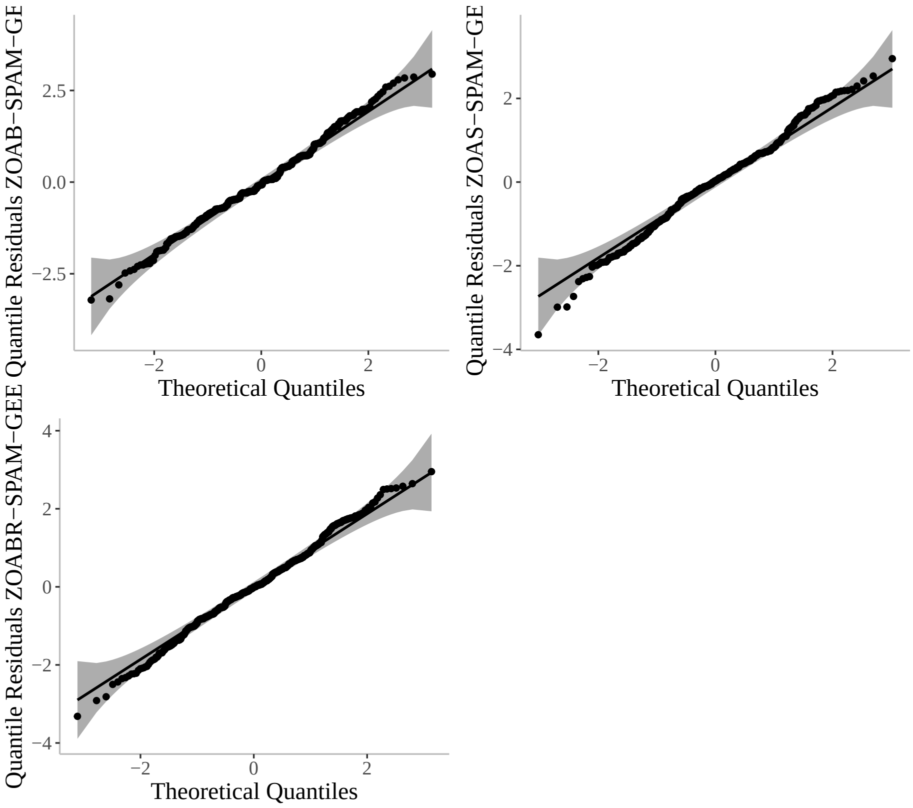
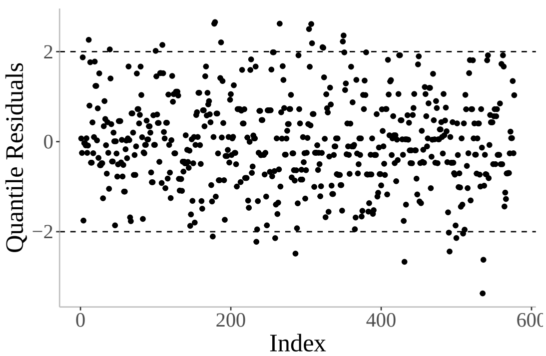


Figure 29 – Quantile-quantile plot with 99% envelopes for all models.

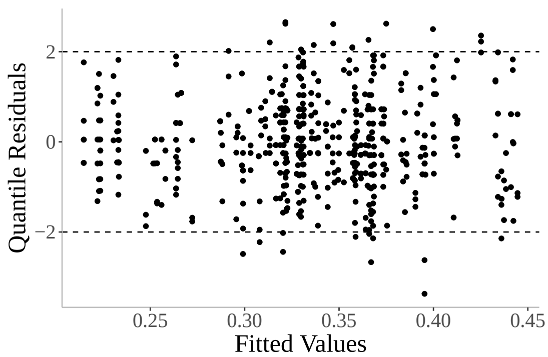
Table 27 – Inferential results for the ZOABR-SPAM-GEE.

Parameters	Estimate	SE	p-value	$IC_{95\%}$
ρ_0	-1.96	0.14	<0.0001	[-2.24;-1.69]
$(\rho_2)_3$	-0.79	0.45	0.0838	[-1.68;0.10]
ρ_4	0.30	0.12	0.0097	[0.07;0.53]
p_1	0.022	0.28	-	[-0.53;0.57]
β_0	-1.07	0.13	<0.0001	[-1.33;-0.81]
$(\beta_1)_2$	0.44	0.24	0.0691	[-0.03;0.92]
$(\beta_1)_3$	0.41	0.20	0.0401	[0.02;0.81]
$(\beta_1)_4$	0.45	0.19	0.0208	[0.07;0.82]
κ_0	2.18	0.12	<0.0001	[1.95;2.41]
$(\kappa_1)_2$	-0.38	0.15	0.0118	[-0.67;-0.08]
$(\kappa_1)_3$	-0.28	0.15	0.0601	[-0.57;0.01]
$(\kappa_1)_4$	-0.26	0.13	0.0472	[-0.52;0]
$(\kappa_2)_1$	-0.45	0.17	0.0083	[-0.79;-0.12]
$(\kappa_2)_3$	-0.30	0.15	0.0537	[-0.60;0]
Correlation parameter ξ	0.39	-	-	-

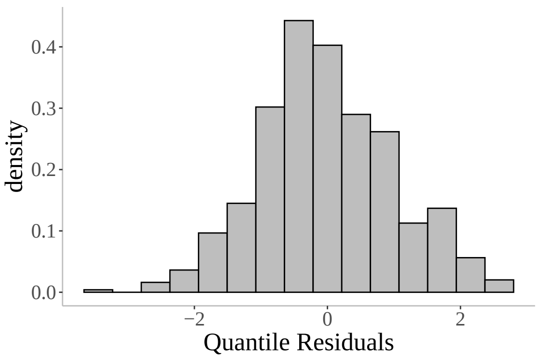
Figure 30 – Estimated curve for ZOAB-SPAM-GEE versus standardized Age.



(a) Index plot of quantile residuals.



(b) Fitted values versus quantile residuals.



(c) Histogram of quantile residuals.

Figure 31 – Results of fitted ZOAB-SPAM-GEE.

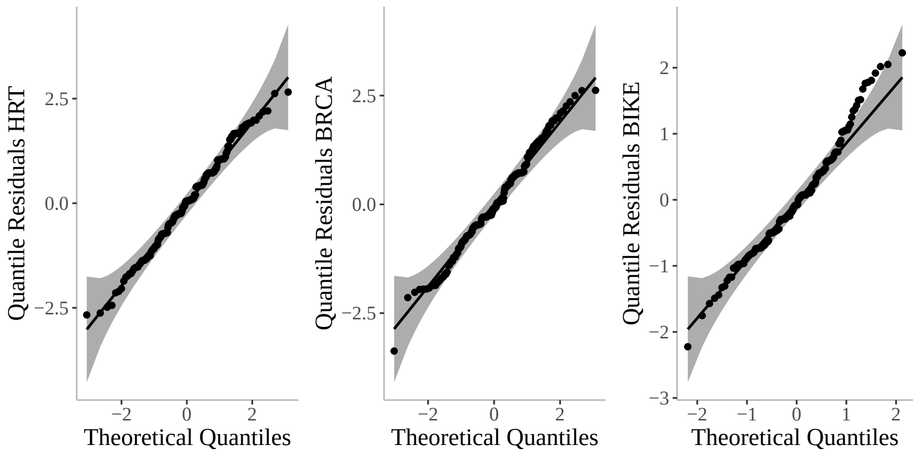
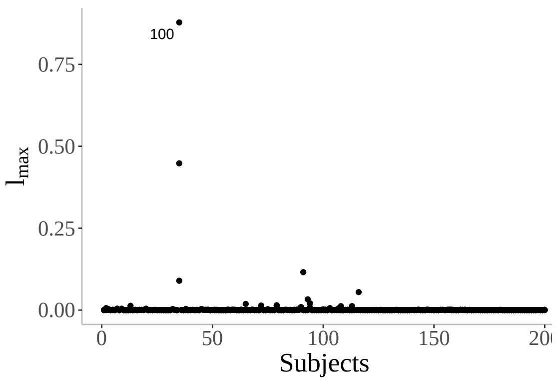
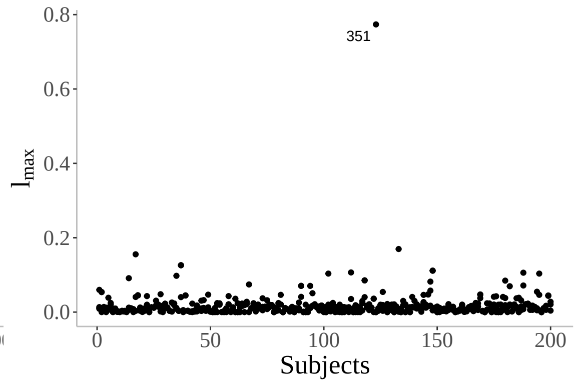


Figure 32 – Quantile-quantile plot with 99% envelopes for each type of Risk (HRT, BRCA, XRAY).

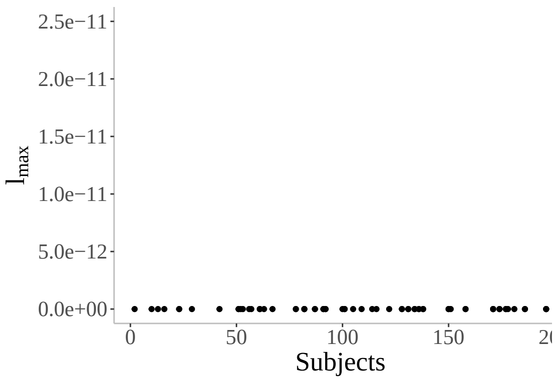
the correlation perturbation scheme (Figure 33d). From these Figures, we see that the observations (with the respective scheme and parameter) #100 (Case-weight for $\boldsymbol{\mu}$),#351 (Case-weight for $\boldsymbol{\phi}$), were and flagged, whereas for the plot related to $\boldsymbol{p}_0$ correlation perturbation scheme no observations were flagged.



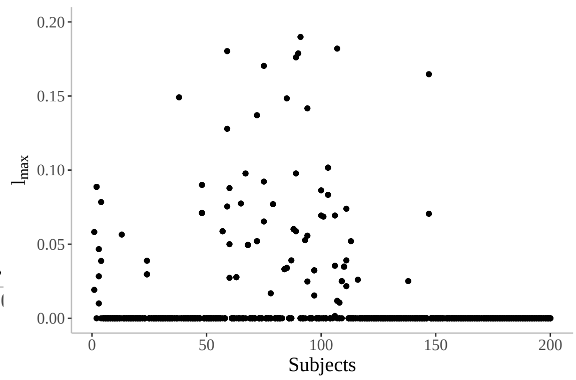
(a) Case-weight perturbation scheme related to $\boldsymbol{\mu}$.



(b) Case-weight perturbation scheme related to $\boldsymbol{\phi}$.



(c) Case-weight perturbation scheme related to $\boldsymbol{p}_0$.



(d) Correlation perturbation scheme.

Figure 33 – Index plot of $\boldsymbol{l}_{max}$ under perturbation schemes for ZOAB-SPAM-GEE.

After removing each flagged observation one at a time, the significance of the parameters are not modified. Also, the behavior of the residuals and QQ plot remained the same. In conclusion, once that the exclusion of the flagged observations did not lead to significant changes on the overall model fit, no further changes are necessary. From the final model, it possible to notice that the Ethinic and the Wvcat are significant covariates attributed to the heterogeneous dispersion. Also, the covariates Age and Wvcat influence significantly in the probability of zero occurrence. From the estimated non-parametric curve, it is possible to see a decay in the Risk as the Age decreasing.

6 Conclusion

In this dissertation we developed Semi-parametric Regression Models for limited (augmented) data ((0,1) and [0,1]) with independent or correlated response variables. We rely on the classical beta and simplex distributions, and on the beta rectangular distribution that is characterized as more robust by having heavier tails than the first two. We developed estimation methods under the frequentist approach, for independent data using the maximum penalized log-likelihood, and for correlated data using Generalized Estimation Equations. We propose, for data lies on (0,1), modeling the mean and dispersion/precision using semi-parametric linear predictors, and for data in [0,1], we additionally proposed to model the probabilities of 0 or 1.

To validate the assumptions of the regression model proposed we develop tools for global and local influence diagnostic analysis. We propose quantile residuals. We present information criteria based on likelihood (for independent data) and quasi-likelihood (for correlated data). Generalized leverage and Cook distance measures were obtained. In addition, local influence perturbation schemes were developed for the cases: case-weight perturbation, continuous covariate perturbation, response variable perturbation, and working correlation matrix perturbation (for GEE).

We conducted simulation studies to evaluate parameter recovery, residual analysis, and misspecification of the distribution and link functions. We fit the models proposed here to real data sets, showing the advantages of using them.

6.1 Future works

As a focus for future work we suggest the following research topics:

1. Propose the use of these models with parametric link functions such as Aranda-Ordaz and Skew-normal, for example.

2. Propose more robust models than those presented, for example, regression models for the median based on the Kumaraswamy distribution.

3. Propose the methodology used here under the Bayesian approach.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723. 42

Altun, E. (2019). The log-weighted exponential regression model: alternative to the beta regression model. *Communications in Statistics-Theory and Methods*, pages 1–16. 14

Altun, E. and Cordeiro, G. M. (2020). The unit-improved second-degree lindley distribution: inference and regression modeling. *Computational Statistics*, 35(1):259–279. 14

Altun, E., El-Morshedy, M., and Eliwa, M. (2021). A new regression model for bounded response variable: An alternative to the beta and unit-lindley regression models. *Plos one*, 16(1). 14

Artes, R.; Botter, D. A. (2005). *Funções de Estimação em Modelos de Regressão*. São Paulo: Insper Instituto de Ensino e Pesquisa - IME - USP. 101

Atkinson, P. (1985). Advances in medical social science, vol. i (book). *Sociology of Health & Illness*, 7(1):132–133. 14

Barndorff-Nielsen, O. and Jørgensen, B. (1991). Some parametric models on the simplex. *Journal of Multivariate Analysis*, 39(1):106–116. 14, 21

Barreto-Souza, W., Mayrink, V. D., and Simas, A. B. (2020). Bessel regression model: Robustness to analyze bounded data. *arXiv preprint arXiv:2003.05157*. 14

Bayer, F. M. and Cribari-Neto, F. (2017). Model selection criteria in beta regression with varying dispersion. *Communications in Statistics-Simulation and Computation*, 46(1):729–746. 15

Bayes, C. and Bazán, J. (2014). An em algorithm for beta-rectangular regression models. *Personal Communication*. 19

Bayes, C., Bazán, J., and Garcia, C. (2012). A new robust regression model for proportions. *Bayesian Analysis*, 7:841 – 866. 14, 19

Bonat, W. H., Petterle, R. R., Hinde, J., and Demétrio, C. G. (2018). Flexible quasi-beta regression models for continuous bounded data. *Statistical Modelling*. 22, 109, 134

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391):580–598. 40, 79

Buja, A., Hastie, T., Tibshirani, R., et al. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17(2):453–510. 25

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208. 41, 55, 86, 109, 135

Cadigan, N. and Farrell, P. (2002). Generalized local influence with applications to fish stock cohort analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(4):469–483. 112, 113, 137

Canterle, D. R. and Bayer, F. M. (2019). Variable dispersion beta regressions with parametric link functions. *Statistical Papers*, 60(5):1541–1567. 16

Carlstrom, L. K., Woodward, J. A., and Palmer, C. G. (2000). Evaluating the simplified conjoint expected risk model: Comparing the use of objective and subjective information. *Risk Analysis*, 20(3):385–392. 91, 142

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 133–169. 49, 83

Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall. 47

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik*, 31(4):377–403. 110, 136

Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika*, 74(3):591–597. 101, 103, 104, 105, 106, 127, 128, 129, 131

Cunha, M. C. C. (2003). Métodos numéricos, editora unicamp. *Campinas-SP, 2ª ed., 280p.* 27

da Paz, R. F., Balakrishnan, N., Bazán, J. L., et al. (2019). L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. *Brazilian Journal of Probability and Statistics*, 33(3):455–479. 15

De Boor, C. (1978). *A Practical Guide to Spline*, volume 27. Applied Mathematical Sciences, New York: Springer, 1978. 27

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22. 32, 68

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford University Press. 100

Duan, X.-D., Zhao, Y.-Y., and Tang, A.-M. (2019). A semiparametric bayesian approach to simplex regression model with heterogeneous dispersion. *Communications in Statistics-Simulation and Computation*, 48(8):2487–2500. 16

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244. 83

Eilers, P. and Marx, B. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2:637 – 653. 29

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102. 28, 42, 80, 190

Eilers, P. H., Marx, B. D., and Durbán, M. (2015). Twenty years of p-splines. *SORT: statistics and operations research transactions*, 39(2):0149–186. 29

Espinheira, P. L., Ferrari, S. L., and Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, 35(4):407–419. 14, 59

Espinheira, P. L. and Silva, A. d. O. (2019). Residual and influence analysis to a general class of simplex regression. *Test*, pages 1–30. 15

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815. 14, 18

Ferrari, S. L., Espinheira, P. L., and Cribari-Neto, F. (2011). Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica*, 65(3):337–351. 14, 59

Freitas, J. V. B. d., Nobre, J. S., Espinheira, P. L., and Rêgo, L. C. (2021). Marginal regression models to analyzing bounded correlated data with varying precision/dispersion. *Submitted to TEST*, 60(4). 15, 103, 107, 112, 113, 133

Galvis, D. M. (2014). *Bayesian analysis of regression models for proportional data in the presence of zeros and ones.* doutorado, Universidade Estadual de Campinas, Instituto de Matemática Estatística e Computação Científica. 24

Galvis, D. M., Bandyopadhyay, D., and Lachos, V. H. (2014). Augmented mixed beta regression models for periodontal proportion data. *Statistics in medicine*, 33(21):3759–3771. 15

Gay, D. M. (1990). Usage summary for selected optimization routines. *Computing science technical report*, 153:1–21. 41, 55, 86

Godambe, V. P. (1991). *Estimating functions.* Oxford University Press. 100

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique*, pages 245–259. 28

Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall/CRC. 40, 110, 136, 189

Green, P. J. and Yandell, B. S. (1985). Semi-parametric generalized linear models. In *Generalized linear models*, pages 44–55. Springer. 16

Guisan, A., Edwards Jr, T. C., and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, 157(2-3):89–100. 25

Hahn, E. (2008). Mixture densities for project management activity times: A robust approach to pert. *European Journal of Operational Research*, 188:450–459. 18, 19

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195. 42

Hardin, J. W. (2005). Generalized estimating equations (gee). *Encyclopedia of statistics in behavioral science*. 111, 136

Hastie, T. and Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386. 25

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman & Hall. 16, 25, 26, 41, 42, 80, 189

Hernando Vanegas, L. and Paula, G. A. (2016). An extension of log-symmetric regression models: R codes and applications. *Journal of Statistical Computation and Simulation*, 86(9):1709–1735. 16

Hossain, A., Rigby, R., Stasinopoulos, M., and Enea, M. (2016). Centile estimation for a proportion response variable. *Statistics in medicine*, 35(6):895–904. 15

Ibacache-Pulgar, G. and Figueroa-Zuniga, J. (2019). Semiparametric additive beta regression models: Inference and local influence diagnostics. In *Preprint Revstat*. https://www.ine.pt/revstat/pdf/SEMIPARAMETRICADDITIVEBETAREGRESSION.pdf. 16

Ibacache-Pulgar, G. and Reyes, S. (2017). Local influence for elliptical partially varying-coefficient model. *Statistical Modelling*, 18(2):149–174. 41, 79

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 49(2):127–145. 14

Jorgensen, B. (1997). *The theory of dispersion models*. CRC Press. 22

Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31(1):93–114. 104, 108, 109, 127, 134

Jørgensen, B. and Labouriau, R. (1995). *Exponential families and theoretical inference*. PhD thesis, University of Aarhus. 102

Koenker, R. and Yoon, J. (2009). Parametric links for binary choice models: A fisherian–bayesian colloquy. *Journal of Econometrics*, 152(2):120–130. 54

Kong, M., Xu, S., Levy, S. M., and Datta, S. (2015). Gee type inference for clustered zero-inflated negative binomial regression with application to dental caries. *Computational statistics & data analysis*, 85:54–66. 132

Kooperberg, C. and Stone, C. J. (1991). A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12(3):327–347. 26

Lehmann, E. L. and Casella, G. (2002). *Theory of point estimation*. Springer Science & Business Media. 187

Lemonte, A. and Bazán, J. (2015). New class of johnson sb distributions and its associated regression model for rates and proportions. *Biometrical Journal*, 58:n/a–n/a. 14

Lemonte, A. and Moreno-Arenas, G. (2019). On a heavy-tailed parametric quantile regression model for limited range response variables. *Computational Statistics*. 15

Lian, H., Liang, H., and Wang, L. (2014). Generalized additive partial linear models for clustered data with diverging number of covariates using gee. *Statistica Sinica*, 24:173–196. 16

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22. 100, 102, 103, 105, 107, 109, 111, 126, 129, 130, 132, 136

Linton, O. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika*, 83(3):529–540. 25

Liu, P., Yuen, K. C., Wu, L.-C., Tian, G.-L., and Li, T. (2020). Zero-one-inflated simplex regression models for the analysis of continuous proportion data. *Statistics and Its Interface*, 13(2):193–208. 15

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. 43

Manghi, R. F., Cysneiros, F. J. A., and Paula, G. A. (2017). On generalized additive partial linear models for correlated data. In *Proceedings of the 32nd International Workshop on Statistical Modelling, Volume II*, page 135–138. 16, 111, 136

Manghi, R. F., Cysneiros, F. J. A., and Paula, G. A. (2019). Generalized additive partial linear models for analyzing correlated data. *Computational Statistics & Data Analysis*, 129:47 – 60. 16, 107, 110, 135, 136

Mazucheli, J., Menezes, A. F. B., and Chakraborty, S. (2019). On the one parameter unit-lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, 46(4):700–714. 14

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press. 54

Meilijson, I. (1989). A fast improvement to the em algorithm on its own terms. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(1):127–138. 44

Meyers, S. M., Ambler, J. S., Tan, M., Werner, J. C., and Huang, S. S. (1992). Variation of perfluoropropane disappearance after vitrectomy. *Retina (Philadelphia, Pa.)*, 12(4):359–363. 118

Mitnik, P. A. and Baek, S. (2013). The kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54(1):177–192. 15

Mousa, A., El-Sheikh, A., and Abdel-Fattah, M. (2016). A gamma regression for bounded continuous variables. *Advances and Applications in Statistics*, 49:305–326. 14

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4):308–313. 41

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A, General*, 135:370–384. 14

Nogarotto, D. C., Azevedo, C. L. N., Bazán, J. L., et al. (2020). Bayesian modeling and prior sensitivity analysis for zero–one augmented beta regression models with an application to psychometric data. *Brazilian Journal of Probability and Statistics*, 34(2):304–322. 15

Ospina, R. and Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6):1609–1623. 15

Ospina, R. M. (2008). *Modelos de regressão beta inflacionados*. Doutorado em estatística, Universidade de São Paulo. 24, 65, 66, 126

Pace, L. and Salvan, A. (1997). *Principles of statistical inference: from a Neo-Fisherian perspective*, volume 4. World scientific. 67

Pereira, T. L. and Cribari-Neto, F. (2014). Detecting model misspecification in inflated beta regressions. *Communications in Statistics - Simulation and Computation*, 43(3):631–656. 16

Petterle, R. R., Bonat, W. H., and Scarpin, C. T. (2019). Quasi-beta longitudinal regression model applied to water quality index data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(2):346–368. 15

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, pages 825–839. 108, 133

Pulgar, G. M. I. (2009). *Modelos mistos aditivos semiparamétricos de contornos elípticos*. Doutorado em estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. 28

Qiu, Z., SONG, P. X.-K., and TAN, M. (2008). Simplex mixed-effects models for longitudinal proportional data. *Scandinavian Journal of Statistics*, 35(4):577–596. 15

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 55, 59, 86

Rigby, R. A. and Stasinopoulos, D. (1996). A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, 6(1):57–65. 16

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554. 16, 25, 41, 78

Rocha, A. V. and Simas, A. B. (2011). Influence diagnostics in a general class of beta regression models. *Test*, 20(1):95–119. 15

Rodrigues, J., Bazán, J. L., and Suzuki, A. K. (2019). A flexible procedure for formulating probability distributions on the unit interval with applications. *Communications in Statistics-Theory and Methods*, pages 1–17. 15

Rudin, W. (1976). *Principles of mathematical analysis (International Series in Pure & Applied Mathematics*. McGraw-Hill Publishing Co. 65

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics*, 11(4):735–757. 28

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464. 42

Sen, P. K. and Singer, J. M. (1994). *Large sample methods in statistics: an introduction with applications*, volume 25. CRC press. 102

Sen, P. K., Singer, J. M., and de Lima, A. C. P. (2010). *From finite sample to asymptotic methods in statistics*, volume 28. Cambridge University Press. 102

Silva, A. R., Azevedo, C. L., Bazán, J. L., and Nobre, J. S. (2020). Augmented-limited regression models with an application to the study of the risk perceived using continuous scales. *Journal of Applied Statistics*, pages 1–24. 15, 32, 42, 78, 83, 91, 99, 137

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge. 26

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, 11(1):54. 59

Song, P.-K. and Tan, M. (2000). Marginal models for longitudinal continuous proportional data. *Biometrics*, 56(2):496–502. 15, 118, 120, 171

Stasinopoulos, D. M., Rigby, R. A., et al. (2007). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46. 46, 55, 86

Thomas, W. and Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics*, 32(1):59–65. 51

Vanegas, L. H. and Paula, G. A. (2016). An extension of log-symmetric regression models: R codes and applications. *Journal of Statistical Computation and Simulation*, 86(9):1709–1735. 31

Venezuela, M. K. (2008). *Equação de estimação generalizada e influência local para modelos de regressão beta com medidas repetidas*. Doutorado em estatística, Universidade de São Paulo. 15, 112

Venezuela, M. K., Sandoval, M. C., and Botter, D. A. (2011). Local influence in estimating equations. *Computational Statistics & Data Analysis*, 55(4):1867–1883. 113, 114, 137, 138

Verkuilen, J. and Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, 37(1):82–113. 15

Vidakovic, B. (2009). *Statistical modeling by wavelets*, volume 503. John Wiley & Sons. 26

Wahba, G. (1990). *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania. 27

Wahba, G. and Wold, S. (1975). A completely automatic french curve: Fitting spline functions by cross validation. *Communications in Statistics-Theory and Methods*, 4(1):1–17. 43, 80

Wang, J. and Luo, S. (2014). Augmented beta rectangular regression models: A bayesian perspective. *Biometrical Journal*, 58(1):206–221. 15

Wang, L., Xiang, L., Liang, H., and Carroll, R. J. (2011). Generalized additive partial linear models–polynomial spline smoothing estimation and variable selection procedures. *The Annals of Statistics*, 39:1827–1851. 16, 25

Wang, L., Xue, L., Qu, A., and Liang, H. (2014). Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *Ann. Statist.*, 42(2):592–624. 16

Wegman, E. J. and Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association*, 78(382):351–365. 26

Wei, B.-C., Hu, Y.-Q., and Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of statistics*, 25(1):25–37. 45

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):495–518. 25

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC. 25, 27, 31, 43, 80, 110, 136

X.-K. Song, P., Qiu, Z., and Tan, M. (2004). Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. *Biometrical Journal*, 46:540 – 553. 15, 103, 106, 107, 118, 120, 131, 132, 133

Yang, J., Yang, H., and Lu, F. (2019). Rank-based shrinkage estimation for identification in semiparametric additive models. *Statistical Papers*, 60(4):1255–1281. 16

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97(460):1042–1054. 16

Zhao, W., Zhang, R., Lv, Y., and Liu, J. (2014). Variable selection for varying dispersion beta regression model. *Journal of Applied Statistics*, 41(1):95–108. 15

Zhu, H.-T. and Lee, S.-Y. (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):111–126. 49, 83

APPENDIX A – Results development for SPAM

This appendix will presents the demonstration of results defined in Section 1.2, such as penalized score function and penalized Fisher information.

A.1 Results related to Beta-SPAM

A.1.1 Penalized score function for Beta-SPAM

This section will presents the demonstration of results defined in Sections 2.3 and 2.4 for Beta-SPAM, such as penalized score function and penalized Fisher information. As seen in Section 2.2, the penalized log-likelihood function for Semi-parametric Additive Beta Regression Model is given by

$$l(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j,$$

and replacing $l(\boldsymbol{\theta})$ by its expression, the result is given by

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} \Big\{ & \log(\Gamma(\phi_i)) - \log(\Gamma(\mu_i\phi_i)) - \log(\Gamma([1-\mu_i]\phi_i)) + [\mu_i\phi_i - 1]\log(y_i) \\ & + ([1-\mu_i]\phi_i - 1)\log[1-y_i] \Big\} - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j, \end{aligned}$$

with $\mu_i = g_1^{-1}(\eta_{1i})$, $\eta_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{b}_{ij}^\top \boldsymbol{\gamma}_j$, $\phi_i = g_2^{-1}(\eta_{2i})$ and $\eta_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$.

The penalized score function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ is given by $\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top\right)^\top$. Initially, the penalized score function for $\boldsymbol{\beta}$ will be calculated, $\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta})$,

$$\begin{aligned} \dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n} \left[\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\beta}} \right] \\ &= \sum_{i=1}^{n} \phi_i \left[\log\left(\frac{y_i}{1-y_i}\right) + \Psi((1-\mu_i)\phi_i) - \Psi(\mu_i\phi_i) \right] \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \boldsymbol{x}_i \\ &= \sum_{i=1}^{n} \phi_i (y_i^* - \mu_i^*) \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \boldsymbol{x}_i. \end{aligned}$$

The above result can be write in matrix form, and this is given by,

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*.$$

For the parameter $\boldsymbol{\gamma}_j,\ j = 1, \ldots, k$, the penalized score function is,

$$\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j} = \sum_{i=1}^{n} \left[\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\gamma}_j} \right] - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j$$
$$= \sum_{i=1}^{n} \phi_i \left[\log\left(\frac{y_i}{1 - y_i} \right) - \Psi(\mu_i \phi_i) + \Psi((1 - \mu_i)\phi_i) \right] \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1} \boldsymbol{b}_{ij} - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j$$
$$= \sum_{i=1}^{n} \phi_i (y_i^* - \mu_i^*) \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1} \boldsymbol{b}_{ij} - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j.$$

In matrix form, we have the below result,

$$\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j} = \phi \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \boldsymbol{\gamma}_j.$$

And, at last, for parameter $\boldsymbol{\kappa}$, its penalized score function is given by

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} = \sum_{i=1}^{n} \left[\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \boldsymbol{\kappa}} \right]$$
$$= \sum_{i=1}^{n} \Big\{ \mu_i (y_i^* - \mu_i^*) + \log(1 - y_i)$$
$$- \Psi[(1 - \mu_i)\phi_i] + \Psi(\phi_i) \Big\} \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \boldsymbol{e}_i.$$

In matrix form,

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a}.$$

A.1.2 Penalized Fisher's Information Matrix for Beta-SPAM

In this section will be presented the calculations to obtain the second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$ and the penalized Fisher's information matrix for Beta-SPAM, both as defined in the Section 2.4.

The second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$ is defined as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

which is equivalent to

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \ldots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\boldsymbol{\kappa}}(\boldsymbol{\theta}) \end{pmatrix},$$

and the penalized Fisher's information matrix was defined as

$$\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right],$$

in which

$$\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{K}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{K}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) \end{pmatrix}.$$

Initially, the penalized Fisher's information matrix $\boldsymbol{K}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta})$ will be calculated, given that,

$$\boldsymbol{K}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right].$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to the $\boldsymbol{\beta}$ is obtained as shown below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\boldsymbol{\beta}^\top}\boldsymbol{x}_i \\
&= \sum_{i=1}^n \left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{x}_i\boldsymbol{x}_i^\top \\
&= -\sum_{i=1}^n \phi_i\Bigg\{\phi_i[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)] \\
&+ (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{x}_i^\top \\
&= -\sum_{i=1}^n q_i\boldsymbol{x}_i\boldsymbol{x}_i^\top,
\end{aligned}$$

which can be rewrite in matrix form as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{X}.$$

Now, to obtain penalized Fisher's information, we have that, since,

$$\mathbb{E}\left[\dot{\boldsymbol{U}}_p^{\boldsymbol{\beta}}(\boldsymbol{\theta})\right] = 0 \Rightarrow \mathbb{E}(y_i^*) = \mu_i^* \Rightarrow \mathbb{E}\left[\frac{l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\right] = \mathbb{E}\left(y_i^* - \mu_i^*\right) = 0,$$

where y_i^* and μ_i^* are defined in Section 2.3.1, thus,

$$\begin{aligned}
\boldsymbol{K}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\boldsymbol{x}_i\boldsymbol{x}_i^\top \\
&= -\sum_{i=1}^n -\phi_i^2[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{x}_i^\top \\
&= \sum_{i=1}^n w_i\boldsymbol{x}_i\boldsymbol{x}_i^\top.
\end{aligned}$$

In matrix form, the penalized Fisher's information is rewrite as

$$\boldsymbol{K}_p^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X}.$$

Second, the penalized Fisher's information matrix $\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta})$ will be calculated. As in above development, it is known that

$$\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top}\right],$$

so the second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\gamma}_j$ is developed below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} &= \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_j^\top}\boldsymbol{b}_{ij} - \lambda_j\boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n \left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j\boldsymbol{\Lambda}_j \\
&= -\sum_{i=1}^n \phi_i\Bigg\{\phi_i[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)] - (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*) \\
&\times\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j\boldsymbol{\Lambda}_j \\
&= -\sum_{i=1}^n q_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j\boldsymbol{\Lambda}_j,
\end{aligned}$$

which can be rewrite in matrix form as

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j.$$

As comment before, since,

$$\mathbb{E}\left[\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta})\right] = 0 \Rightarrow \mathbb{E}(y_i^*) = \mu_i^* \Rightarrow \mathbb{E}\left[\frac{l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\right] = \mathbb{E}\left(y_i^* - \mu_i^*\right) = 0,$$

thereby, it is possible to obtain the penalized Fisher's information as follows,

$$\begin{aligned}
\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j\boldsymbol{\Lambda}_j \\
&= -\sum_{i=1}^n -\phi_i^2[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)]\left(\frac{\partial g(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j\boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n w_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j\boldsymbol{\Lambda}_j.
\end{aligned}$$

In the matrix form, the penalized Fisher's information can be rewrite as

$$\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{W}\boldsymbol{B}_j + \lambda_j\boldsymbol{\Lambda}_j.$$

And, at last, the penalized Fisher's information matrix for $\boldsymbol{\kappa}$ will be calculated, being known that

$$\boldsymbol{K}_p^{\boldsymbol{\kappa}\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top}\right].$$

Therefore, the second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to ϕ is shown below,

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^\top} &= \sum_{i=1}^{n} \frac{\partial}{\partial \phi_i} \left(\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \right) \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\boldsymbol{\kappa}^\top} \boldsymbol{e}_i \\
&= \sum_{i=1}^{n} \left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i^2} \frac{\partial \phi_i}{\partial \eta_{2i}} + \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i} \frac{\partial^2 \phi_i}{\partial \phi_i \partial \eta_{2i}} \right) \frac{\partial \phi_i}{\partial \eta_{2i}} \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n} \Bigg[\Psi'(\mu_i \phi_i) \mu_i^2 + \Psi'((1-\mu_i)\phi_i)(1-\mu_i)^2 - \Psi'(\phi_i) \\
&\quad - u_i \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \left(\frac{\partial^2 g_2(\phi_i)}{\partial \phi_i^2} \right) \Bigg] \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-2} \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n} s_i \boldsymbol{e}_i \boldsymbol{e}_i^\top,
\end{aligned}
$$

which can be rewrite in matrix form as

$$
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = -\boldsymbol{E}^\top \boldsymbol{S} \boldsymbol{E}.
$$

Thus, the penalized Fisher's information matrix is given by

$$
\begin{aligned}
\boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^\top} \right] &= -\sum_{i=1}^{n} \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i^2} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n} - \left[\Psi'(\mu_i \phi_i) \mu_i^2 + \Psi'((1-\mu_i)\phi_i)(1-\mu_i)^2 - \Psi'(\phi_i) \right] \left(\frac{\partial g_2(\mu_i)}{\partial \mu_i} \right)^{-2} \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= \sum_{i=1}^{n} p_i \boldsymbol{e}_i \boldsymbol{e}_i^\top,
\end{aligned}
$$

and, rewriting in matrix form, the result is,

$$
\boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}} = \boldsymbol{E}^\top \boldsymbol{P} \boldsymbol{E}.
$$

For cross second derivative with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$, the penalized Fisher's information is given by,

$$
\boldsymbol{K}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\kappa}^\top} \right].
$$

The second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ is give by

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\kappa}^\top} &= \sum_{i=1}^{n} \frac{\partial l_p^2(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i \partial \phi_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{x}_i \\
&= \sum_{i=1}^{n} \frac{\partial l_p^2(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i \partial \phi_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \phi_i}{\partial \eta_{2i}} \boldsymbol{x}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n} \{ \phi_i [\Psi'(\mu_i \phi_i) \mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)] - (y_i^* - \mu_i^*) \} \\
&\quad \times \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1} \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \boldsymbol{x}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n} q_i^* \boldsymbol{x}_i \boldsymbol{e}_i^\top.
\end{aligned}
$$

The last result can be rewrite in matrix form as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\boldsymbol{\kappa}}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top \boldsymbol{Q}^* \boldsymbol{E}.$$

Since $\mathbb{E}(y_i^*) = \mu_i^*$, it implies that,

$$\begin{aligned}
\boldsymbol{K}_p^{\boldsymbol{\beta}\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\boldsymbol{x}_i\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n -\{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)]\} \\
&\times \left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{x}_i\boldsymbol{e}_i^\top \\
&= \sum_{i=1}^n w_i^*\boldsymbol{x}_i\boldsymbol{e}_i^\top.
\end{aligned}$$

In matrix notation, penalized Fisher's information matrix is rewrite as

$$\boldsymbol{K}_p^{\boldsymbol{\beta}\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{W}^* \boldsymbol{E}.$$

Furthermore, for cross derivative with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\kappa}$, the penalized Fisher's information is obtained as shown below

$$\boldsymbol{K}_p^{\boldsymbol{\gamma}_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right].$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\kappa}$ is given by

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\gamma}_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top} &= \sum_{i=1}^n \frac{\partial l_p^2(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^n \frac{\partial l_p^2(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n\{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)] - (y_i^* - \mu_i^*)\} \\
&\times \left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n q_i^*\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top
\end{aligned}$$

which can be rewrite in matrix form as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\gamma}_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top \boldsymbol{Q}^* \boldsymbol{E}.$$

Since $\mathbb{E}(y_i^*) = \mu_i^*$, it implies at

$$
\begin{aligned}
\boldsymbol{K}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n -\{\phi_i[\Psi'(\mu_i\phi_i)\mu_i - \Psi'((1-\mu_i)\phi_i)(1-\mu_i)]\} \\
&\times \left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= \sum_{i=1}^n w_i^*\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top.
\end{aligned}
$$

In matrix notation, penalized Fisher's information matrix is rewrite as

$$
\boldsymbol{K}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{W}^*\boldsymbol{E}.
$$

For cross derivative with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\gamma}_{j'}$, the penalized Fisher's information is obtained as shown below

$$
\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top}\right].
$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\gamma}_j'$ is given by

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} &= \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_{j'}^\top}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^n\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= -\sum_{i=1}^n \phi_i\Bigg\{\phi_i[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)] + (y_i^* - \mu_i^*) \\
&\times \left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= -\sum_{i=1}^n q_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top,
\end{aligned}
$$

which, in matrix form is,

$$
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_{j'}
$$

As comment before, since $\mathbb{E}(y_i^*) = \mu_i^*$, it implies that,

$$
\begin{aligned}
\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= -\sum_{i=1}^n -\phi_i^2[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= \sum_{i=1}^n w_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top.
\end{aligned}
$$

In the matrix form, the penalized Fisher's information matrix is rewrite as

$$\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top \boldsymbol{W}\boldsymbol{B}_{j'}.$$

And, at last, for cross derivative with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$, the penalized Fisher's information is obtained as shown below

$$\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top}\right].$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$ is given by,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} &= \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_j^\top}\boldsymbol{x}_i \\
&= \sum_{i=1}^n \left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= -\sum_{i=1}^n \phi_i\Bigg\{\phi_i[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)] \\
&+ (\boldsymbol{y}_i^* - \boldsymbol{\mu}_i^*)\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= -\sum_{i=1}^n q_i\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top,
\end{aligned}$$

which can be rewrite in matrix form as,

$$\ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{B}_j.$$

Since $\mathbb{E}(y_i^*) = \mu_i^*$, thus,

$$\begin{aligned}
\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= -\sum_{i=1}^n -\phi_i^2[\Psi'(\mu_i\phi_i) + \Psi'((1-\mu_i)\phi_i)]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= \sum_{i=1}^n w_i\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top.
\end{aligned}$$

In the matrix form, the penalized Fisher's information matrix is

$$\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{B}_j.$$

A.2 Results related to Simplex-SPAM

A.2.1 Penalized score function for Simplex-SPAM

This section will presents the demonstration of results defined in Sections 2.3 and 2.4 for Simplex-SPAM, such as penalized score function and penalized Fisher information.

As seen in Section 2.2, the penalized log-likelihood function for Semi-parametric Additive Simplex Regression Model is given by

$$l(\boldsymbol{\theta},\boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \sum_{j=1}^{k}\frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top\boldsymbol{\Lambda}_j^d\boldsymbol{\gamma}_j,$$

and replacing $l(\boldsymbol{\theta})$ by its expression, the result is given by

$$\begin{aligned} l(\boldsymbol{\theta},\boldsymbol{\lambda}) = \sum_{i=1}^{n}\Bigg\{ & -\frac{1}{2}\left\{\log(2\pi\phi_i) + 3\log[y_i(1-y_i)]\right\} \\ & - \frac{(y-\mu_i)^2}{2\phi_i y_i(1-y_i)\mu_i^2(1-\mu_i)^2}\Bigg\} - \sum_{j=1}^{k}\frac{\lambda_j}{2}\boldsymbol{\gamma}_j^\top\boldsymbol{\Lambda}_j^d\boldsymbol{\gamma}_j, \end{aligned}$$

with $\mu_i = g_1^{-1}(\eta_{1i})$, $\eta_{1i} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{k}\boldsymbol{b}_{ij}^\top\boldsymbol{\gamma}_j$, $\phi_i = g_2^{-1}(\eta_{2i})$ and $\eta_{2i} = \boldsymbol{e}_i^\top\boldsymbol{\kappa}$.

The penalized score function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top)^\top$ is given by $\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta})^\top\right)^\top$. Initially, penalized score function for $\boldsymbol{\beta}$ will be calculated, $\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})$,

$$\begin{aligned} \dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\boldsymbol{\beta}}\right] \\ &= \sum_{i=1}^{n}\frac{1}{\phi_i}\left\{\frac{(y_i-\mu_i)}{\mu_i(1-\mu_i)}\left[\frac{(y_i-\mu_i)^2}{y_i(1-y_i)\mu_i^2(1-\mu_i)^2} + \frac{1}{\mu_i^2(1-\mu_i)^2}\right]\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{x}_i \\ &= \sum_{i=1}^{n}\frac{1}{\phi_i}u_i(y_i-\mu_i)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{x}_i. \end{aligned} \tag{A.1}$$

The above result can be write in matrix form, and this is given by,

$$\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) = \boldsymbol{X}^\top\boldsymbol{T}_1\boldsymbol{f}^*.$$

For the parameter $\boldsymbol{\gamma}_j,\ j = 1,\ldots,k$, the penalized score function is given by

$$\begin{aligned} \dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j} = \sum_{i=1}^{n}\left[\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\boldsymbol{\gamma}_j}\right] - \lambda_j\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_j \\ &= \sum_{i=1}^{n}\frac{1}{\phi_i}\left\{\frac{(y_i-\mu_i)}{\mu_i(1-\mu_i)}\left[\frac{(y_i-\mu_i)^2}{y_i(1-y_i)\mu_i^2(1-\mu_i)^2} + \frac{1}{\mu_i^2(1-\mu_i)^2}\right]\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{b}_{ij} \\ &\quad - \lambda_j\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_j = \sum_{i=1}^{n}\frac{1}{\phi_i}u_i(y_i-\mu_i)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{b}_{ij} - \lambda_j\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_j. \end{aligned}$$

In matrix form, we have the below result,

$$\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{T}_1\boldsymbol{f}^* - \lambda_j\boldsymbol{\Lambda}_j\boldsymbol{\gamma}_j.$$

And, at last, for parameter $\boldsymbol{\kappa}$, the penalized score function is given by

$$\begin{aligned}
\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}} &= \sum_{i=1}^{n}\left[\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \phi_i}\frac{\partial \phi_i}{\partial \eta_{2i}}\frac{\partial \eta_{2i}}{\partial \boldsymbol{\kappa}}\right] \\
&= \sum_{i=1}^{n}\left[-\frac{1}{2\phi}+\frac{(y_i-\mu_i)^2}{2\phi^2 y_i(1-y_i)\mu_i^2(1-\mu_i)^2}\right]\left(\frac{\partial g_2(\phi_i)}{\partial \phi_i}\right)^{-1}\boldsymbol{e}_i \\
&= \sum_{i=1}^{n} a_i \boldsymbol{T}_2 \boldsymbol{e}_i,
\end{aligned}$$

which can be rewrite in matrix form as

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \boldsymbol{E}^{\top}\boldsymbol{T}_2\boldsymbol{a}.$$

A.2.2 Penalized Fisher's Information Matrix for Simplex-SPAM

In this section will be presented the calculations to obtain the second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$ and the penalized Fisher's information matrix for Simplex-SPAM, both as defined in Section 2.4.

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$ is defined as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}},$$

which is equivalent to

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \begin{pmatrix}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) \\
\ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta})
\end{pmatrix},$$

and the penalized Fisher's information matrix was defined as

$$\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\top}}\right],$$

in which

$$\boldsymbol{K}_p^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}) = \begin{pmatrix}
\boldsymbol{K}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) \\
\boldsymbol{K}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\boldsymbol{K}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\
\boldsymbol{K}_p^{\boldsymbol{\kappa\beta}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \boldsymbol{K}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta})
\end{pmatrix}.$$

Initially, it will be defined some quantities that will be useful to the demonstrations. Let $d(y;\mu)$ be defined as

$$d(y;\mu) = \frac{(y-\mu)^2}{y(1-y)\mu^2(1-\mu)^2}, \tag{A.2}$$

and

$$\frac{\partial d(y;\mu)}{\partial \mu} = -2u(y-\mu),$$

where u_i is defined in Equation 2.12, this relation is obtained from result presented in Equation A.1, because

$$\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu} = -\frac{1}{2}\frac{\partial d(y;\mu)}{\partial \mu} = u(y-\mu). \tag{A.3}$$

Now, some properties will be shown, for more details and proofs, see Song and Tan (2000).

Proposition 1. *Let $y \sim Simplex^{-}(\mu, \phi)$. Then,*

(1) $\mathbb{E}[d(y;\mu)] = \phi$;

(2) $\mathbb{E}[(y-\mu)^2 u] = -2\phi$;

(3) $\mathbb{E}[(y-\mu)d(y;\mu)] = 0$;

(4) $\mathbb{E}\left[\frac{\partial^2 d(y;\mu)}{\partial \mu^2}\right] = 2\left[\frac{3\phi}{\mu(1-\mu)} + \frac{1}{\mu^3(1-\mu)^3}\right]$

Besides that, another result that is consequence of item (3) in Proposition 1 and will useful is

$$\mathbb{E}[(y-\mu)u] = 0. \tag{A.4}$$

Proof: based on the relation shown in Equation A.3 and given that u is defined as

$$u = \frac{1}{\mu(1-\mu)}\left[d(y;\mu) + \frac{1}{\mu^2(1-\mu)^2}\right],$$

and considering $\mathbb{E}(y) = \mu$, it follows that

$$\mathbb{E}\left[\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu}\right] = \mathbb{E}[u(y-\mu)] = \frac{\mathbb{E}[(y-\mu)d(y;\mu)]}{\mu(1-\mu)} + \frac{\mathbb{E}(y-\mu)}{\mu^3(1-\mu)^3} = 0.$$

The penalized Fisher's information matrix $\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta})$ will be calculated. It known that

$$\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\top}}\right].$$

The second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to the $\boldsymbol{\beta}$ is obtained as shown below,

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \sum_{i=1}^n \frac{\partial}{\partial \mu_i} \left(\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \right) \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\beta}^\top} \boldsymbol{x}_i \\
&= \sum_{i=1}^n \left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i^2} \frac{\partial \mu_i}{\partial \eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i} \frac{\partial^2 \mu_i}{\partial \mu_i \partial \eta_{1i}} \right) \frac{\partial \mu_i}{\partial \eta_{1i}} \boldsymbol{x}_i \boldsymbol{x}_i^\top \\
&= -\sum_{i=1}^n \frac{1}{\phi_i} \left\{ u_i + (y_i - \mu_i) \left[\frac{2(y_i - \mu_i) u_i}{\mu_i (1-\mu_i)} + \frac{3(1-2\mu_i)}{\mu_i^4 (1-\mu_i)^4} + \frac{(1-2\mu_i) d(y_i; \mu_i)}{\mu_i^2 (1-\mu_i)^2} \right] \right. \\
&\left. + u_i (y_i - \mu_i) \left(\frac{\partial^2 g_1(\mu_i)}{\partial \mu_i^2} \right) \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1} \right\} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-2} \boldsymbol{x}_i \boldsymbol{x}_i^\top \\
&= -\sum_{i=1}^n q_i \boldsymbol{x}_i \boldsymbol{x}_i^\top.
\end{aligned}
$$

In matrix form, it can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\beta\beta}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top \boldsymbol{Q} \boldsymbol{X}.$$

Thus, using the results given by Proposition 1, and given that $\dfrac{\partial^2 d(y_i; \mu_i)}{\partial \mu^2} = (y_i - \mu_i) \left[\dfrac{2(y_i - \mu_i) u_i}{\mu_i (1-\mu_i)} + \dfrac{3(1-2\mu_i)}{\mu_i^4 (1-\mu_i)^4} + \dfrac{(1-2\mu_i) d(y_i; \mu_i)}{\mu_i^2 (1-\mu_i)^2} \right]$, the penalized Fisher's information matrix is obtained as shown below

$$
\begin{aligned}
\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right] = -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2 \boldsymbol{x}_i \boldsymbol{x}_i^\top \\
&= \sum_{i=1}^n \frac{1}{\phi_i} \left\{ \mathbb{E}(u_i) + \mathbb{E}\left(\frac{\partial^2 d(y_i; \mu_i)}{\partial \mu^2} \right) \right\} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-2} \boldsymbol{x}_i \boldsymbol{x}_i^\top \\
&= \sum_{i=1}^n \frac{1}{\phi_i} \left[\frac{3\phi_i}{\mu_i (1-\mu_i)} + \frac{1}{\mu_i^3 (1-\mu_i)^3} \right] \boldsymbol{x}_i \boldsymbol{x}_i^\top \\
&= \sum_{i=1}^n \frac{1}{\phi_i} w_i \boldsymbol{x}_i \boldsymbol{x}_i^\top.
\end{aligned}
$$

In matrix form, it can be rewrite as

$$\boldsymbol{K}_p^{\beta\beta}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}.$$

Now, the penalized Fisher's information matrix $\boldsymbol{K}_p^{\gamma_j \gamma_j}(\boldsymbol{\theta})$ will be calculated. As in above development, it is known that

$$\boldsymbol{K}_p^{\gamma_j \gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^\top} \right],$$

thereby, the second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\gamma}_j$ is developed below,

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^\top} &= \sum_{i=1}^n \frac{\partial}{\partial \mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_{1i}}\right)\frac{\partial \mu_i}{\partial \eta_{1i}}\frac{\partial \eta_{1i}}{\boldsymbol{\gamma}_j^\top}\boldsymbol{b}_{ij} - \lambda_j \boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n \left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i^2}\frac{\partial \mu_i}{\partial \eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i}\frac{\partial^2 \mu_i}{\partial \mu_i \partial \eta_{1i}}\right)\frac{\partial \mu_i}{\partial \eta_{1i}}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j \boldsymbol{\Lambda}_j \\
&= -\sum_{i=1}^n \frac{1}{\phi_i}\Bigg\{u_i + (y_i - \mu_i)\left[\frac{2(y_i - \mu_i)u_i}{\mu_i(1-\mu_i)} + \frac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4} + \frac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right] \\
&+ u_i(y_i - \mu_i)\left(\frac{\partial^2 g_1(\mu_i)}{\partial \mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j \boldsymbol{\Lambda}_j \\
&= -\sum_{i=1}^n q_i \boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j \boldsymbol{\Lambda}_j.
\end{aligned}
$$

In matrix form, it be shown that

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top \boldsymbol{Q}\boldsymbol{B}_j - \lambda_j \boldsymbol{\Lambda}_j.$$

Thus, using the results given by Proposition 1, and given that $\frac{\partial^2 d(y_i;\mu_i)}{\partial \mu^2} = (y_i - \mu_i)\left[\frac{2(y_i - \mu_i)u_i}{\mu_i(1-\mu_i)} + \frac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4} + \frac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right]$, the penalized Fisher's information matrix is obtained as shown below

$$
\begin{aligned}
\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_j^\top}\right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i^2}\right)\left(\frac{\partial \mu_i}{\partial \eta_{1i}}\right)^2 \boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j \boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n \frac{1}{\phi_i}\left\{\mathbb{E}(u_i) + \mathbb{E}\left(\frac{\partial^2 d(y_i;\mu_i)}{\partial \mu^2}\right)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j \boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n \frac{1}{\phi_i}\left[\frac{3\phi_i}{\mu_i(1-\mu_i)} + \frac{1}{\mu_i^3(1-\mu_i)^3}\right]\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j \boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n \frac{1}{\phi_i}\left[\frac{3\phi_i}{\mu_i(1-\mu_i)} + \frac{1}{\mu_i^3(1-\mu_i)^3}\right]\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j \boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n w_i \boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top + \lambda_j \boldsymbol{\Lambda}_j.
\end{aligned}
$$

The penalized Fisher's information matrix in matrix form is given by

$$\boldsymbol{K}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top \boldsymbol{W}\boldsymbol{B}_j.$$

And, at last, the penalized Fisher's information matrix for parameter $\boldsymbol{\kappa}$ will be calculated, being known that

$$\boldsymbol{K}_p^{\boldsymbol{\kappa}\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa}\partial \boldsymbol{\kappa}^\top}\right].$$

Therefore, the second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\kappa}$ is shown below,

$$\begin{aligned}\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^\top} &= \sum_{i=1}^n \frac{\partial}{\partial \phi_i} \left(\frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \right) \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\boldsymbol{\kappa}^\top} \boldsymbol{e}_i \\
&= \sum_{i=1}^n \left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i^2} \frac{\partial \phi_i}{\partial \eta_{2i}} + \frac{\partial l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i} \frac{\partial^2 \phi_i}{\partial \phi_i \partial \eta_{2i}} \right) \frac{\partial \phi_i}{\partial \eta_{2i}} \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n \left[-\frac{1}{2\phi_i^2} + \frac{1}{\phi_i^3} d(y_i; \mu_i) + a_i \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \left(\frac{\partial^2 g_2(\phi_i)}{\partial \phi_i^2} \right) \right] \\
&\times \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-2} \boldsymbol{e}_i \boldsymbol{e}_i^\top = -\sum_{i=1}^n s_i \boldsymbol{e}_i \boldsymbol{e}_i^\top.\end{aligned}$$

In matrix form, it can rewrite as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = -\boldsymbol{E}^\top \boldsymbol{S} \boldsymbol{E}^\top.$$

Thus, using the results given by Proposition 1, the penalized Fisher's information matrix is given by

$$\begin{aligned}\boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\kappa} \partial \boldsymbol{\kappa}^\top} \right] &= -\sum_{i=1}^n \mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \phi_i^2} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n - \left[-\frac{1}{2\phi_i^2} + \frac{1}{\phi_i^3} \mathbb{E}[d(y_i; \mu_i)] \right] \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \boldsymbol{e}_i \boldsymbol{e}_i^\top \\
&= \sum_{i=1}^n \left[-\frac{1}{2\phi_i^2} + \frac{1}{\phi_i^2} \right] \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \boldsymbol{e}_i \boldsymbol{e}_i^\top = \sum_{i=1}^n \frac{1}{2\phi_i^2} \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2 \boldsymbol{e}_i \boldsymbol{e}_i^\top,\end{aligned}$$

which can be rewrite in matrix form as

$$\boldsymbol{K}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \boldsymbol{E}^\top \boldsymbol{P} \boldsymbol{E}.$$

For cross second derivative with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$, the penalized Fisher's information is given by,

$$\boldsymbol{K}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\kappa}^\top} \right].$$

The second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ is give by

$$\begin{aligned}\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\kappa}^\top} &= \sum_{i=1}^n \frac{\partial l_p^2(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i \partial \phi_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \boldsymbol{\kappa}^\top} \boldsymbol{x}_i \\
&= \sum_{i=1}^n \frac{\partial l_p^2(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \mu_i \partial \phi_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \phi_i}{\partial \eta_{2i}} \boldsymbol{x}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n \left[\frac{1}{\phi_i^2} u_i (y_i - \mu_i) \right] \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1} \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \boldsymbol{x}_i \boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n q_i^* \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i} \right)^{-1} \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \boldsymbol{x}_i \boldsymbol{e}_i^\top.\end{aligned}$$

In matrix form, the quantity can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\beta\kappa}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top \boldsymbol{Q}^* \boldsymbol{E}.$$

Thus, using the results given by Equation A.4, the penalized Fisher's information matrix is shown below,

$$\begin{aligned}
\boldsymbol{K}_p^{\beta\kappa}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right] = -\sum_{i=1}^{n}\mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\boldsymbol{x}_i\boldsymbol{e}_i^\top \\
&= \sum_{i=1}^{n}\frac{1}{\phi_i^2}\mathbb{E}[u_i(y_i-\mu_i)]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\boldsymbol{x}_i\boldsymbol{e}_i^\top = \boldsymbol{0}
\end{aligned}$$

For cross second derivative with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\kappa}$, the penalized Fisher's information is given by,

$$\boldsymbol{K}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right].$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\kappa}$ is give by

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top} &= \sum_{i=1}^{n}\frac{\partial l_p^2(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^{n}\frac{\partial l_p^2(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n}\left[\frac{1}{\phi_i^2}u_i(y_i-\mu_i)\right]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n}q_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top.
\end{aligned}$$

Rewrite in matrix form, we obtain that

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top \boldsymbol{Q}^* \boldsymbol{E}.$$

Thus, using the results given by Equation A.4, the penalized Fisher's information matrix is shown below,

$$\begin{aligned}
\boldsymbol{K}_p^{\gamma_j\kappa}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right] = -\sum_{i=1}^{n}\mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= \sum_{i=1}^{n}\frac{1}{\phi_i^2}\mathbb{E}[u_i(y_i-\mu_i)]\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\left(\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top = \boldsymbol{0}
\end{aligned}$$

The penalized Fisher's information matrix $\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta})$ will be calculated. As in above development, it is known that

$$\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top}\right].$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\gamma}_j$ and $\boldsymbol{\gamma}_{j'}$ is developed below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} = \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_{j'}^\top}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^n\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= -\sum_{i=1}^n\frac{1}{\phi_i}\Bigg\{u_i + (y_i-\mu_i)\left[\frac{2(y_i-\mu_i)u_i}{\mu_i(1-\mu_i)} + \frac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4} + \frac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right] \\
&+ u_i(y_i-\mu_i)\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= -\sum_{i=1}^n q_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top.
\end{aligned}$$

And in matrix form, we have that

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top\boldsymbol{Q}\boldsymbol{B}_{j'}.$$

Thus, using the results given by Proposition 1, and given that $\dfrac{\partial^2 d(y_i;\mu_i)}{\partial\mu^2} = (y_i-\mu_i)\left[\dfrac{2(y_i-\mu_i)u_i}{\mu_i(1-\mu_i)}+\dfrac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4}+\dfrac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right]$, the penalized Fisher's information matrix is obtained as shown below

$$\begin{aligned}
\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) &= \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top}\right] = -\sum_{i=1}^n\mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= \sum_{i=1}^n\frac{1}{\phi_i}\left\{\mathbb{E}(u_i) + \mathbb{E}\left(\frac{\partial^2 d(y_i;\mu_i)}{\partial\mu^2}\right)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= \sum_{i=1}^n\frac{1}{\phi_i}\left[\frac{3\phi_i}{\mu_i(1-\mu_i)} + \frac{1}{\mu_i^3(1-\mu_i)^3}\right]\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= \sum_{i=1}^n w_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top.
\end{aligned}$$

The penalized Fisher's information matrix in matrix form is given by

$$\boldsymbol{K}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{W}\boldsymbol{B}_{j'}.$$

The penalized Fisher's information matrix $\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta})$ will be calculated. As in above development, it is known that

$$\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}^\top\partial\boldsymbol{\gamma}_j^\top}\right].$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$ is developed below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} &= \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_j^\top}\boldsymbol{x}_i \\
&= \sum_{i=1}^n\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= -\sum_{i=1}^n\frac{1}{\phi_i}\Bigg\{u_i + (y_i-\mu_i)\left[\frac{2(y_i-\mu_i)u_i}{\mu_i(1-\mu_i)} + \frac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4} + \frac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right] \\
&+ u_i(y_i-\mu_i)\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= -\sum_{i=1}^n q_i\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top.
\end{aligned}$$

Rewrite in matrix form, we obtain that

$$\ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top\boldsymbol{Q}\boldsymbol{B}_j$$

Thus, using the results given by Proposition 1, and given that $\dfrac{\partial^2 d(y_i;\mu_i)}{\partial\mu^2} = (y_i-\mu_i)\left[\dfrac{2(y_i-\mu_i)u_i}{\mu_i(1-\mu_i)} + \dfrac{3(1-2\mu_i)}{\mu_i^4(1-\mu_i)^4} + \dfrac{(1-2\mu_i)d(y_i;\mu_i)}{\mu_i^2(1-\mu_i)^2}\right]$, the penalized Fisher's information matrix is obtained as shown below

$$\begin{aligned}
\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top}\right] &= -\sum_{i=1}^n\mathbb{E}\left(\frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\mu_i^2}\right)\left(\frac{\partial\mu_i}{\partial\eta_{1i}}\right)^2\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= \sum_{i=1}^n\frac{1}{\phi_i}\left\{\mathbb{E}(u_i) + \mathbb{E}\left(\frac{\partial^2 d(y_i;\mu_i)}{\partial\mu^2}\right)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= \sum_{i=1}^n w_i\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top.
\end{aligned}$$

The penalized Fisher's information matrix in matrix form is given by

$$\boldsymbol{K}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{B}_j.$$

A.3 Results related to BR-SPAM

A.3.1 Penalized gradient vector for BR-SPAM

This section will presents the demonstration of results defined in Sections 2.3 and 2.4 for BR-SPAM, such as penalized gradient. As seen in Section 2.3.3, for Semiparametric Additive Beta Rectangular Regression Model, we will based the calculus on the quantity $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$, which is given by

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)}) = \mathbb{E}[l_p^c(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\theta}^{(m)}) = \sum_{i=1}^{n} &\hat{v}_i^{(m)} \log \varepsilon_i + (1-\hat{v}_i^{(m)}) \log(1-\varepsilon_i) \\ &+ (1-\hat{v}_i^{(m)})[\log(\Gamma(\phi_i)) - \log(\Gamma(\delta_i\phi_i)) - \log(\Gamma((1-\delta_i)\phi_i)) \\ &+ (\delta_i\phi_i - 1)\log y_i + ((1-\delta_i)\phi_i - 1)\log(1-y_i)] - \boldsymbol{\gamma}^\top \boldsymbol{\Lambda}\boldsymbol{\gamma},\end{aligned}$$

where $\hat{v}_i$ is defined in Equation 2.10, $\mu_i = g_1^{-1}(\eta_{1i})$, $\eta_{1i} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{k} \boldsymbol{b}_{ij}^\top \boldsymbol{\gamma}_j$, $\phi_i = g_2^{-1}(\eta_{2i})$, $\eta_{2i} = \boldsymbol{e}_i^\top \boldsymbol{\kappa}$, $\varepsilon_i = 1 - \sqrt{1-4\alpha\mu_i(1-\mu_i)}$ and $\delta_i = \dfrac{\mu_i - \frac{1}{2} + \frac{1}{2}\sqrt{1-4\alpha\mu_i(1-\mu_i)}}{\sqrt{1-4\alpha\mu_i(1-\mu_i)}}$.

The penalized gradient vector for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_k^\top, \boldsymbol{\kappa}^\top, \alpha)^\top$ is given by $\dot{\boldsymbol{U}}_p^{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\gamma_1}(\boldsymbol{\theta})^\top, \ldots, \dot{\boldsymbol{U}}_p^{\gamma_k}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}), \dot{U}_p^{\alpha}(\boldsymbol{\theta})\right)^\top$. Initially, penalized gradient vector for $\boldsymbol{\beta}$ will be calculated, $\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta})$,

$$\begin{aligned}\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) &= \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \left[\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\beta}}\right] \\ &= \sum_{i=1}^{n} \Bigg\{ \frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i} \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] \\ &+ (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^* - \delta_i^*)\Bigg\} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \boldsymbol{x}_i \\ &= \sum_{i=1}^{n} f_i^* \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \boldsymbol{x}_i,\end{aligned}$$

with $\delta_i^* = \Psi(\delta_i\phi_i) - \Psi[(1-\delta_i)\phi_i]$, $y_i^* = \log[y_i/(1-y_i)]$, for $i = 1, \ldots, n$. The above result can be write in matrix form, and this is given by,

$$\dot{\boldsymbol{U}}_p^{\beta}(\boldsymbol{\theta}) = \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{T}_1 \boldsymbol{f}^*.$$

For the parameter $\boldsymbol{\gamma}_j,\ j = 1, \ldots, k$, the penalized gradient vector is given by

$$\begin{aligned}\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\gamma}_j} = \sum_{i=1}^{n} \left[\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \boldsymbol{\gamma}_j}\right] - \lambda_j \boldsymbol{\Lambda}_j \gamma_j \\ &= \sum_{i=1}^{n} \Bigg\{ \frac{2\alpha(1-2\mu_i)}{1-\varepsilon_i} \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] \\ &+ (1-\hat{v}_i)(1-\alpha)\frac{\phi_i}{(1-\varepsilon_i)^3}(y_i^* - \delta_i^*)\Bigg\} \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \boldsymbol{b}_{ij} \\ &- \lambda_j \boldsymbol{\Lambda}_j \gamma_j = \sum_{i=1}^{n} f_i^* \left(\frac{\partial g_1(\mu_i)}{\partial \mu_i}\right)^{-1} \boldsymbol{b}_{ij} - \lambda_j \boldsymbol{\Lambda}_j \gamma_j.\end{aligned}$$

In matrix form, we have the below result,

$$\dot{\boldsymbol{U}}_p^{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\gamma}_j} = \boldsymbol{B}_j^\top \boldsymbol{T}_1 \boldsymbol{f}^* - \lambda_j \boldsymbol{\Lambda}_j \gamma_j.$$

For the parameter $\boldsymbol{\kappa}$, the penalized gradient vector is given by

$$\begin{aligned} \boldsymbol{U}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\kappa}} &= \sum_{i=1}^{n} \left[\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \boldsymbol{\kappa}} \right] \\ &= \sum_{i=1}^{n} (1-\hat{v}_i)\{\delta_i(y_i^* - \delta_i^*) + \log(1-y_i) \\ &\quad - \Psi[(1-\delta_i)\phi_i] + \Psi(\phi_i)\} \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \boldsymbol{e}_i \\ &= \sum_{i=1}^{n} a_i \left(\frac{\partial g_2(\phi_i)}{\partial \phi_i} \right)^{-1} \boldsymbol{e}_i, \end{aligned}$$

which in matrix form, can be rewrite as

$$\boldsymbol{U}_p^{\boldsymbol{\kappa}}(\boldsymbol{\theta}) = \boldsymbol{E}^\top \boldsymbol{T}_2 \boldsymbol{a}.$$

At last, for the parameter α, the penalized gradient vector is given by

$$\begin{aligned} U_p^{\alpha}(\boldsymbol{\theta}) &= \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \alpha} \\ &= \sum_{i=1}^{n} \frac{2\mu_i(1-\mu_i)}{(1-\varepsilon_i)} \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)} \right] - (1-\hat{v}_i) \left[\frac{\phi_i\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^3} \right] (y_i^* - \delta_i^*) \\ &= \sum_{i=1}^{n} d_i = trace(\boldsymbol{D}). \end{aligned}$$

A.3.2 Penalized Hessian matrix for BR-SPAM

In this section will be presented the calculations to obtain the second derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with respect to $\boldsymbol{\theta}$ defined in Section 2.4 for BR-SPAM.

The second derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with respect to $\boldsymbol{\theta}$ is defined as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top},$$

which is equivalent to

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta}\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_1\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_1\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\gamma_k\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\gamma_k\boldsymbol{\kappa}}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_1}(\boldsymbol{\theta}) & \dots & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\gamma_k}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa}\boldsymbol{\kappa}}(\boldsymbol{\theta}) \end{pmatrix}.$$

The second derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with respect to the $\boldsymbol{\beta}$ is obtained as shown below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\partial\boldsymbol{\beta}^\top}\boldsymbol{x}_i \\
&= \sum_{i=1}^n \left(\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{x}_i\boldsymbol{x}_i^\top \\
&= \sum_{i=1}^n\Bigg\{\left[\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] \\
&\times\left[-\frac{4\alpha}{(1-\varepsilon_i)} + \frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^3}\right] + (1-\alpha)(1-\hat{v}_i) \\
&\times\left[\frac{-(1-\alpha)\phi_i^2}{(1-\varepsilon_i)^6}w_i^* + \frac{6\alpha\phi_i(1-2\mu_i)}{(1-\varepsilon_i)^5}(y_i^*-\delta_i^*)\right] - f_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1} \\
&\times\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{x}_i^\top \\
&= \sum_{i=1}^n r_i\boldsymbol{x}_i\boldsymbol{x}_i^\top.
\end{aligned}$$

In matrix form, it can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) = \boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{X}.$$

The second derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with respect to $\boldsymbol{\gamma}_j$ is developed below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_j^\top} = \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_j^\top}\boldsymbol{b}_{ij} - \lambda_j\boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n \left(\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j\boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n\Bigg\{\left[\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] \\
&\times\left[-\frac{4\alpha}{(1-\varepsilon_i)} + \frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^3}\right] + (1-\alpha)(1-\hat{v}_i) \\
&\times\left[\frac{-(1-\alpha)\phi_i^2}{(1-\varepsilon_i)^6}w_i^* + \frac{6\alpha\phi_i(1-2\mu_i)}{(1-\varepsilon_i)^5}(y_i^*-\delta_i^*)\right] - f_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1} \\
&\times\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\Bigg\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j\boldsymbol{\Lambda}_j \\
&= \sum_{i=1}^n r_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij}^\top - \lambda_j\boldsymbol{\Lambda}_j.
\end{aligned}$$

In matrix form, it be shown that

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{R}\boldsymbol{B}_j - \lambda_j\boldsymbol{\Lambda}_j.$$

Therefore, the second derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with respect to $\boldsymbol{\kappa}$ is shown below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top} &= \sum_{i=1}^n \frac{\partial}{\partial\phi_i}\left(\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\phi_i}\frac{\partial\phi_i}{\partial\eta_{2i}}\right)\frac{\partial\phi_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\boldsymbol{\kappa}^\top}\boldsymbol{e}_i \\
&= \sum_{i=1}^n\left(\frac{\partial^2 Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\phi_i^2}\frac{\partial\phi_i}{\partial\eta_{2i}} + \frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})}{\partial\phi_i}\frac{\partial^2\phi_i}{\partial\phi_i\partial\eta_{2i}}\right)\frac{\partial\phi_i}{\partial\eta_{2i}}\boldsymbol{e}_i\boldsymbol{e}_i^\top \\
&= \sum_{i=1}^n\Bigg[-(1-\hat{v}_i)[(1-\delta_i)^2\Psi'((1-\delta_i)\phi_i) + \delta^2\Psi'(\delta_i\phi_i) - \Psi'(\phi_i)] \\
&- p_i\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\left(\frac{\partial^2 g_2(\phi_i)}{\partial\phi_i^2}\right)\Bigg]\times\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-2}\boldsymbol{e}_i\boldsymbol{e}_i^\top \\
&= \sum_{i=1}^n s_i\boldsymbol{e}_i\boldsymbol{e}_i^\top.
\end{aligned}$$

In matrix form, it can rewrite as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\kappa\kappa}}(\boldsymbol{\theta}) = \boldsymbol{E}^\top\boldsymbol{S}\boldsymbol{E}.$$

The second derivative of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(m)})$ with respect to α is shown below,

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\alpha\alpha}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\alpha^2} &= \sum_{i=1}^n\Bigg\{\left[\frac{4\mu_i^2(1-\mu_i)^2}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right] + \left[\frac{\hat{v}_i}{\varepsilon_i} - \frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right] \\
&\times\left[\frac{4\mu_i^2(1-\mu_i)^2}{(1-\varepsilon_i)^3}\right] - (1-\hat{v}_i)\left[\frac{\phi_i\mu_i^2(1-\mu_i)^2(1-2\mu_i)}{(1-\varepsilon_i)^5}\right] \\
&\times\left[\frac{\phi_i(1-2\mu_i)}{(1-\varepsilon_i)}w_i^* - 6(y_i^* - \delta_i^*)\right]\Bigg\} = \sum_{i=1}^n j_i
\end{aligned}$$

In matrix form, it can rewrite as

$$\ddot{\boldsymbol{U}}_p^{\alpha\alpha}(\boldsymbol{\theta}) = trace(\boldsymbol{J}).$$

The second derivative of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ with respect to the $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ is give by

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top} &= \sum_{i=1}^n\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{x}_i \\
&= \sum_{i=1}^n\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\boldsymbol{x}_i\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n(1-\hat{v}_i)\left\{\frac{(1-\alpha)}{(1-\varepsilon_i)^3}[\phi_i(\delta_i w_i^* - \Psi'((1-\delta_i)\phi_i)) + (y_i^* - \delta_i^*)]\right\} \\
&\times\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{x}_i\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^n r_i^*\boldsymbol{x}_i\boldsymbol{e}_i^\top.
\end{aligned}$$

In matrix form, the quantity can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}) = -\boldsymbol{X}^\top\boldsymbol{R}^*\boldsymbol{E}.$$

The second derivative of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ with respect to the $\boldsymbol{\gamma}_j$ and $\boldsymbol{\kappa}$ is give by

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\kappa}^\top} = \sum_{i=1}^{n}\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\frac{\partial\eta_{2i}}{\partial\boldsymbol{\kappa}^\top}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^{n}\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\phi_i}{\partial\eta_{2i}}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n}(1-\hat{\upsilon}_i)\left\{\frac{(1-\alpha)}{(1-\varepsilon_i)^3}[\phi_i(\delta_i w_i^* - \Psi'((1-\delta_i)\phi_i)) + (y_i^* - \delta_i^*)]\right\} \\
&\times \left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top \\
&= -\sum_{i=1}^{n} r_i^*\boldsymbol{b}_{ij}\boldsymbol{e}_i^\top.
\end{aligned}
$$

In matrix form, the quantity can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\boldsymbol{\kappa}}(\boldsymbol{\theta}) = -\boldsymbol{B}_j^\top\boldsymbol{R}^*\boldsymbol{E}.$$

The second derivative of $l_p(\boldsymbol{\theta},\boldsymbol{\lambda})$ with respect to $\boldsymbol{\gamma}_j$ and $\boldsymbol{\gamma}_{j'}$ is developed below,

$$
\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}^\top} = \sum_{i=1}^{n}\frac{\partial}{\partial\mu_i}\left(\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_{j'}^\top}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^{n}\left(\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= \sum_{i=1}^{n}\left\{\left[\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{\upsilon}_i}{\varepsilon_i^2} - \frac{(1-\hat{\upsilon}_i)}{(1-\varepsilon_i)^2}\right] + \left[\frac{\hat{\upsilon}_i}{\varepsilon_i} - \frac{(1-\hat{\upsilon}_i)}{(1-\varepsilon_i)}\right]\right. \\
&\times \left[-\frac{4\alpha}{(1-\varepsilon_i)} + \frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^3}\right] + (1-\alpha)(1-\hat{\upsilon}_i) \\
&\times \left[\frac{-(1-\alpha)\phi_i^2}{(1-\varepsilon_i)^6}w_i^* + \frac{6\alpha\phi_i(1-2\mu_i)}{(1-\varepsilon_i)^5}(y_i^* - \delta_i^*)\right] + f_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1} \\
&\times \left.\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top \\
&= \sum_{i=1}^{n} r_i\boldsymbol{b}_{ij}\boldsymbol{b}_{ij'}^\top.
\end{aligned}
$$

And in matrix form, we have that

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\gamma_{j'}}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{R}\boldsymbol{B}_{j'}.$$

The second derivative of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$ is developed below,

$$\begin{aligned}\ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) &= \frac{\partial^2 l_p(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_j^\top} = \sum_{i=1}^n \frac{\partial}{\partial\mu_i}\left(\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\frac{\partial\eta_{1i}}{\boldsymbol{\gamma}_j^\top}\boldsymbol{x}_i \\
&= \sum_{i=1}^n\left(\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_{1i}} + \frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\mu_i\partial\eta_{1i}}\right)\frac{\partial\mu_i}{\partial\eta_{1i}}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= \sum_{i=1}^n\left\{\left[\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right]+\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\right. \\
&\times\left[-\frac{4\alpha}{(1-\varepsilon_i)}+\frac{4\alpha^2(1-2\mu_i)^2}{(1-\varepsilon_i)^3}\right]+(1-\alpha)(1-\hat{v}_i) \\
&\times\left[\frac{-(1-\alpha)\phi_i^2}{(1-\varepsilon_i)^6}w_i^*+\frac{6\alpha\phi_i(1-2\mu_i)}{(1-\varepsilon_i)^5}(y_i^*-\delta_i^*)\right]+f_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1} \\
&\times\left.\left(\frac{\partial^2 g_1(\mu_i)}{\partial\mu_i^2}\right)\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-2}\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top \\
&= \sum_{i=1}^n r_i\boldsymbol{x}_i\boldsymbol{b}_{ij}^\top.\end{aligned}$$

Rewrite in matrix form, we obtain that

$$\ddot{\boldsymbol{U}}_p^{\beta\gamma_j}(\boldsymbol{\theta}) = \boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{B}_j.$$

The second derivative of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ with respect to the $\boldsymbol{\beta}$ and α is give by

$$\begin{aligned}\ddot{\boldsymbol{U}}_p^{\beta\alpha}(\boldsymbol{\theta}) &= \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\beta}\partial\alpha} = \sum_{i=1}^n\left\{\left[\frac{4\alpha\mu_i(1-\mu_i)(1-2\delta_i)}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right]\right. \\
&+\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\left\{2(1-2\mu_i)\left[\frac{1}{(1-\varepsilon_i)}+\frac{2\alpha\mu_i(1-\mu_i)}{(1-\varepsilon_i)^3}\right]\right\} \\
&+(1-\hat{v}_i)\left\{\left[\frac{(1-\alpha)\phi_i^2\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^6}w_i^*\right]+(y_i^*-\delta_i^*)\right. \\
&\times\left.\left.\left\{-\phi\left[\frac{1}{(1-\varepsilon_i)^4}-\frac{6(1-\alpha)\mu_i(1-\mu_i)}{(1-\varepsilon_i)^6}\right]\right\}\right\}\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{x}_i \\
&= \sum_{i=1}^n s_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{x}_i.\end{aligned}$$

In matrix form, the quantity can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\beta\alpha}(\boldsymbol{\theta}) = \boldsymbol{X}^\top\boldsymbol{T}_1\boldsymbol{s}^*.$$

The second derivative of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ with respect to the $\boldsymbol{\gamma}_j$ and α is give by

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\gamma_j\alpha}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\gamma}_j\partial\alpha} &= \sum_{i=1}^{n}\left\{\left[\frac{4\alpha\mu_i(1-\mu_i)(1-2\delta_i)}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right]\right. \\
&+\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\left\{2(1-2\mu_i)\left[\frac{1}{(1-\varepsilon_i)}+\frac{2\alpha\mu_i(1-\mu_i)}{(1-\varepsilon_i)^3}\right]\right\} \\
&+(1-\hat{v}_i)\left\{\left[\frac{(1-\alpha)\phi_i^2\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^6}w_i^*\right]+(y_i^*-\delta_i^*)\right. \\
&\left.\left.\times\left\{-\phi\left[\frac{1}{(1-\varepsilon_i)^4}-\frac{6(1-\alpha)\mu_i(1-\mu_i)}{(1-\varepsilon_i)^6}\right]\right\}\right\}\right\}\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{b}_{ij} \\
&= \sum_{i=1}^{n} s_i^*\left(\frac{\partial g_1(\mu_i)}{\partial\mu_i}\right)^{-1}\boldsymbol{b}_{ij}.
\end{aligned}$$

In matrix form, the quantity can be rewrite as

$$\ddot{\boldsymbol{U}}_p^{\gamma_j\alpha}(\boldsymbol{\theta}) = \boldsymbol{B}_j^\top\boldsymbol{T}_1\boldsymbol{s}^*.$$

The second derivative of $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ with respect to the $\boldsymbol{\kappa}$ and α is give by

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\kappa\alpha}(\boldsymbol{\theta}) = \frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})}{\partial\boldsymbol{\kappa}\partial\alpha} &= \sum_{i=1}^{n}\left[\frac{4\alpha\mu_i(1-\mu_i)(1-2\delta_i)}{(1-\varepsilon_i)^2}\right]\left[-\frac{\hat{v}_i}{\varepsilon_i^2}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)^2}\right] \\
&+\left[\frac{\hat{v}_i}{\varepsilon_i}-\frac{(1-\hat{v}_i)}{(1-\varepsilon_i)}\right]\left\{2(1-2\mu_i)\left[\frac{1}{(1-\varepsilon_i)}+\frac{2\alpha\mu_i(1-\mu_i)}{(1-\varepsilon_i)^3}\right]\right\} \\
&+(1-\hat{v}_i)\left\{\left[\frac{(1-\alpha)\phi_i^2\mu_i(1-\mu_i)(1-2\mu_i)}{(1-\varepsilon_i)^6}w_i^*\right]+(y_i^*-\delta_i^*)\right. \\
&\left.\times\left\{-\phi_i\left[\frac{1}{(1-\varepsilon_i)^4}-\frac{6(1-\alpha)\mu_i(1-\mu_i)}{(1-\varepsilon_i)^6}\right]\right\}\right\}\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1} \\
&= \sum_{i=1}^{n} c_i^*\left(\frac{\partial g_2(\phi_i)}{\partial\phi_i}\right)^{-1}\boldsymbol{e}_i,
\end{aligned}$$

in matrix form,

$$\ddot{\boldsymbol{U}}_p^{\kappa\alpha}(\boldsymbol{\theta}) = \boldsymbol{E}^\top\boldsymbol{T}_2\boldsymbol{c}^*.$$

A.4 Results related to discrete part of Zero-and/or-One Augmented Models

A.4.1 Penalized score function for Zero-and/or-One Augmented for discrete part of the Models

As all models share the same structure for discrete part, all development of calculations that involves these components, will be found in this and next section. Thus, given that is possible split the likelihood for $\boldsymbol{\theta}$, that is the vector containing all model parameters, into two parts, one that is related with the parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, which regards the discrete components and the other that is related with $\boldsymbol{\varphi}$, which regards the continuous components, it will be discussed here only about the subvector $\boldsymbol{\vartheta}$. In general, the penalized likehood is defined as

$$l(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l_1(\boldsymbol{\vartheta}) + l_2(\boldsymbol{\varphi}) - \sum_{j=1}^{k} \frac{\lambda_j}{2} \boldsymbol{\gamma}_j^\top \boldsymbol{\Lambda}_j^d \boldsymbol{\gamma}_j,$$

where, in particular,

$$l_1(\boldsymbol{\vartheta}) = \sum_{i=1}^{n} l_i(p_{0i}, p_{1i})$$

with

$$l_i(p_{0i}, p_{1i}) = z_i^*[(1 - y_i)\log(p_{0i}) + y_i \log(p_{1i})] + (1 - z_i^*)\log(1 - p_{0i} - p_{1i}),$$

with p_{0i} and p_{1i} defined, for example, in the Equation (3.2), because all models are using the same definition for this parameters.

Given that equations system $(\zeta_{0i}, \zeta_{1i}) = (h_0(p_{0i}, p_{1i}), h_1(p_{0i}, p_{1i}))$ defines a bijective transformation, it is possible solve in a unique way the equations $\zeta_{0i} = h_0(p_{0i}, p_{1i})$ and $\zeta_{1i} = h_0(p_{0i}, p_{1i})$ in terms of p_{0i} and p_{1i}. Denoting this inverse transformation as $p_{0i} = h_0^*(\zeta_{0i}, \zeta_{1i})$, $p_{1i} = h_1^*(\zeta_{0i}, \zeta_{1i})$, thus, it is possible to define

$$\begin{aligned}
\frac{\partial p_{0i}}{\partial \zeta_{0i}} &= \frac{\partial h_0^*(\zeta_{0i}, \zeta_{1i})}{\partial \zeta_{0i}}, \\
\frac{\partial p_{0i}}{\partial \zeta_{1i}} &= \frac{\partial h_0^*(\zeta_{0i}, \zeta_{1i})}{\partial \zeta_{1i}}, \\
\frac{\partial p_{1i}}{\partial \zeta_{0i}} &= \frac{\partial h_1^*(\zeta_{0i}, \zeta_{1i})}{\partial \zeta_{0i}}, \\
\frac{\partial p_{1i}}{\partial \zeta_{1i}} &= \frac{\partial h_1^*(\zeta_{0i}, \zeta_{1i})}{\partial \zeta_{1i}}.
\end{aligned}$$

Particularly,

$$\begin{aligned}
p_{0i} = h_0^*(\zeta_{0i}, \zeta_{1i}) &= \frac{e^{\zeta_{0i}}}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}}, \\
p_{1i} = h_1^*(\zeta_{0i}, \zeta_{1i}) &= \frac{e^{\zeta_{1i}}}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}}, \\
1 - p_{0i} - p_{1i} &= \frac{1}{1 + e^{\zeta_{0i}} + e^{\zeta_{1i}}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial p_{0i}}{\partial \zeta_{0i}} &= \left(\frac{e^{\zeta_{0i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right)\left(\frac{1+e^{\zeta_{1i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right) = p_{0i}(1-p_{0i}),\\
\frac{\partial p_{0i}}{\partial \zeta_{1i}} &= -\left(\frac{e^{\zeta_{0i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right)\left(\frac{e^{\zeta_{1i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right) = -p_{0i}p_{1i},\\
\frac{\partial p_{1i}}{\partial \zeta_{0i}} &= -\left(\frac{e^{\zeta_{0i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right)\left(\frac{e^{\zeta_{1i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right) = -p_{0i}p_{1i},\\
\frac{\partial p_{1i}}{\partial \zeta_{1i}} &= \left(\frac{e^{\zeta_{1i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right)\left(\frac{1+e^{\zeta_{0i}}}{1+e^{\zeta_{0i}}+e^{\zeta_{1i}}}\right) = p_{1i}(1-p_{1i}).
\end{aligned} \tag{A.5}$$

Thereby, based on the $l_1(\boldsymbol{\vartheta})$, the penalized score function for $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$ is given by $\dot{\boldsymbol{U}}_p^{\boldsymbol{\vartheta}}(\boldsymbol{\theta}) = (\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta})^\top, \dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta})^\top)^\top$, where $\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta})$ is defined as

$$\begin{aligned}
\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta}) = \frac{\partial l_1(\boldsymbol{\rho},\boldsymbol{\tau})}{\partial \boldsymbol{\rho}} &= \sum_{i=1}^{n}\left[\frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{0i}}\frac{\partial p_{0i}}{\partial \zeta_{0i}}\frac{\partial \zeta_{0i}}{\partial \boldsymbol{\rho}} + \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{1i}}\frac{\partial p_{1i}}{\partial \zeta_{0i}}\frac{\partial \zeta_{0i}}{\partial \boldsymbol{\rho}}\right]\\
&= \sum_{i=1}^{n}\left[\left(\frac{z_i^*(1-y_i)}{p_{0i}} - \frac{(1-z_i^*)}{1-p_{0i}-p_{1i}}\right)\frac{\partial p_{0i}}{\zeta_{0i}}\boldsymbol{f}_i + \left(\frac{z_i^* y_i}{p_{1i}} - \frac{(1-z_i^*)}{1-p_{0i}-p_{1i}}\right)\frac{\partial p_{1i}}{\zeta_{0i}}\boldsymbol{f}_i\right],
\end{aligned}$$

and replacing $\frac{\partial p_{0i}}{\zeta_{0i}}$ and $\frac{\partial p_{1i}}{\zeta_{0i}}$ by the quantities defined in the Equation (A.5) and making some algebrics manipulations, it is possible to reduce the result to:

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta}) = \sum_{i=1}^{n}[z_i^*(1-y_i) - p_{0i}]\boldsymbol{f}_i = \sum_{i=1}^{n} d_{1i}\boldsymbol{f}_i,$$

which in matrix form, can be rewrite as

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\rho}}(\boldsymbol{\theta}) = \boldsymbol{F}^\top \boldsymbol{d}_1,$$

and $\dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta})$ is defined as

$$\begin{aligned}
\dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = \frac{\partial l_1(\boldsymbol{\rho},\boldsymbol{\tau})}{\partial \boldsymbol{\tau}} &= \sum_{i=1}^{n}\left[\frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{0i}}\frac{\partial p_{0i}}{\partial \zeta_{1i}}\frac{\partial \zeta_{1i}}{\partial \boldsymbol{\tau}} + \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{1i}}\frac{\partial p_{1i}}{\partial \zeta_{1i}}\frac{\partial \zeta_{1i}}{\partial \boldsymbol{\tau}}\right]\\
&= \sum_{i=1}^{n}\left[\left(\frac{z_i^*(1-y_i)}{p_{0i}} - \frac{(1-z_i^*)}{1-p_{0i}-p_{1i}}\right)\frac{\partial p_{0i}}{\zeta_{1i}}\boldsymbol{m}_i + \left(\frac{z_i^* y_i}{p_{1i}} - \frac{(1-z_i^*)}{1-p_{0i}-p_{1i}}\right)\frac{\partial p_{1i}}{\zeta_{1i}}\boldsymbol{m}_i\right],
\end{aligned}$$

and replacing $\frac{\partial p_{0i}}{\zeta_{1i}}$ and $\frac{\partial p_{1i}}{\zeta_{1i}}$ by the quantities defined in the Equation (A.5) and making some algebrics manipulations, it is possible to reduce the result to:

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = \sum_{i=1}^{n}[z_i^* y_i - p_{1i}]\boldsymbol{m}_i = \sum_{i=1}^{n} d_{2i}\boldsymbol{m}_i,$$

which in matrix form, can be rewrite as

$$\dot{\boldsymbol{U}}_p^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = \boldsymbol{M}^\top \boldsymbol{d}_2.$$

A.4.2 Penalized Fisher's Information Matrix for Zero-and/or-One Augmented for discrete part of the Models

As the previously section, this one will also to approach about the discrete part of all models in the Section 3. Here, it will be presented the development to obtain second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\vartheta}$ and the Penalized Fisher's Information Matrix for discrete part of Zero-and/or-One Augmented models.

First, defining some important quantities that will be needed later, as

$$\begin{aligned}
\frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{0i}^2} &= -\frac{z_i^*(1-y_i)}{p_{0i}^2} - \frac{(1-z_i^*)}{(1-p_{0i}-p_{1i})^2} \\
\frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{1i}^2} &= -\frac{z_i^* y_i}{p_{1i}^2} - \frac{(1-z_i^*)}{(1-p_{0i}-p_{1i})^2} \\
\frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{0i}\partial p_{1i}} &= -\frac{(1-z_i^*)}{(1-p_{0i}-p_{1i})^2}.
\end{aligned} \tag{A.6}$$

And, under appropriate regularity conditions given in Lehmann and Casella (2002), it is given that

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{0i}}\right) &= 0 \\
\mathbb{E}\left(\frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{1i}}\right) &= 0 \\
\mathbb{E}\left(-\frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{0i}^2}\right) &= \frac{1}{p_{0i}} + \frac{1}{(1-p_{0i}-p_{1i})} \\
\mathbb{E}\left(-\frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{1i}^2}\right) &= \frac{1}{p_{1i}} + \frac{1}{(1-p_{0i}-p_{1i})} \\
\mathbb{E}\left(-\frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{0i}\partial p_{1i}}\right) &= \frac{1}{(1-p_{0i}-p_{1i})}
\end{aligned} \tag{A.7}$$

The second derivative of $l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\vartheta} = (\boldsymbol{\rho}^\top, \boldsymbol{\tau}^\top)^\top$, which is a subvector of θ related only to discrete part, is defined as

$$\ddot{\boldsymbol{U}}_p^{\vartheta\vartheta}(\boldsymbol{\theta}) = \frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\top} = \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\top},$$

which is equivalent to

$$\ddot{\boldsymbol{U}}_p^{\vartheta\vartheta}(\boldsymbol{\theta}) = \begin{pmatrix} \ddot{\boldsymbol{U}}_p^{\rho\rho}(\boldsymbol{\theta}) & \boldsymbol{U}_p^{\rho\tau}(\boldsymbol{\theta}) \\ \ddot{\boldsymbol{U}}_p^{\tau\rho}(\boldsymbol{\theta}) & \ddot{\boldsymbol{U}}_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix},$$

and the penalized Fisher's information matrix was defined as

$$\boldsymbol{K}_p^{\vartheta\vartheta}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\top}\right] = \mathbb{E}\left[-\frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}^\top}\right],$$

in which

$$\boldsymbol{K}_p^{\boldsymbol{\vartheta\vartheta}}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{K}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta}) \\ \boldsymbol{K}_p^{\boldsymbol{\tau\rho}}(\boldsymbol{\theta}) & \boldsymbol{K}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta}) \end{pmatrix}.$$

It is known that the penalized Fisher's information matrix $\boldsymbol{K}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta})$,

$$\boldsymbol{K}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}) = \mathbb{E}\left[-\frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}^\top}\right]$$

thus, the second derivative of $l_1(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\rho}$ is developed, using results of the Equations (A.5) and (A.6), below

$$\begin{aligned}\ddot{\boldsymbol{U}}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}) = \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}^\top} &= \sum_{i=1}^{n}\Bigg\{\left(\frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{0i}^2}\frac{\partial p_{0i}}{\partial\zeta_{0i}} + \frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{0i}\partial p_{1i}}\frac{\partial p_{1i}}{\partial\zeta_{0i}}\right)\frac{\partial p_{0i}}{\partial\zeta_{0i}}\frac{\partial\zeta_{0i}}{\partial\boldsymbol{\rho}^\top}\boldsymbol{f}_i \\ &+ \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{0i}}\frac{\partial^2 p_{0i}}{\partial\zeta_{0i}^2}\frac{\partial\zeta_{0i}}{\partial\boldsymbol{\rho}^\top}\boldsymbol{f}_i + \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{0i}}\frac{\partial p_{0i}}{\partial\zeta_{0i}}\frac{\partial^2\zeta_{0i}}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}^\top} \\ &+ \left(\frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{1i}^2}\frac{\partial p_{1i}}{\partial\zeta_{0i}} + \frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{0i}\partial p_{1i}}\frac{\partial p_{0i}}{\partial\zeta_{0i}}\right)\frac{\partial p_{1i}}{\partial\zeta_{0i}}\frac{\partial\zeta_{0i}}{\partial\boldsymbol{\rho}^\top}\boldsymbol{f}_i \\ &+ \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{1i}}\frac{\partial^2 p_{1i}}{\partial\zeta_{0i}^2}\frac{\partial\zeta_{0i}}{\partial\boldsymbol{\rho}^\top}\boldsymbol{\rho} + \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{1i}}\frac{\partial p_{1i}}{\partial\zeta_{0i}}\frac{\partial^2\zeta_{0i}}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}^\top}\Bigg\} \\ &= -\sum_{i=1}^{n} p_{0i}(1-p_{0i})\boldsymbol{f}_i\boldsymbol{f}_i^\top = -\sum_{i=1}^{n} d_{1i}^*\boldsymbol{f}_i\boldsymbol{f}_i^\top.\end{aligned}$$

In matrix form,

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}) = -\boldsymbol{F}^\top\boldsymbol{D}_{\mathbf{1}}^*\boldsymbol{F}.$$

As the the quantity $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta})$ do not depend on any random variable, the $\boldsymbol{K}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta})$ is defined as

$$\boldsymbol{K}_p^{\boldsymbol{\rho\rho}}(\boldsymbol{\theta}) = \boldsymbol{F}^\top\boldsymbol{D}_{\mathbf{1}}^*\boldsymbol{F}.$$

Now,the second derivative of $l_1(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\tau}$ is developed, using results of the Equations (A.5) and (A.6), below

$$\begin{aligned}\ddot{\boldsymbol{U}}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta}) = \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^\top} &= \sum_{i=1}^{n}\Bigg\{\left(\frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{0i}^2}\frac{\partial p_{0i}}{\partial\zeta_{1i}} + \frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{0i}\partial p_{1i}}\frac{\partial p_{1i}}{\partial\zeta_{1i}}\right)\frac{\partial p_{0i}}{\partial\zeta_{1i}}\frac{\partial\zeta_{1i}}{\partial\boldsymbol{\tau}^\top}\boldsymbol{m}_i \\ &+ \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{0i}}\frac{\partial^2 p_{0i}}{\partial\zeta_{1i}^2}\frac{\partial\zeta_{1i}}{\partial\boldsymbol{\tau}^\top}\boldsymbol{m}_i + \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{0i}}\frac{\partial p_{0i}}{\partial\zeta_{1i}}\frac{\partial^2\zeta_{1i}}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^\top} \\ &+ \left(\frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{1i}^2}\frac{\partial p_{1i}}{\partial\zeta_{1i}} + \frac{\partial^2 l_i(p_{0i},p_{1i})}{\partial p_{0i}\partial p_{1i}}\frac{\partial p_{0i}}{\partial\zeta_{1i}}\right)\frac{\partial p_{1i}}{\partial\zeta_{1i}}\frac{\partial\zeta_{1i}}{\partial\boldsymbol{\tau}^\top}\boldsymbol{m}_i \\ &+ \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{1i}}\frac{\partial^2 p_{1i}}{\partial\zeta_{1i}^2}\frac{\partial\zeta_{1i}}{\partial\boldsymbol{\tau}^\top}\boldsymbol{\tau} + \frac{\partial l_i(p_{0i},p_{1i})}{\partial p_{1i}}\frac{\partial p_{1i}}{\partial\zeta_{1i}}\frac{\partial^2\zeta_{1i}}{\partial\boldsymbol{\tau}\partial\boldsymbol{\tau}^\top}\Bigg\} \\ &= -\sum_{i=1}^{n} p_{1i}(1-p_{1i})\boldsymbol{m}_i\boldsymbol{m}_i^\top = -\sum_{i=1}^{n} d_{2i}^*\boldsymbol{m}_i\boldsymbol{m}_i^\top.\end{aligned}$$

In matrix form,

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta}) = -\boldsymbol{M}^\top \boldsymbol{D}_{\boldsymbol{2}}^* \boldsymbol{M}.$$

As the the quantity $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta})$ do not depend on any random variable, the $\boldsymbol{K}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta})$ is defined as

$$\boldsymbol{K}_p^{\boldsymbol{\tau\tau}}(\boldsymbol{\theta}) = \boldsymbol{M}^\top \boldsymbol{D}_{\boldsymbol{2}}^* \boldsymbol{M}.$$

The cross derivative of $l_1(\boldsymbol{\vartheta})$ with respect to $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$ is presented, also using results of the Equations (A.5) and (A.6), below below

$$\begin{aligned}
\ddot{\boldsymbol{U}}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta}) = \frac{\partial^2 l_1(\boldsymbol{\vartheta})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\tau}^\top} &= \sum_{i=1}^n \Bigg\{ \left(\frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{0i}^2} \frac{\partial p_{0i}}{\partial \zeta_{1i}} + \frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{0i} \partial p_{1i}} \frac{\partial p_{1i}}{\partial \zeta_{1i}} \right) \frac{\partial p_{0i}}{\partial \zeta_{0i}} \frac{\partial \zeta_{1i}}{\partial \boldsymbol{\tau}^\top} \boldsymbol{m}_i \\
&+ \frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{0i}} \frac{\partial^2 p_{0i}}{\partial \zeta_{1i} \partial \zeta_{0i}} \frac{\partial \zeta_{1i}}{\partial \boldsymbol{\tau}^\top} \boldsymbol{f}_i + \frac{\partial l_i(p_{0i}, p_{1i})}{\partial p_{1i}} \frac{\partial^2 p_{1i}}{\partial \zeta_{0i} \partial \zeta_{1i}} \frac{\partial \zeta_{1i}}{\partial \boldsymbol{\tau}^\top} \boldsymbol{f}_i \\
&+ \left(\frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{1i}^2} \frac{\partial p_{1i}}{\partial \zeta_{1i}} + \frac{\partial^2 l_i(p_{0i}, p_{1i})}{\partial p_{0i} \partial p_{1i}} \frac{\partial p_{0i}}{\partial \zeta_{1i}} \right) \frac{\partial p_{1i}}{\partial \zeta_{0i}} \frac{\partial \zeta_{1i}}{\partial \boldsymbol{\tau}^\top} \boldsymbol{m}_i \\
&= -\sum_{i=1}^n -p_{0i} p_{1i} \boldsymbol{m}_i \boldsymbol{m}_i^\top = -\sum_{i=1}^n d_{3i}^* \boldsymbol{f}_i \boldsymbol{m}_i^\top.
\end{aligned}$$

In matrix form,

$$\ddot{\boldsymbol{U}}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta}) = -\boldsymbol{F}^\top \boldsymbol{D}_{\boldsymbol{3}}^* \boldsymbol{M}.$$

As the the quantity $\ddot{\boldsymbol{U}}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta})$ do not depend on any random variable, the $\boldsymbol{K}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta})$ is defined as

$$\boldsymbol{K}_p^{\boldsymbol{\rho\tau}}(\boldsymbol{\theta}) = \boldsymbol{F}^\top \boldsymbol{D}_{\boldsymbol{3}}^* \boldsymbol{M}.$$

A.5 Effective degrees of freedom for BR-SPAM and ZOABR-SPAM

The expression for the effective degrees of freedom is based on the estimated linear predictor (see Hastie and Tibshirani (1990) and Green and Silverman (1993)), then it is necessary an analytical form for this quantity. Given that, for the parameter estimation of the BR-SPAM (Chapter 2) and the continuous part of ZOABR-SPAM (Chapter 3), we used the L-BFGS-B algorithm, which has only numerical results for the estimated predictors. Then, we can approximate the linear predictor(s) of interest by using the Newton-Raphson algorithm at the convergence step. That is, once we obtain the estimates through the L-BFGS-B algorithm, we replaced the estimates on the expression of the linear predictor given by the Newton-Raphson algorithm, which is defined below

$$\hat{\boldsymbol{\eta}}_1 = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \sum_{j=1}^k \boldsymbol{B}_j \hat{\boldsymbol{\gamma}}_j = \widehat{\boldsymbol{H}}(\boldsymbol{\lambda}) \hat{\boldsymbol{z}}^{**},$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^\top$, $\boldsymbol{z}^{**} = \boldsymbol{N}^*\boldsymbol{\xi} - (\dot{\boldsymbol{R}})^{-1}\boldsymbol{T}_1\boldsymbol{f}^*$, $\boldsymbol{H}(\boldsymbol{\lambda}) = \boldsymbol{N}^* \left[\boldsymbol{N}^{*\top}\dot{\boldsymbol{R}}\boldsymbol{N}^* - \boldsymbol{\Lambda}(\boldsymbol{\lambda})\right]^{-1}$ $\boldsymbol{N}^{*\top}\dot{\boldsymbol{R}}$, $\boldsymbol{N}^*\boldsymbol{\xi} = [\boldsymbol{X}, \boldsymbol{B}_1, \dots, \boldsymbol{B}_k]\boldsymbol{\xi}$, $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \dots, \boldsymbol{\gamma}_k^\top)^\top$. Also, $\dot{\boldsymbol{R}} = \boldsymbol{R}$ for the BR-SPAM and $\dot{\boldsymbol{R}} = \boldsymbol{R}^\dagger$ for the ZOABR-SPAM. Then, based on that the effective degrees of freedom for $\boldsymbol{\gamma}_j, j = 1, \dots, k$, following Eilers and Marx (1996), are obtained as the trace of $\boldsymbol{H}(\boldsymbol{\lambda})$.

APPENDIX B – Simulation results

B.1 Simulation results of Chapter 2: Study 1

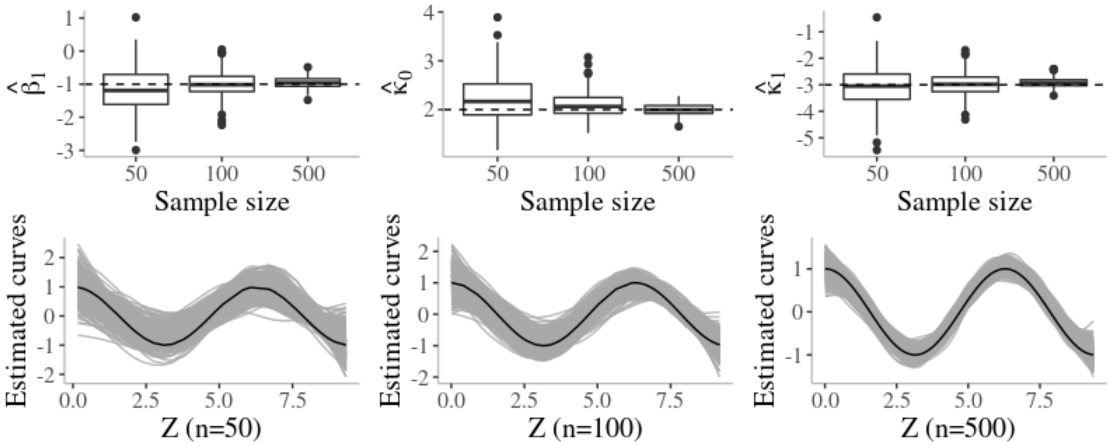


Figure 34 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Beta-SPAM with logit as link function for mean by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

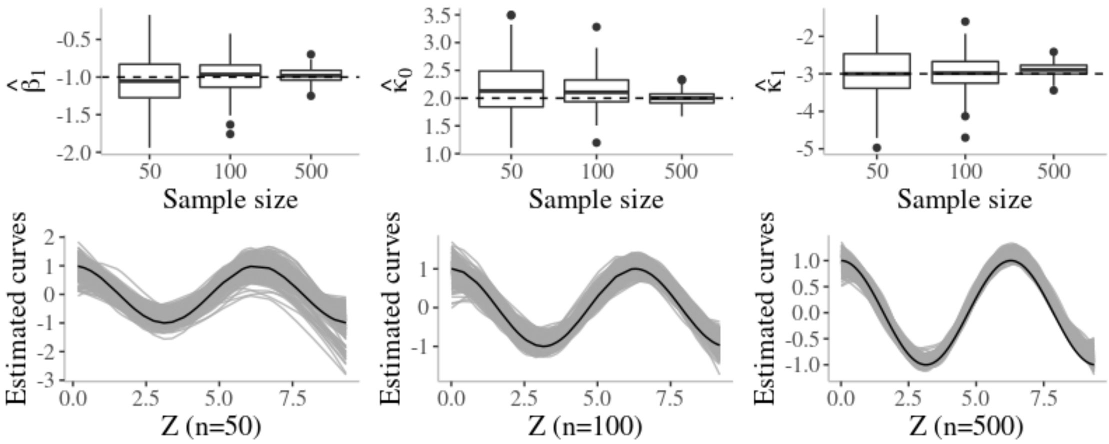


Figure 35 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Beta-SPAM with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

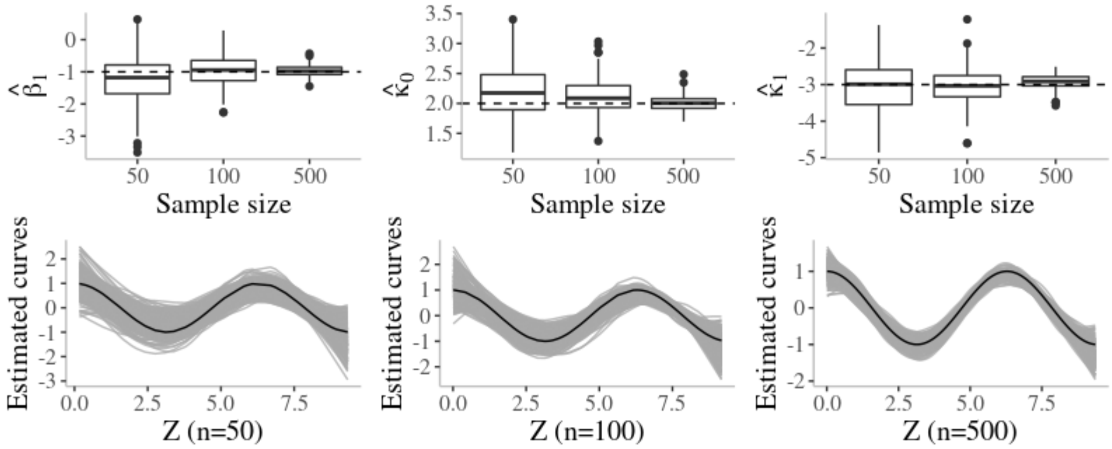


Figure 36 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Beta-SPAM with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

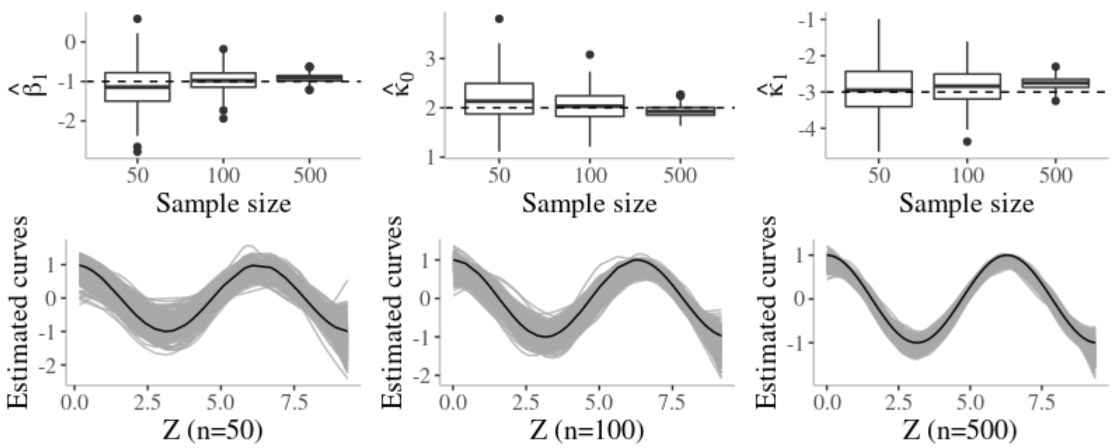


Figure 37 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Beta-SPAM with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

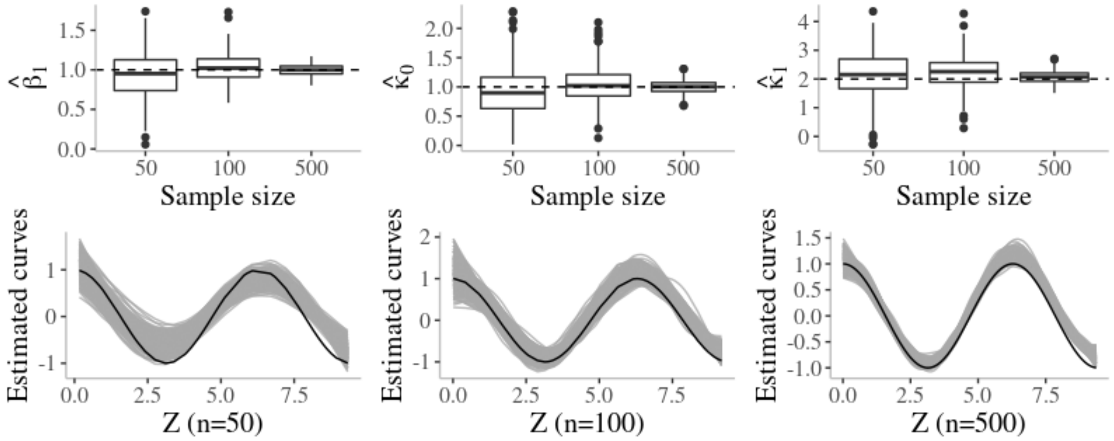


Figure 38 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Beta-SPAM with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

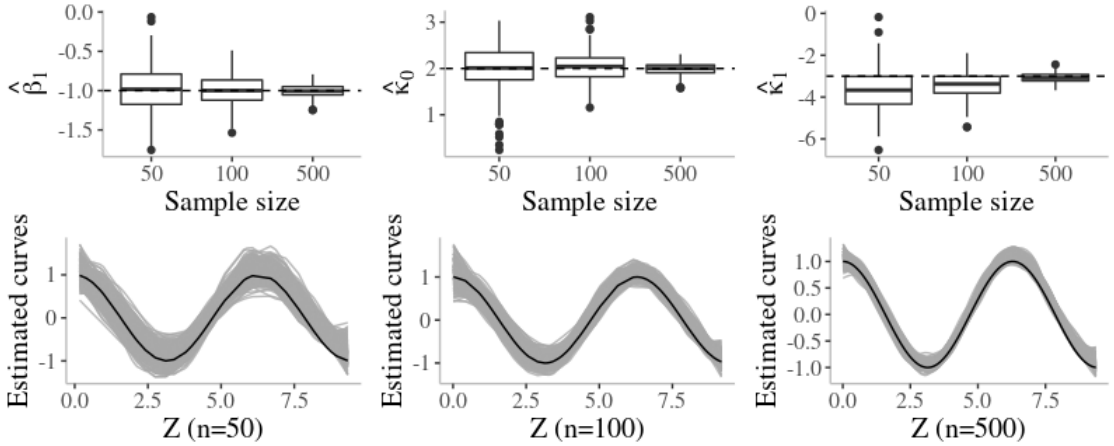


Figure 39 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Simplex-SPAM with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

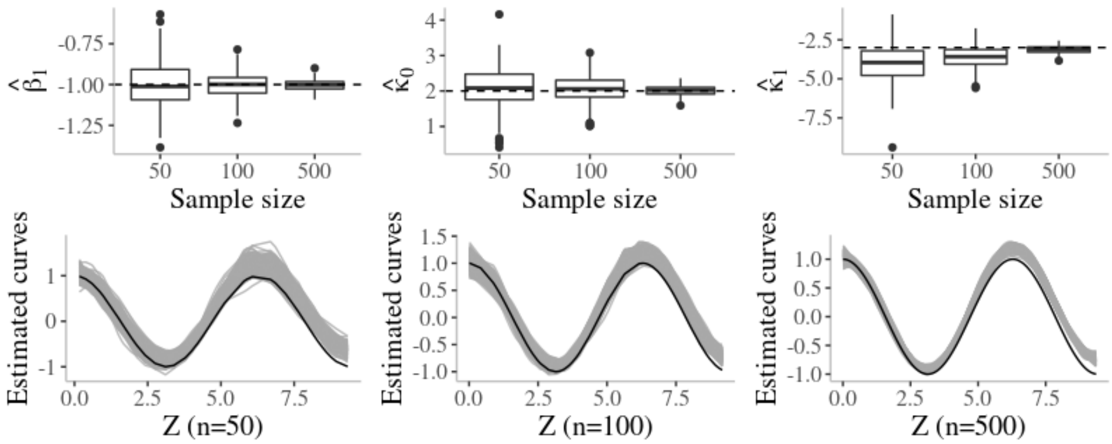


Figure 40 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Simplex-SPAM with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

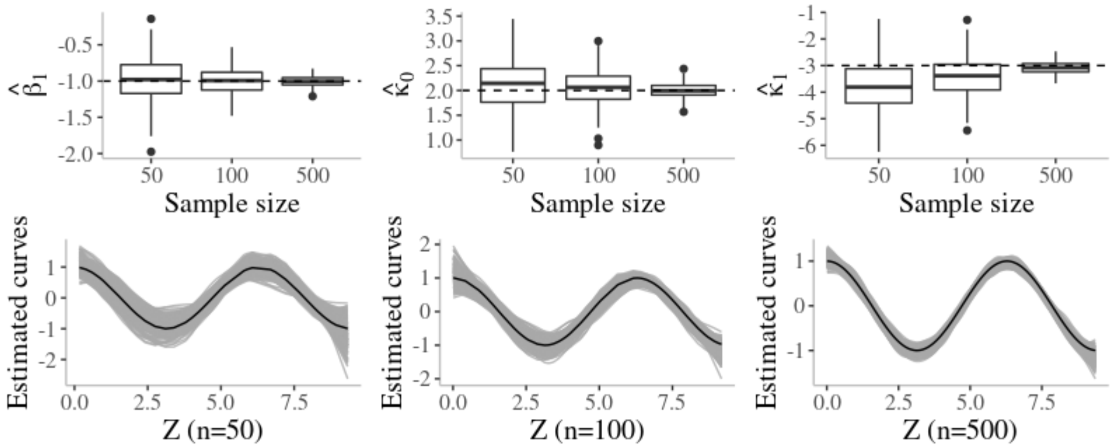


Figure 41 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Simplex-SPAM with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

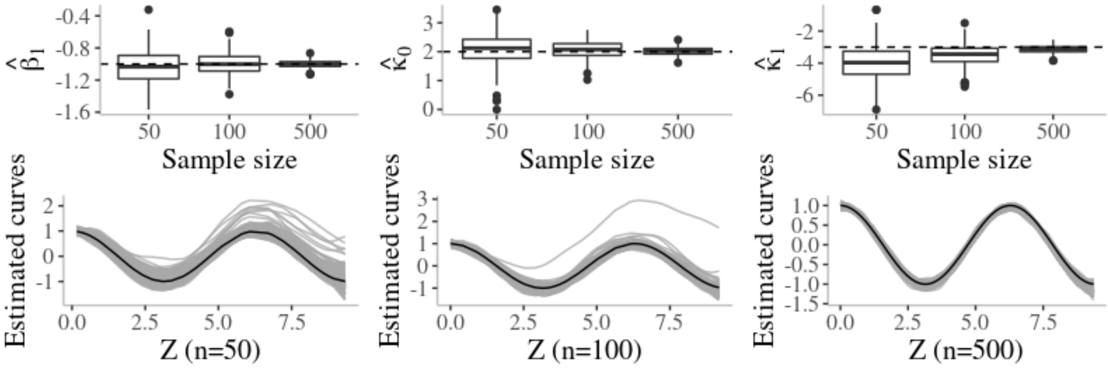


Figure 42 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the Simplex-SPAM with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

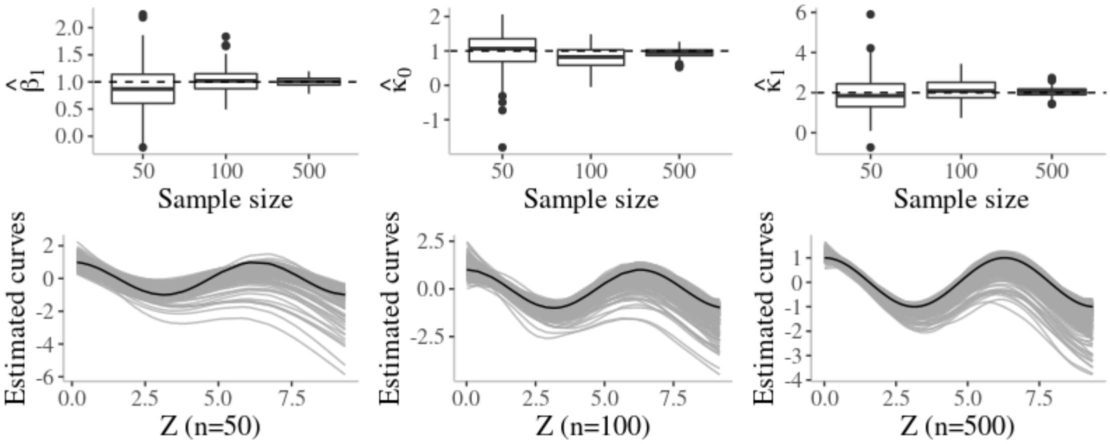


Figure 43 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 for the model based on distribution Simplex with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

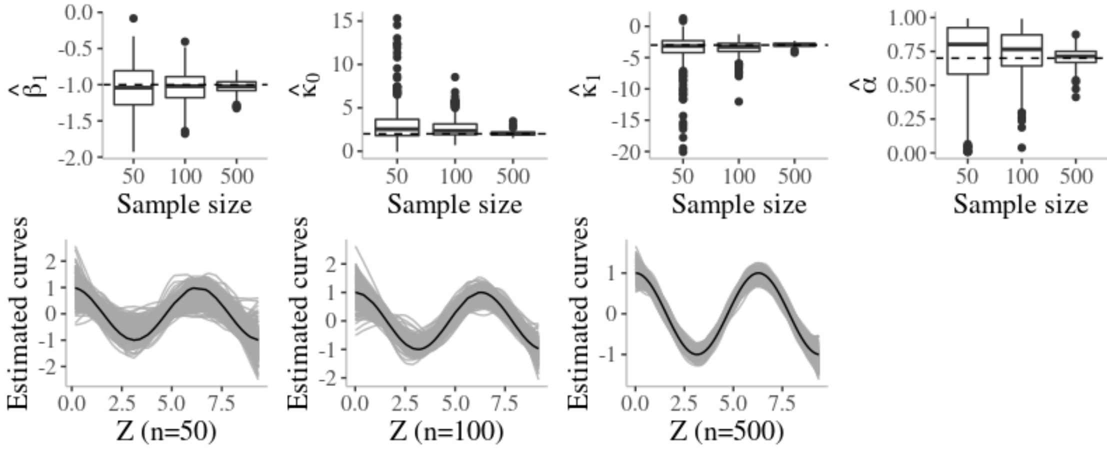


Figure 44 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and α for the BRSPAM with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

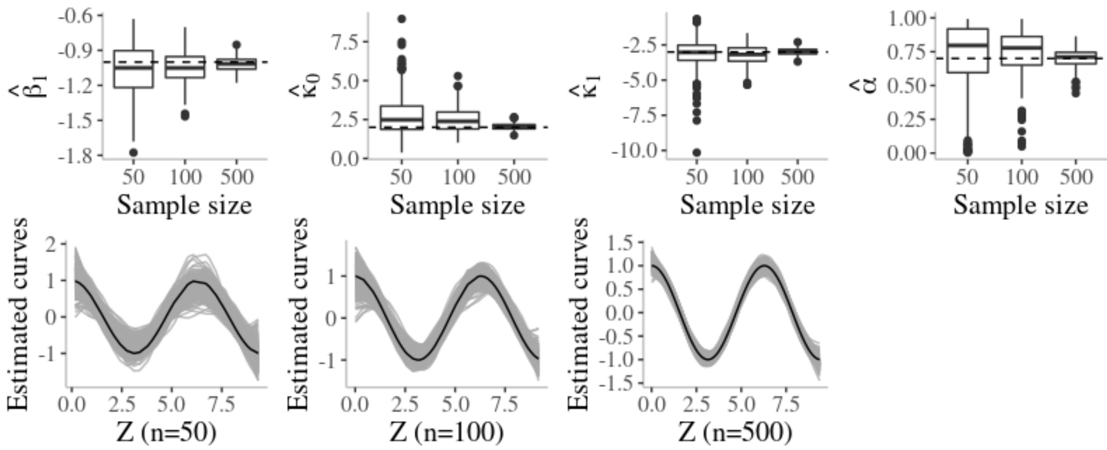


Figure 45 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and α for the BRSPAM with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

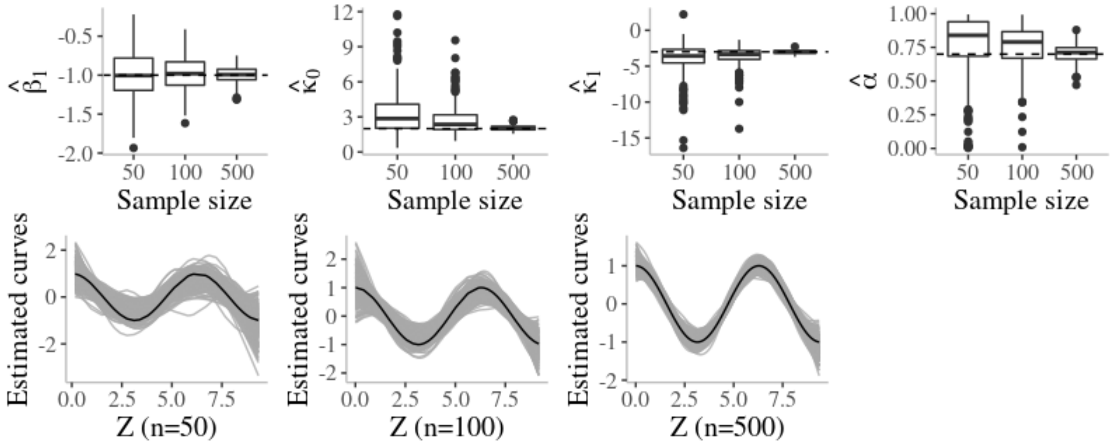


Figure 46 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and α for the BR-SPAM with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

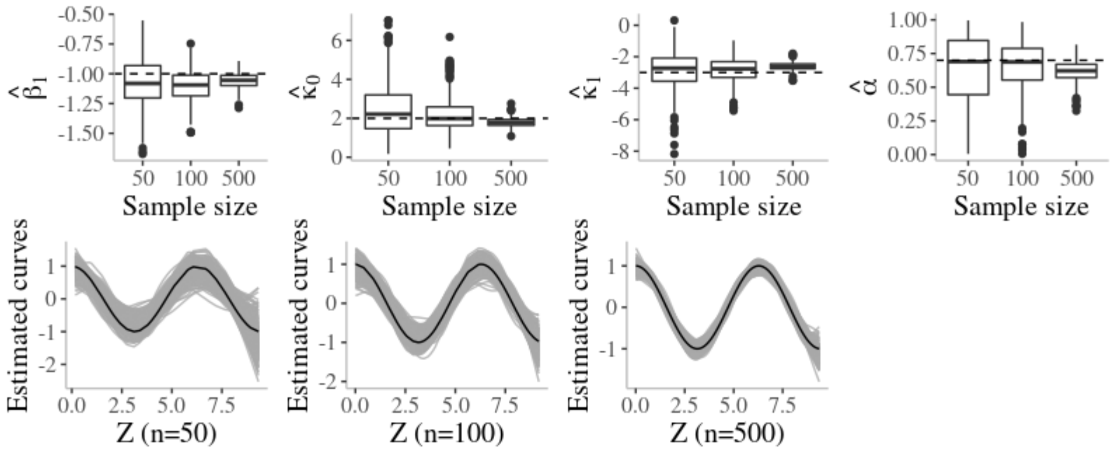


Figure 47 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and α for the BR-SPAM with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

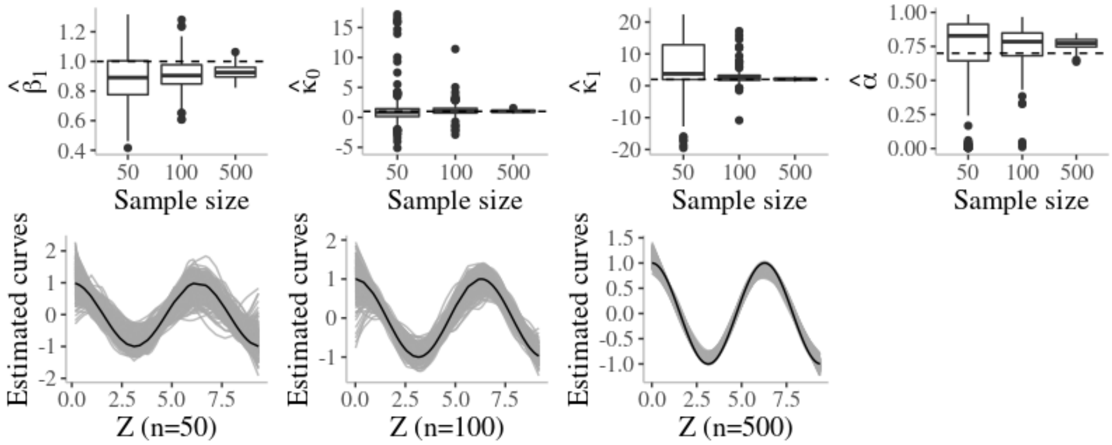


Figure 48 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and α for the BR-SPAM with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

B.2 Simulation results of Chapter 3: Study 1

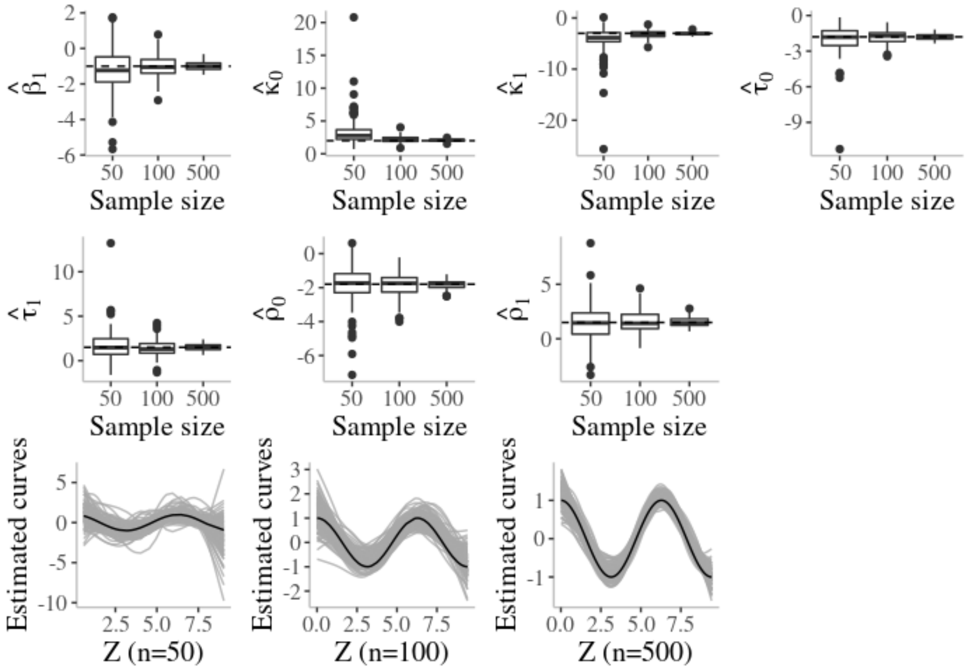


Figure 49 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAB-SPAM with logit as link function for mean by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

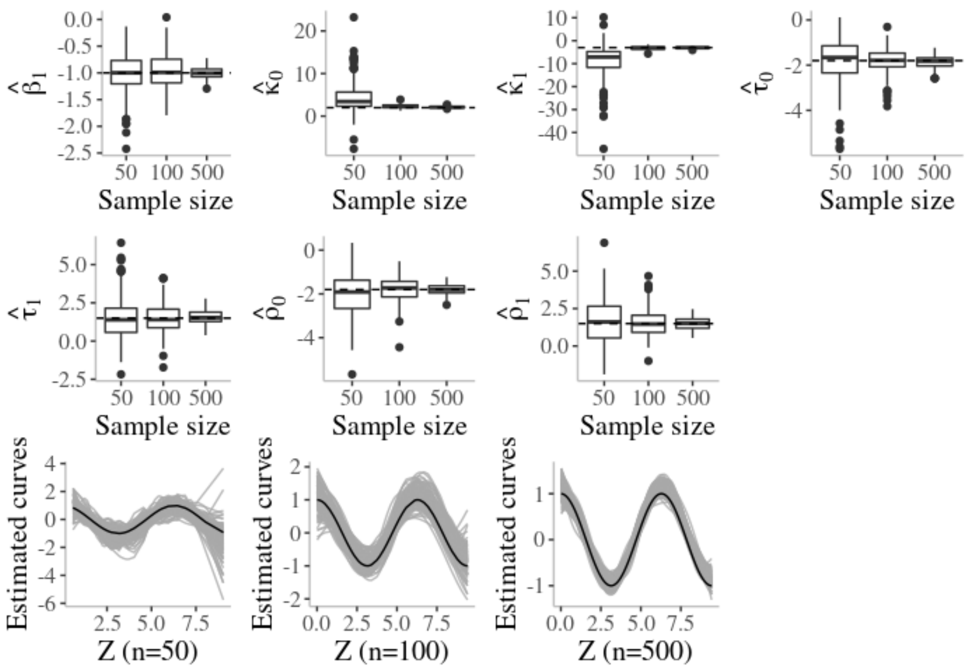


Figure 50 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAB-SPAM with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

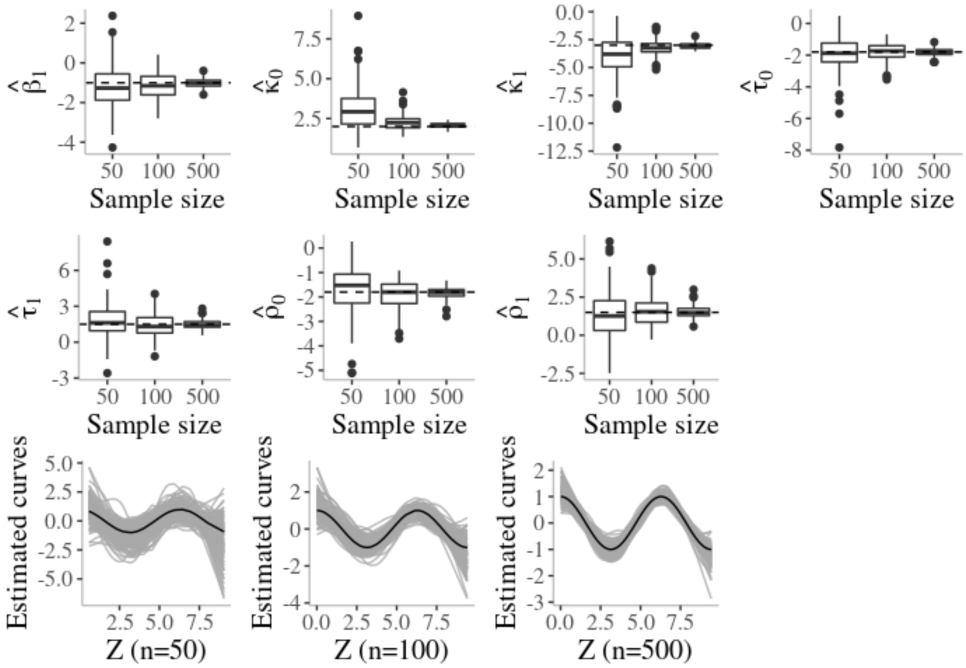


Figure 51 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAB-SPAM with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

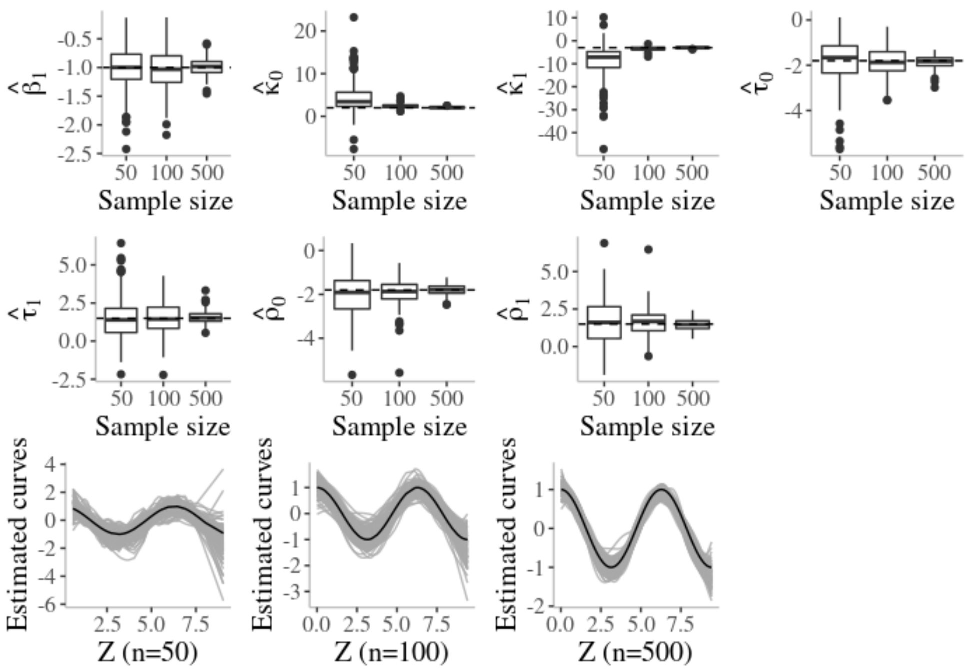


Figure 52 – The first plots are the estimates of parametersβ_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAB-SPAM with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

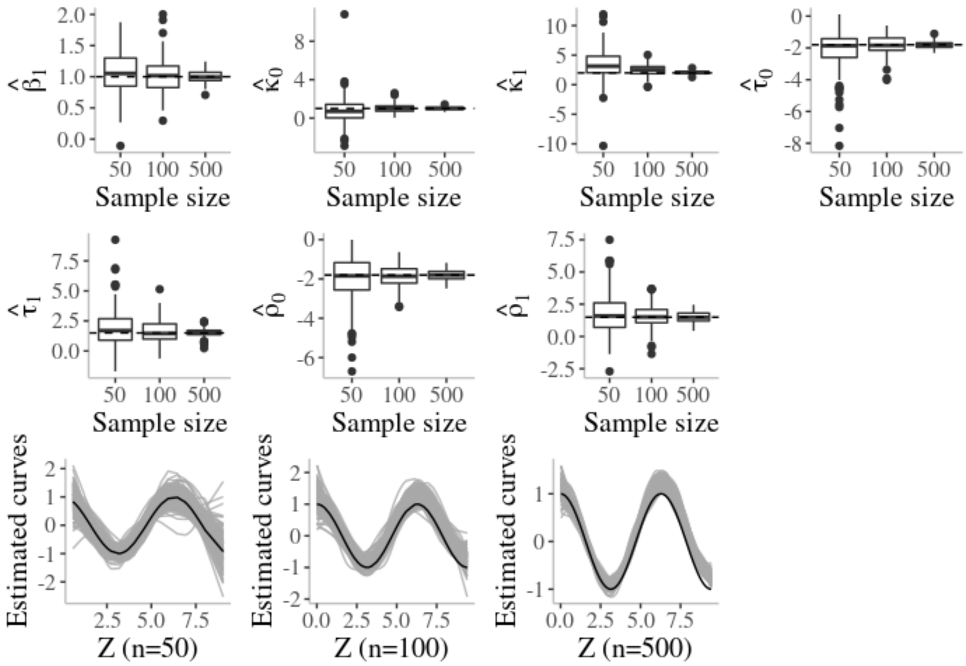


Figure 53 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAB-SPAM with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

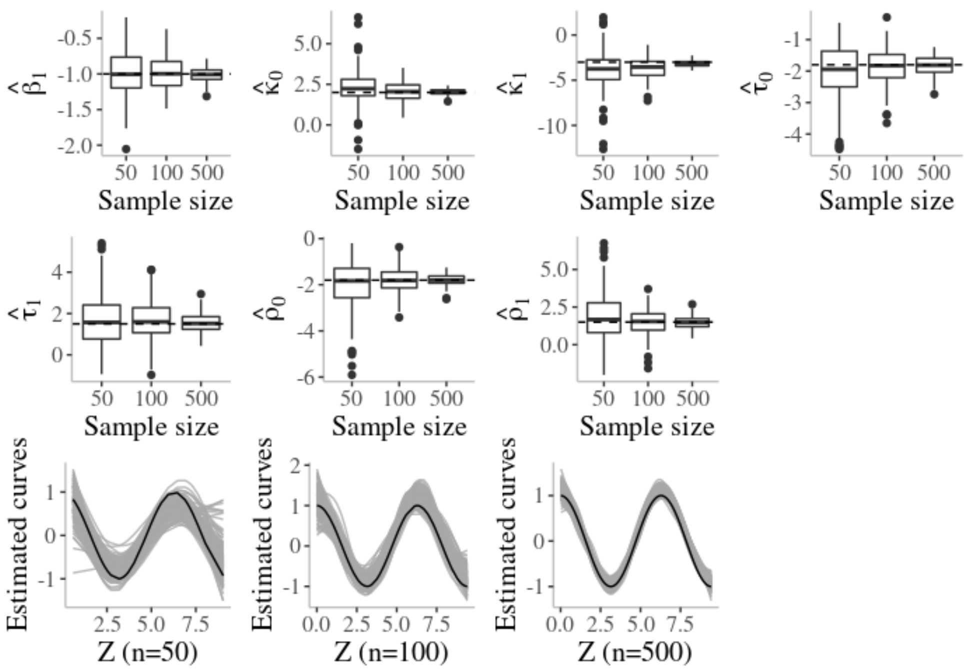


Figure 54 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAS-SPAM with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

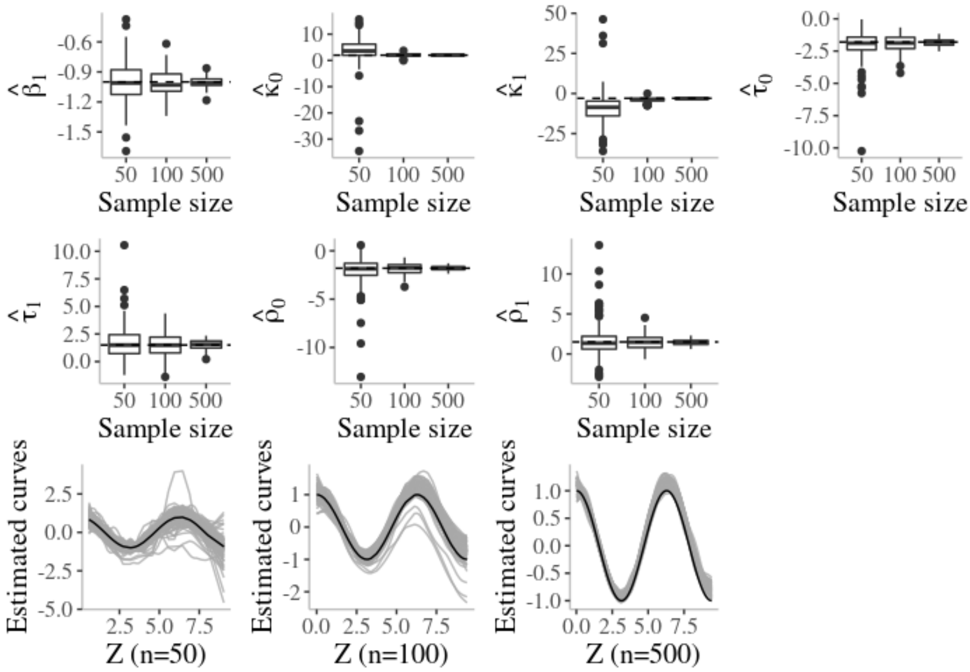


Figure 55 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAS-SPAM with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

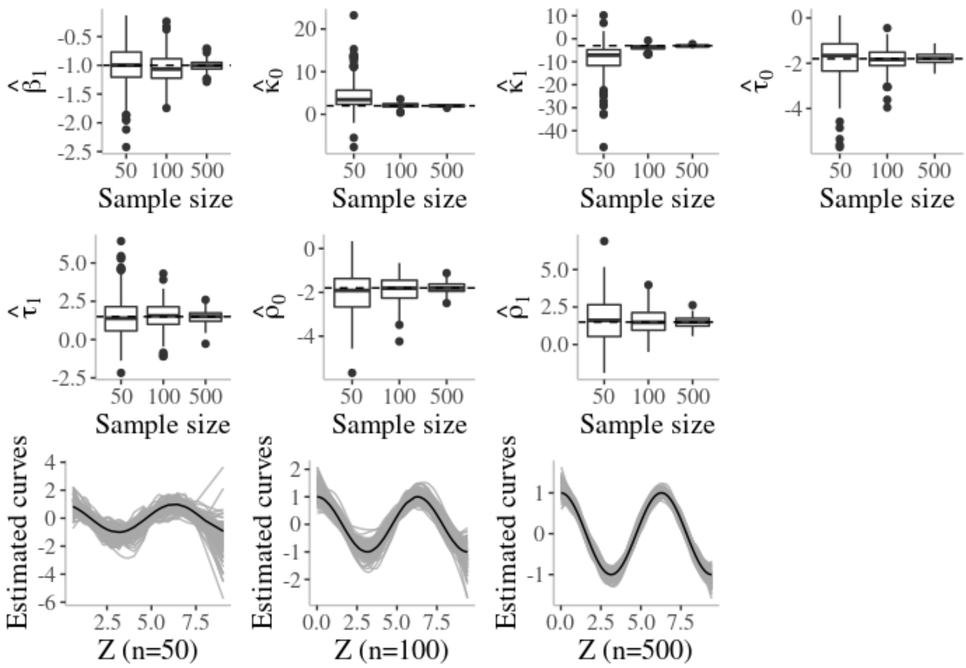


Figure 56 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAS-SPAM with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

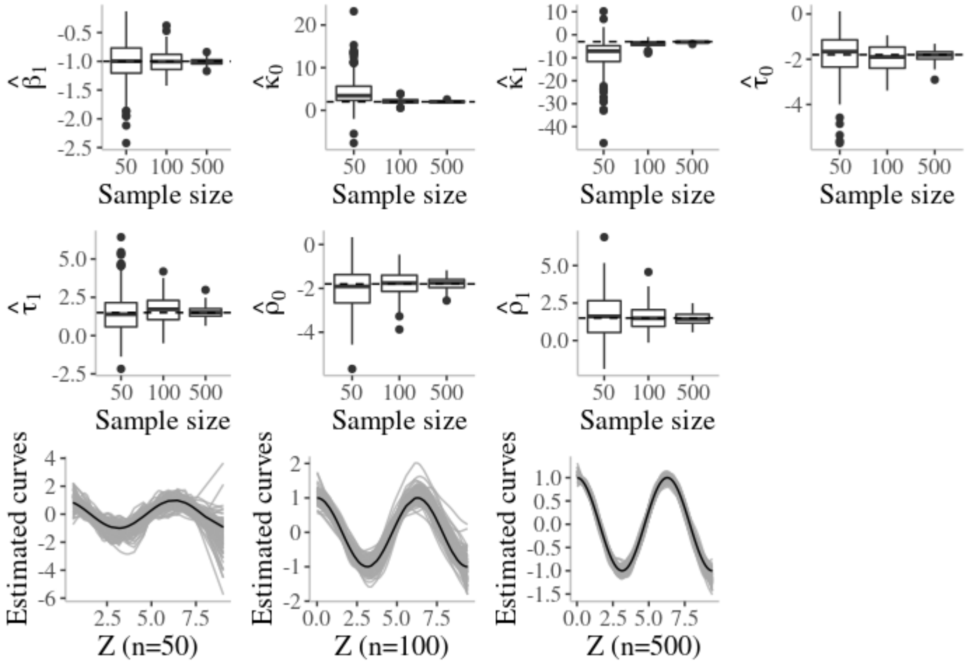


Figure 57 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAS-SPAM with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

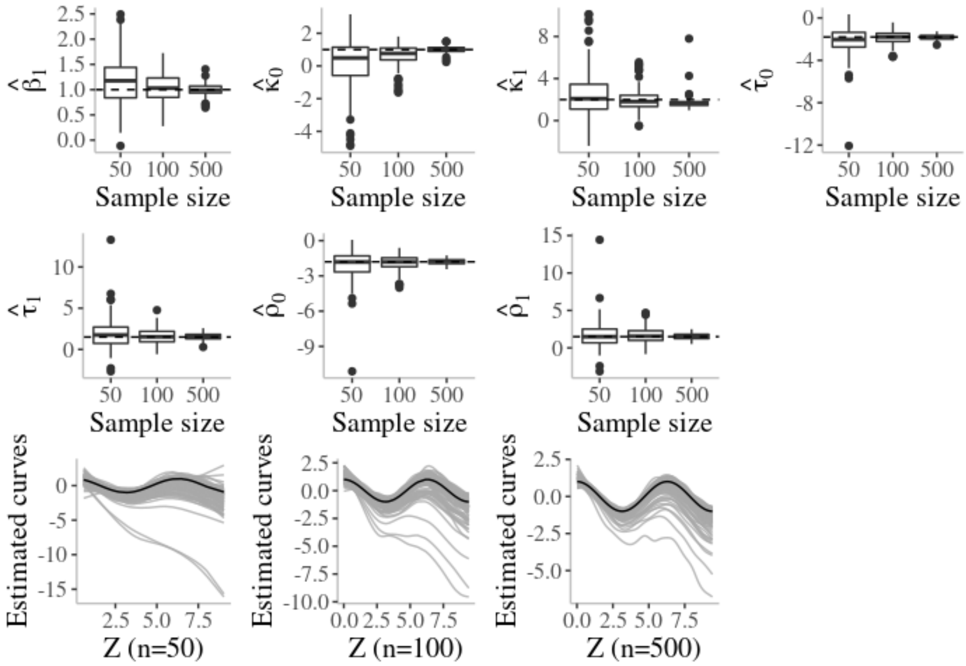


Figure 58 – The first plots are the estimates of parameters β_1, κ_0 and κ_1 β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0 and τ_1 for ZOAS-SPAM with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

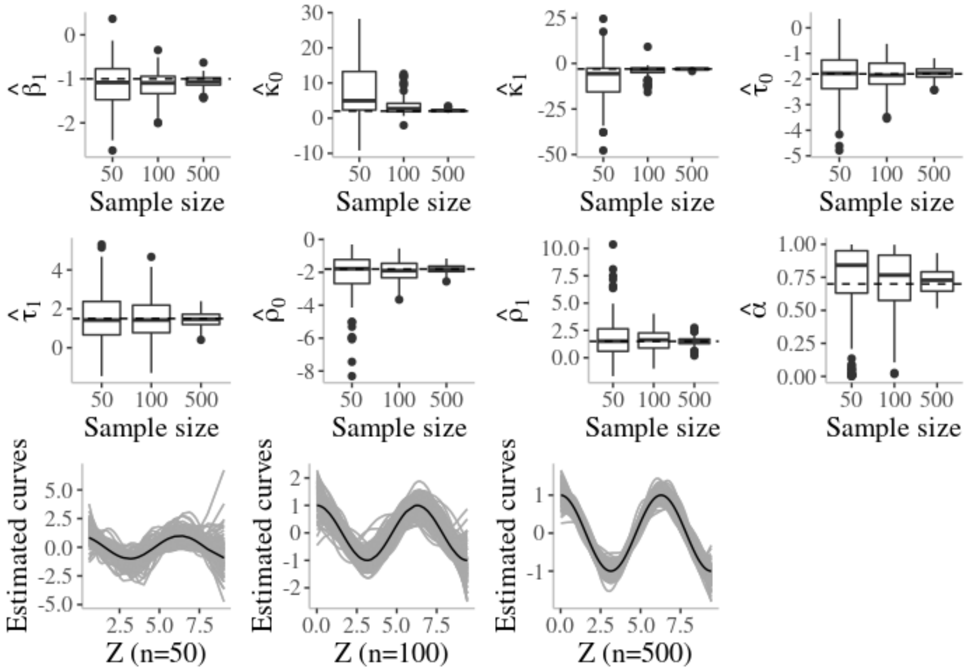


Figure 59 – The first plots are the estimates of parameters β_1, κ_0, κ_1, τ_0, τ_1,ρ_0, ρ_1 and α for ZOABR-SPAM with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

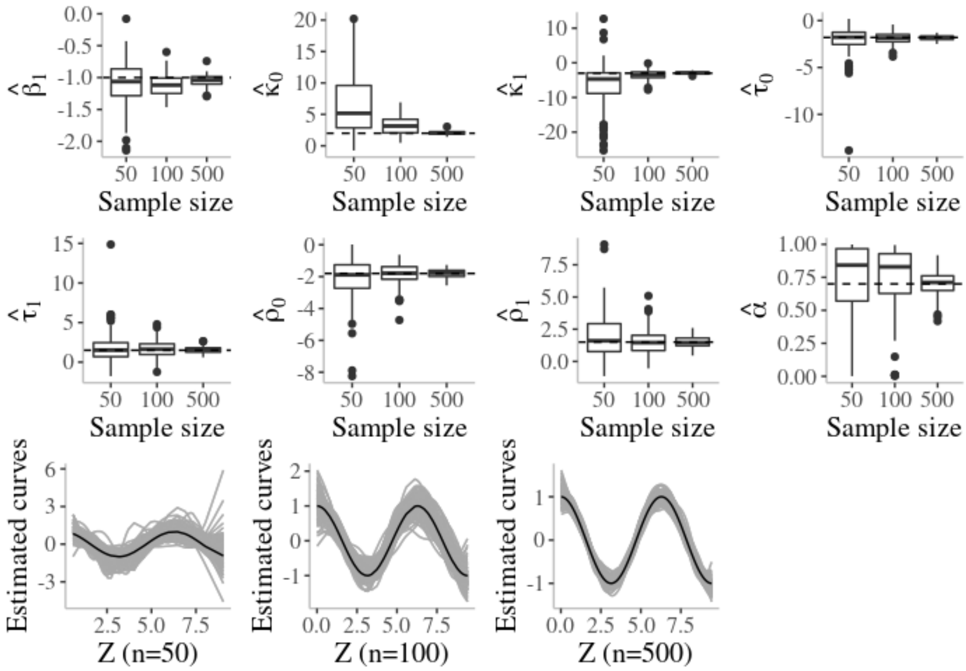


Figure 60 – The first plots are the estimates of parameters β_1, κ_0, κ_1, τ_0, τ_1,ρ_0, ρ_1 and α for ZOABR-SPAM with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

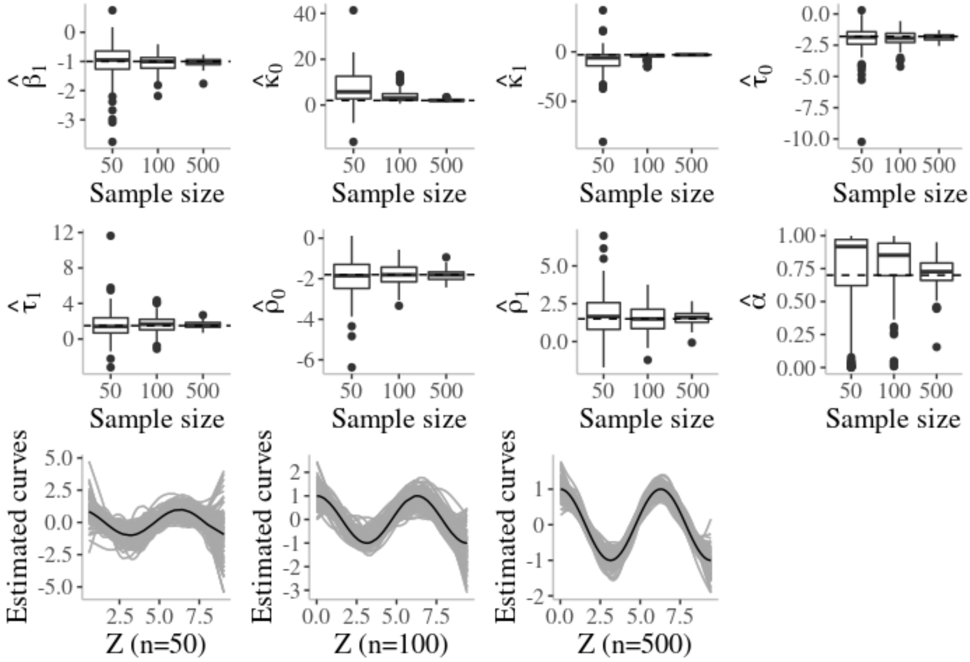


Figure 61 – The first plots are the estimates of parameters β_1, κ_0, κ_1, $\tau_0, \tau_1, \rho_0, \rho_1$ and α for ZOABR-SPAM with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

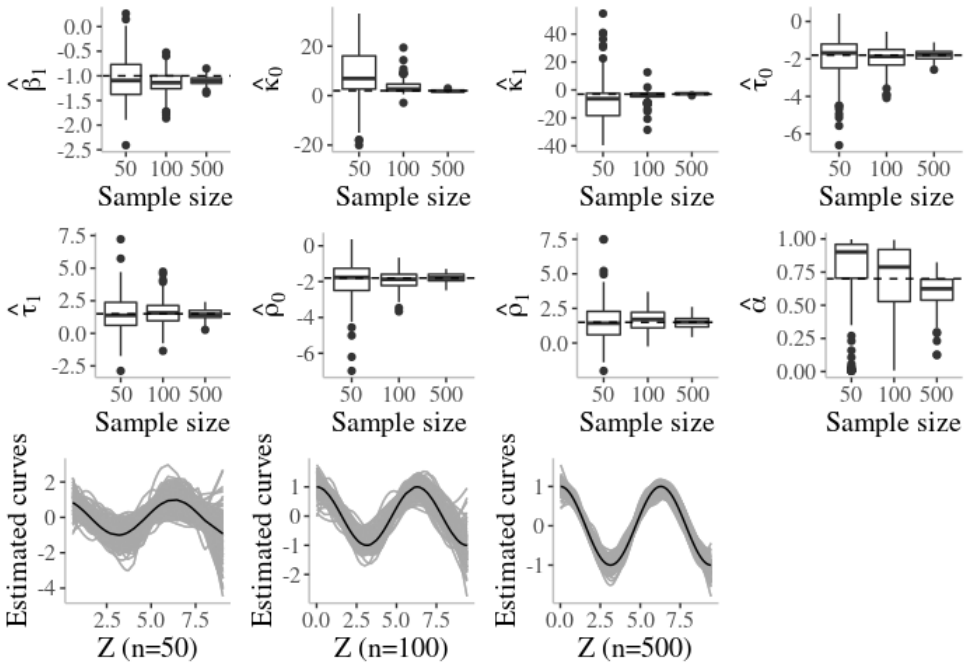


Figure 62 – The first plots are the estimates of parameters β_1, κ_0, κ_1, $\tau_0, \tau_1, \rho_0, \rho_1$ and α for ZOABR-SPAM with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

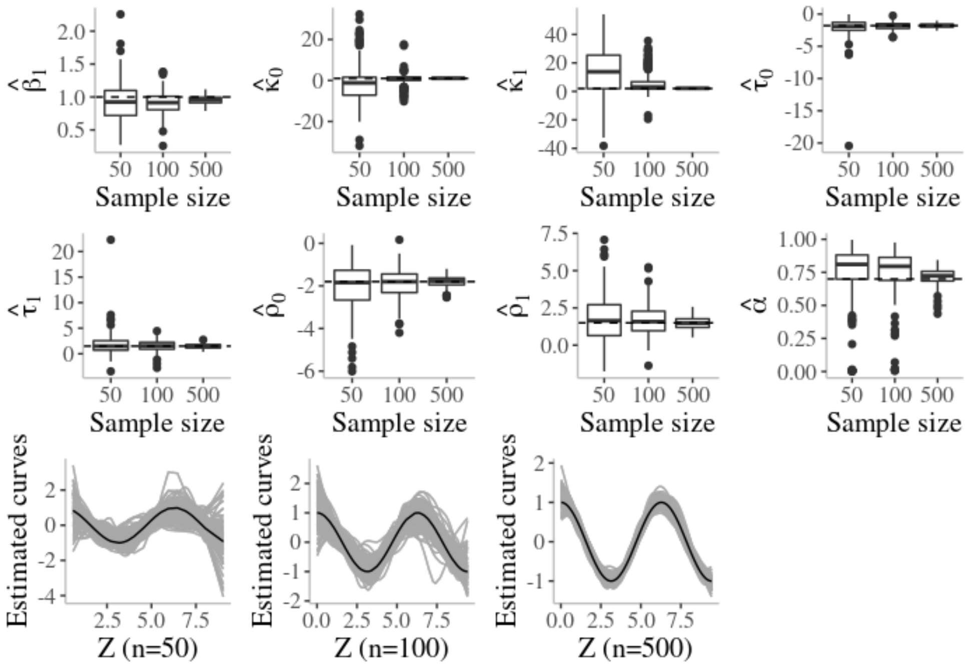


Figure 63 – The first plots are the estimates of parameters $\beta_1, \kappa_0,\ \kappa_1,\ \tau_0, \tau_1, \rho_0, \rho_1$ and α for ZOABR-SPAM with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

B.3 Simulation results of Chapter 4: Study 1

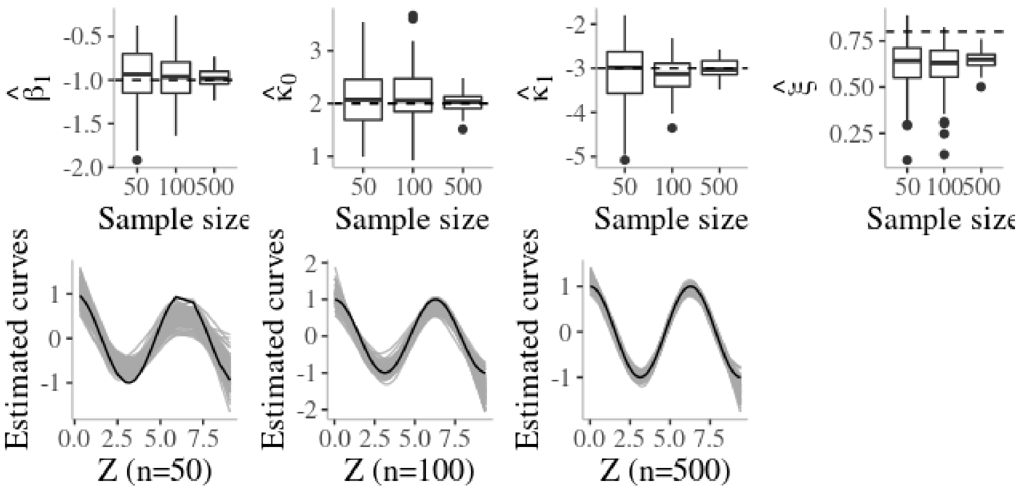


Figure 64 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Beta-SPAM-GEE with logit as link function for mean by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

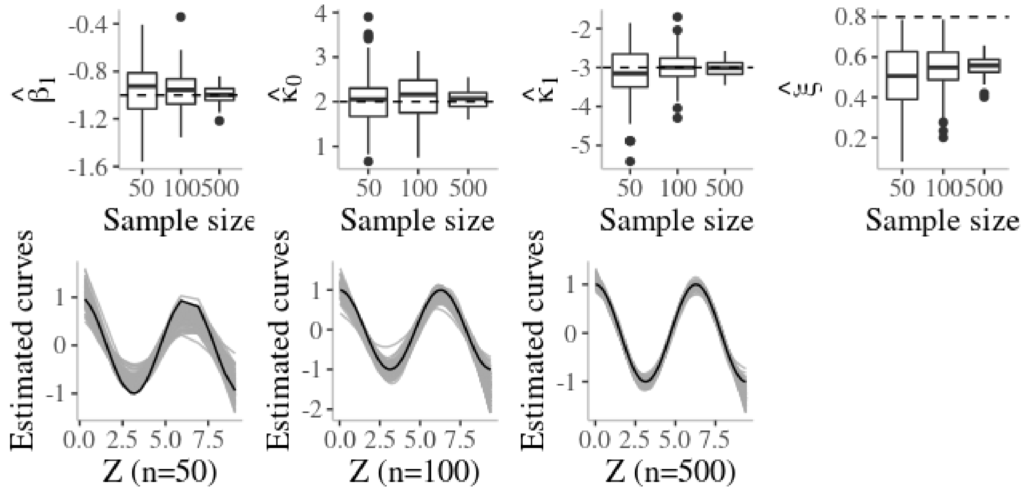


Figure 65 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Beta-SPAM-GEE with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

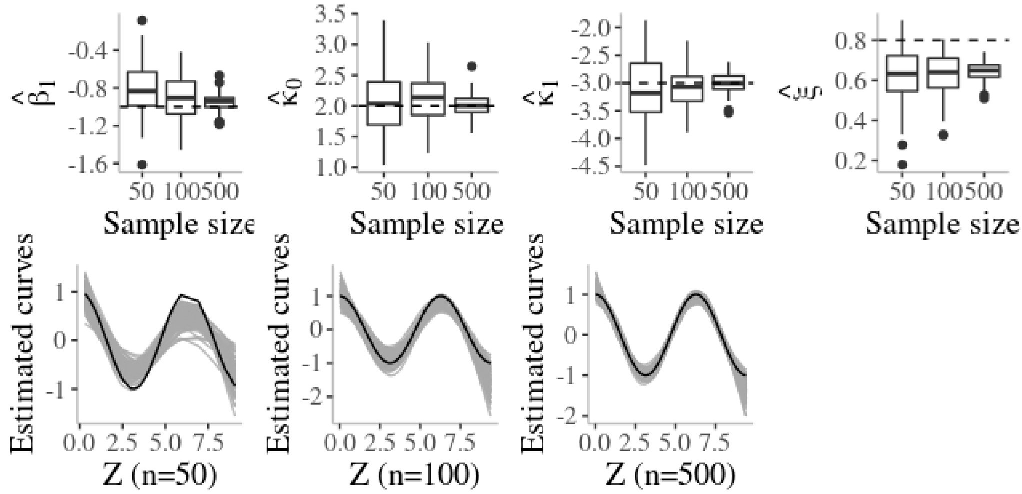


Figure 66 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Beta-SPAM-GEE with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

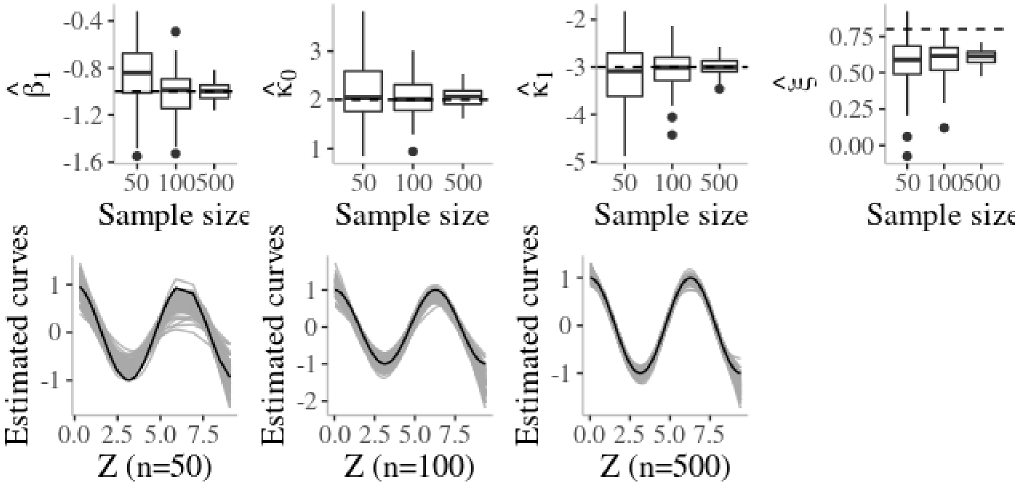


Figure 67 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Beta-SPAM-GEE with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

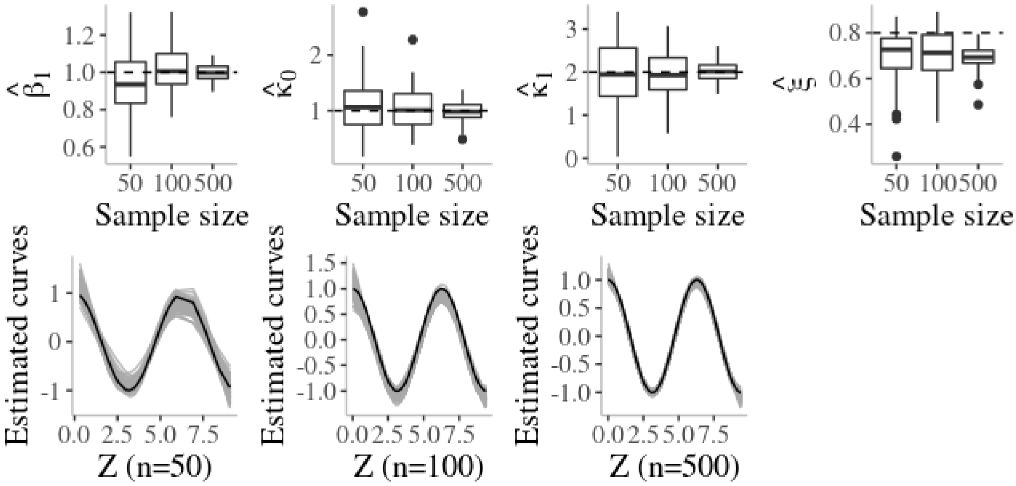


Figure 68 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Beta-SPAM-GEE with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

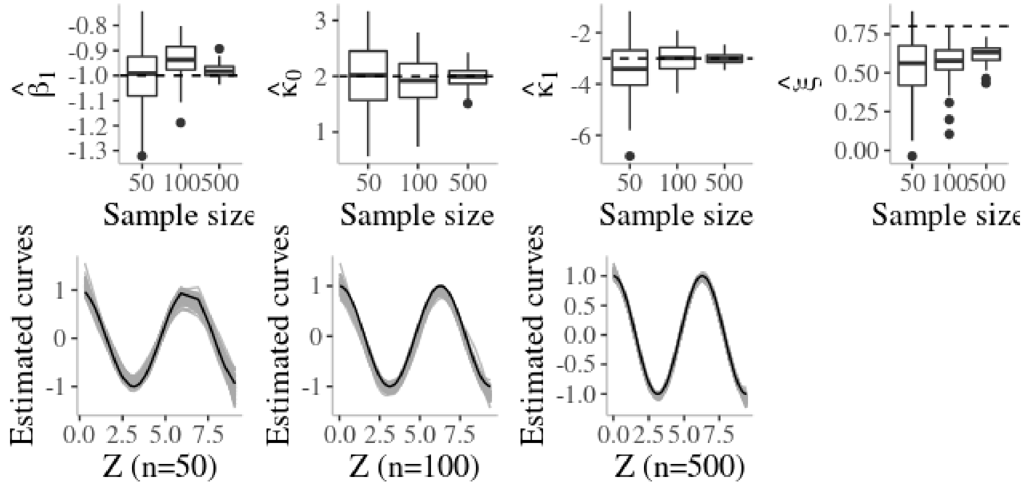


Figure 69 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Simplex-SPAM-GEE with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

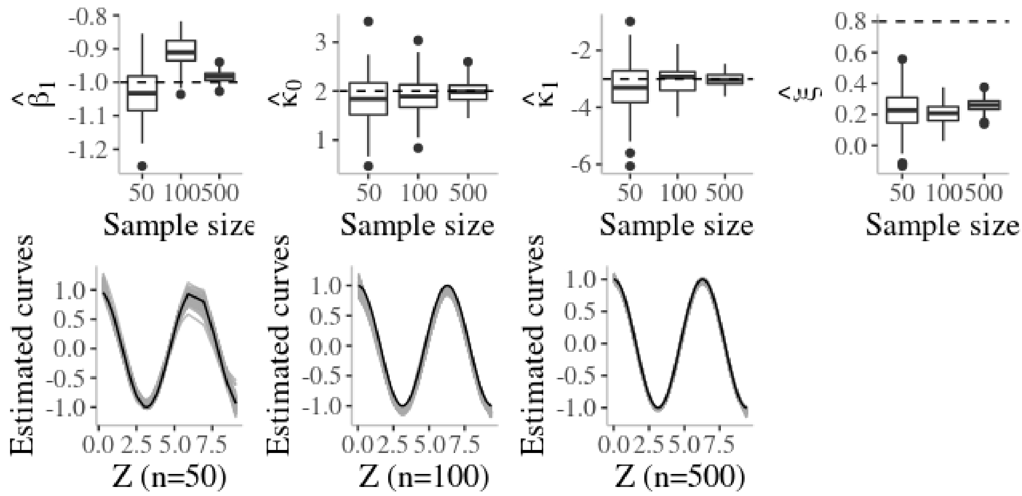


Figure 70 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Simplex-SPAM-GEE with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

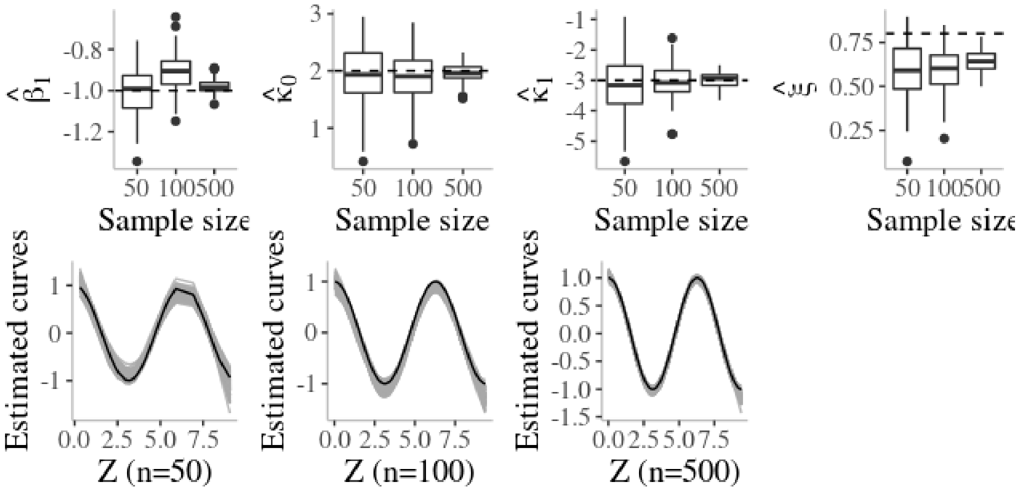


Figure 71 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Simplex-SPAM-GEE with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

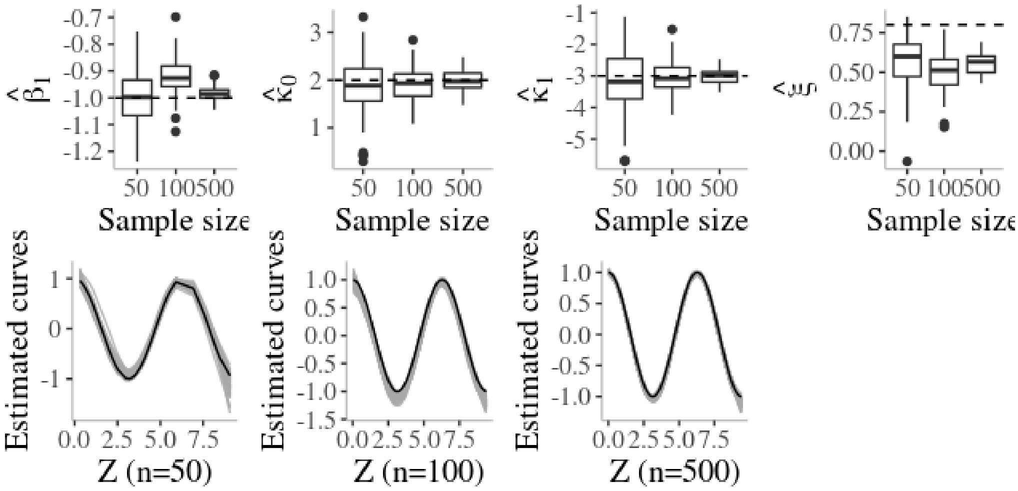


Figure 72 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Simplex-SPAM-GEE with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

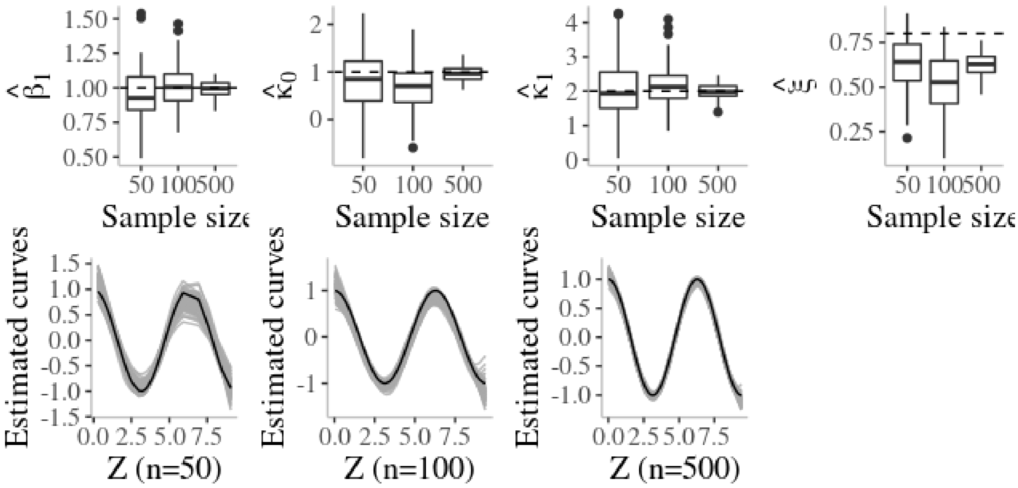


Figure 73 – The first plots are the estimates of parameters β_1, κ_0, κ_1 and ξ for Simplex-SPAM-GEE with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

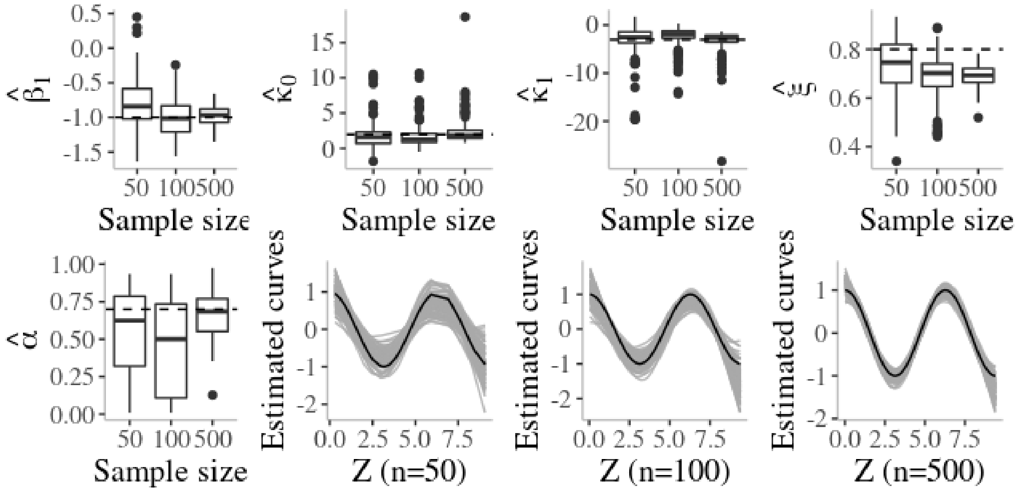


Figure 74 – The first plots are the estimates of parameters β_1, κ_0, κ_1, α and ξ for BR-SPAM-GEE with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

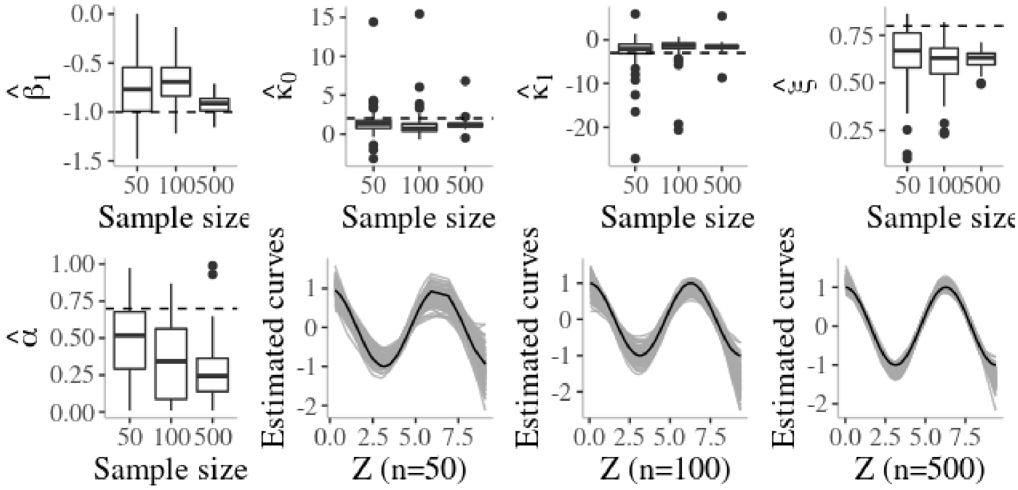


Figure 75 – The first plots are the estimates of parameters β_1, κ_0, κ_1, α and ξ for BR-SPAM-GEE with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

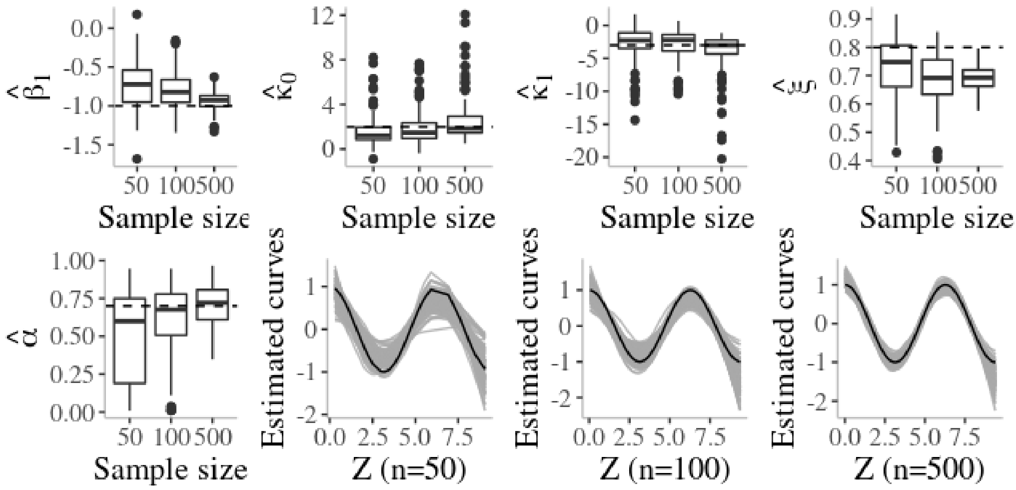


Figure 76 – The first plots are the estimates of parameters β_1, κ_0, κ_1, α and ξ for BR-SPAM-GEE with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

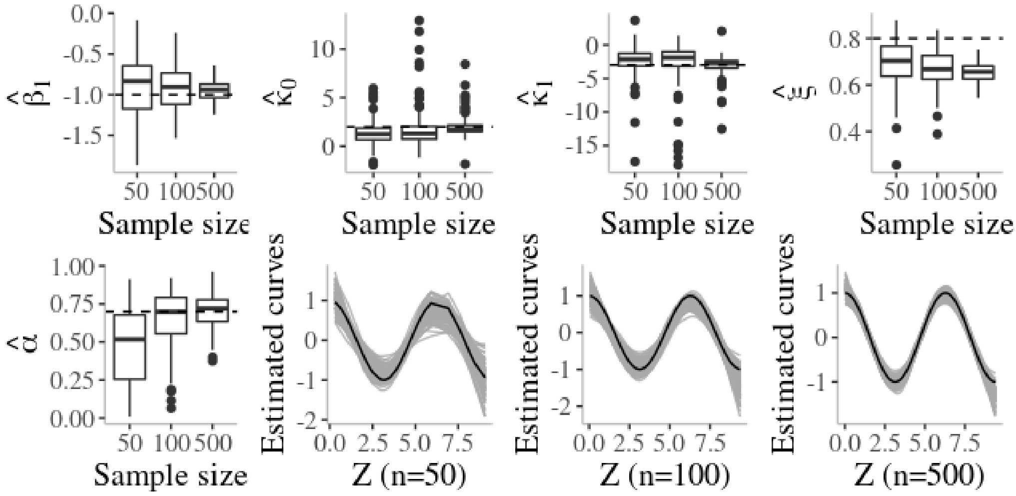


Figure 77 – The first plots are the estimates of parameters β_1, κ_0, κ_1, α and ξ for BR-SPAM-GEE with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

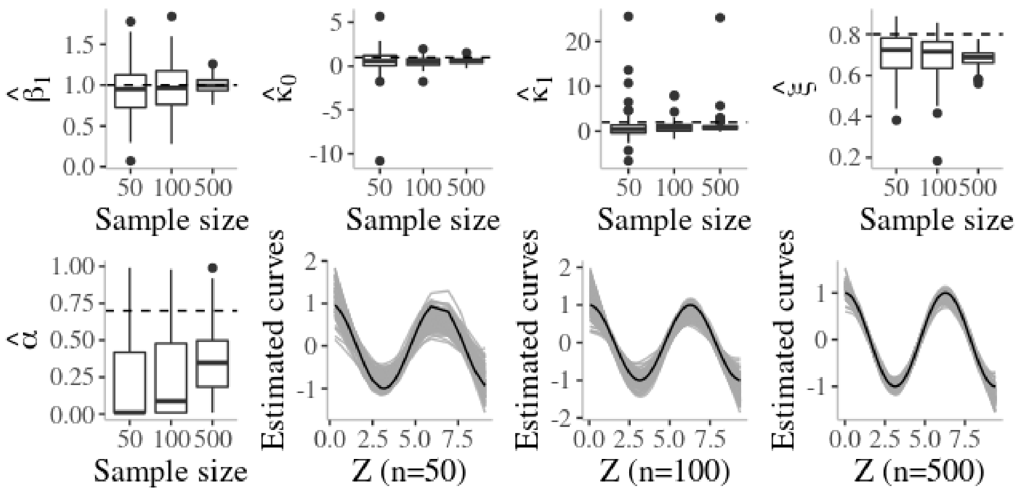


Figure 78 – The first plots are the estimates of parameters β_1, κ_0, κ_1, α and ξ for BR-SPAM-GEE with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

B.4 Simulation results of Chapter 5: Study 1

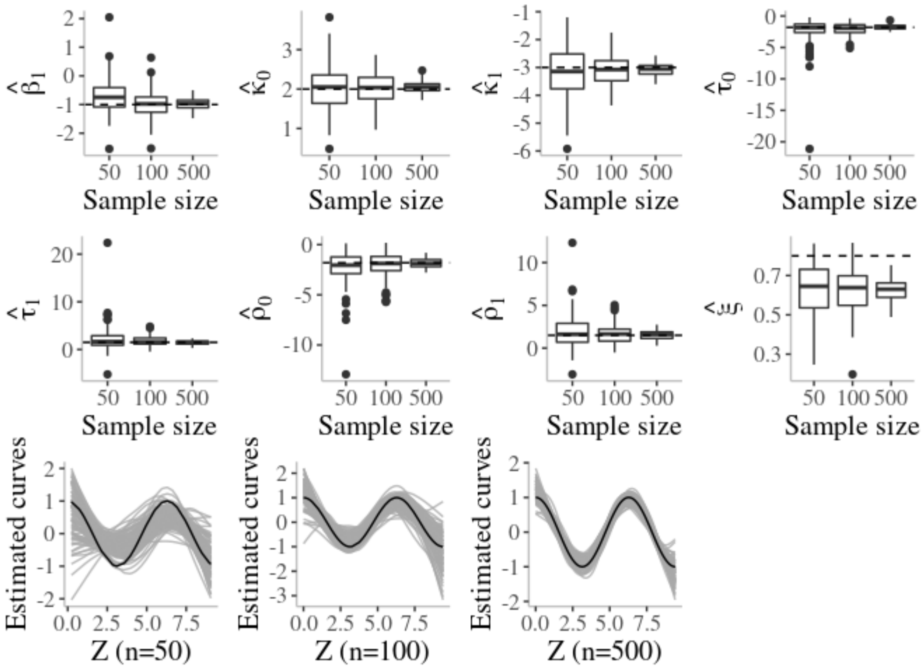


Figure 79 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAB-SPAM-GEE with logit as link function for mean by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

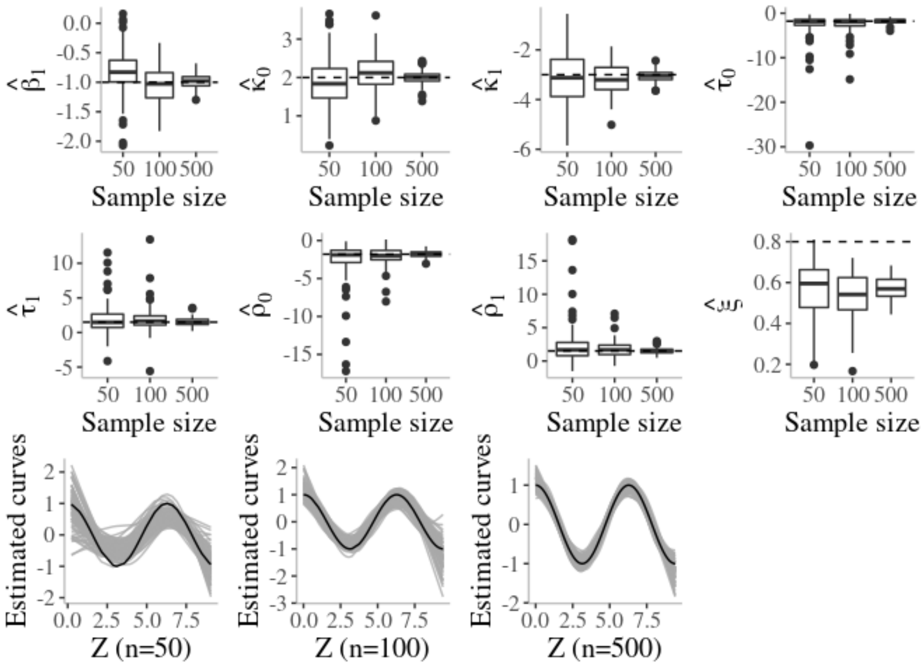


Figure 80 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAB-SPAM-GEE with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

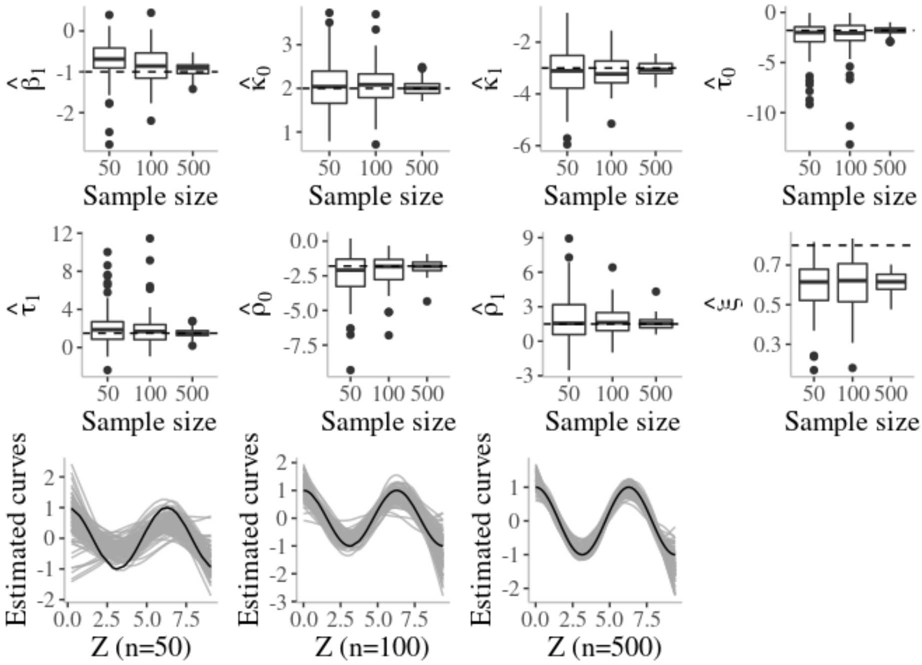


Figure 81 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAB-SPAM-GEE with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

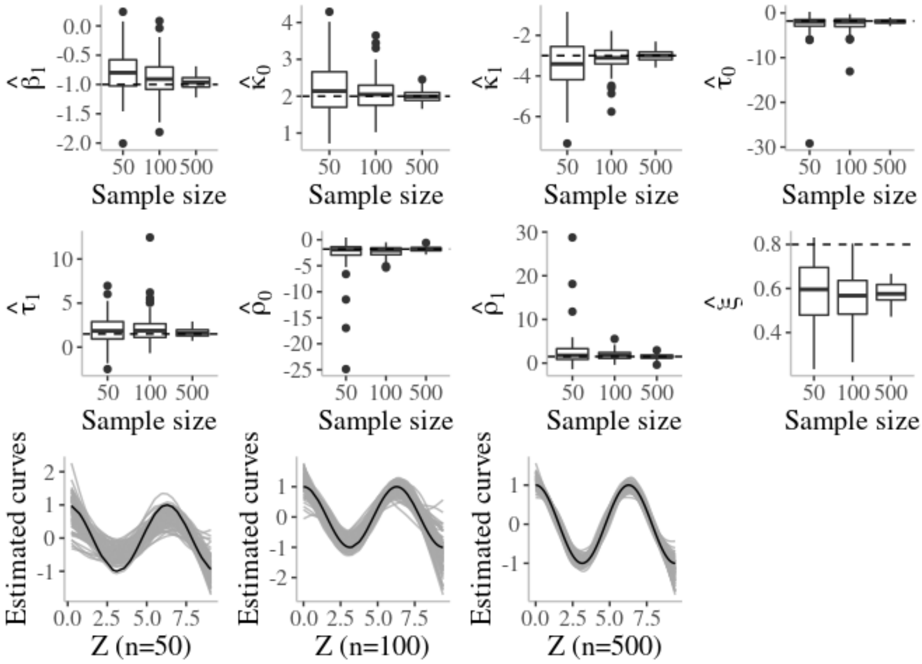


Figure 82 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAB-SPAM-GEE with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

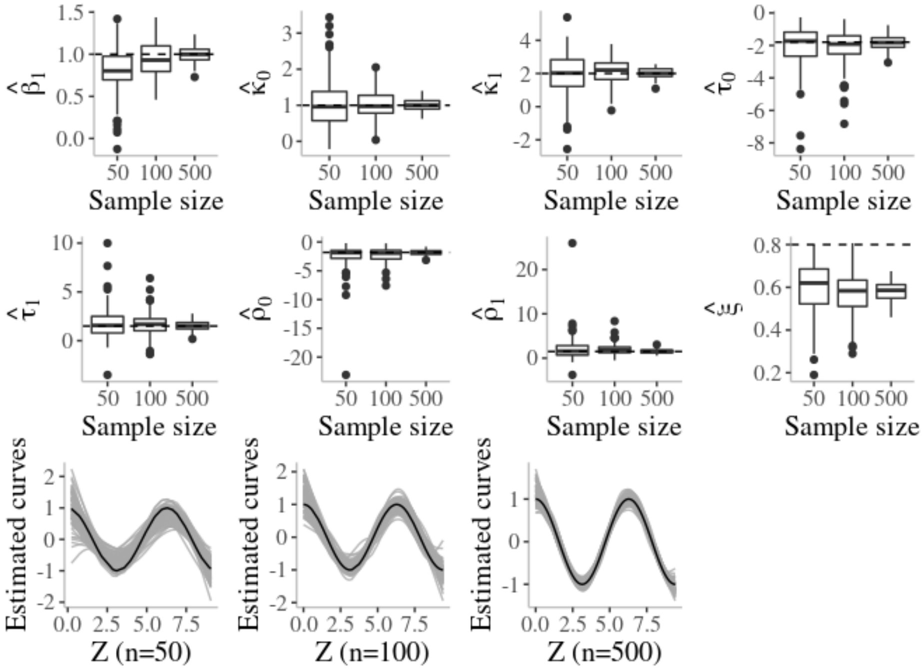


Figure 83 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAB-SPAM-GEE with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

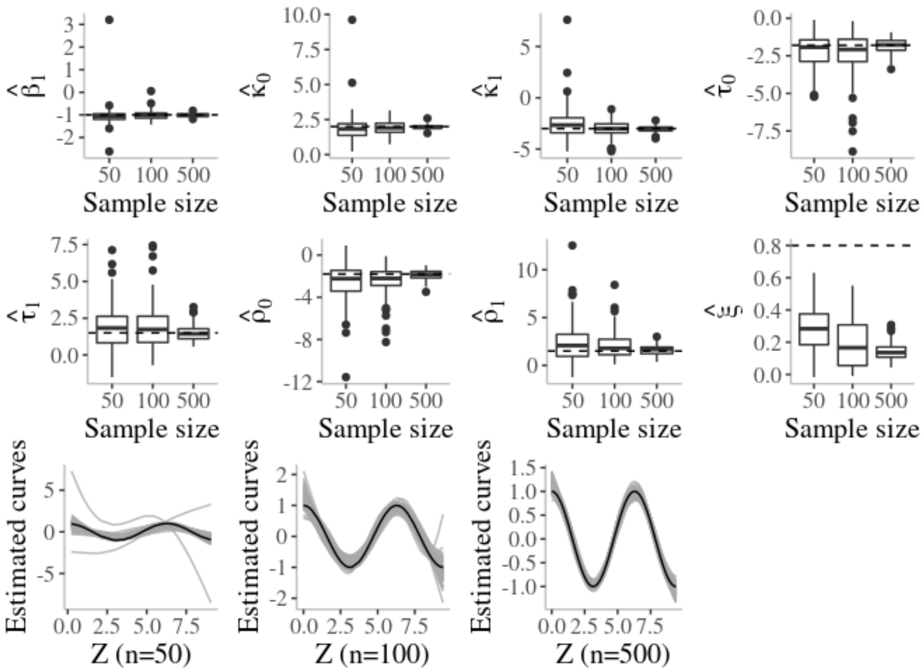


Figure 84 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAS-SPAM-GEE with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

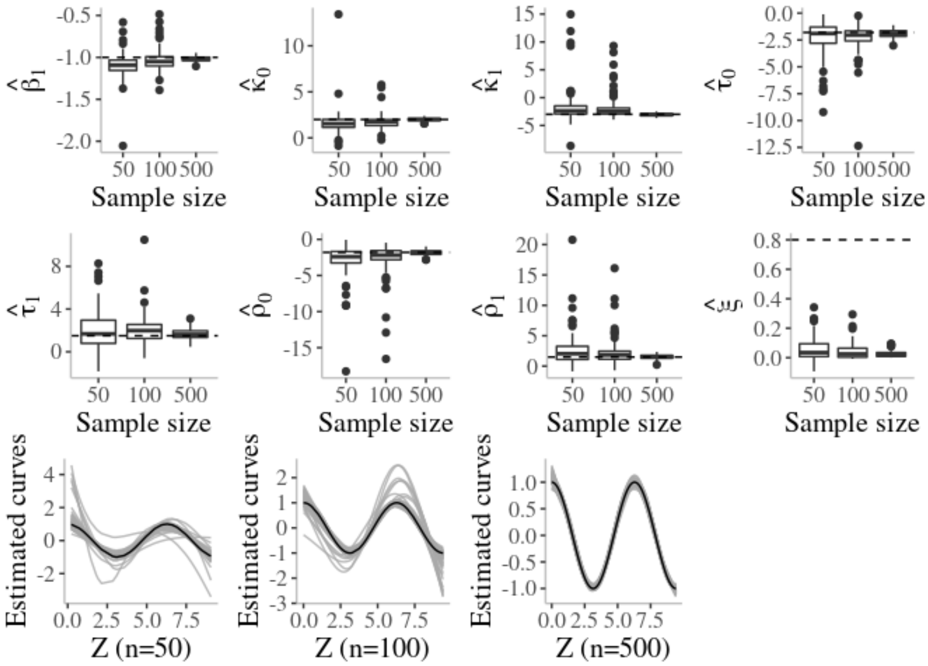


Figure 85 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAS-SPAM-GEE with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

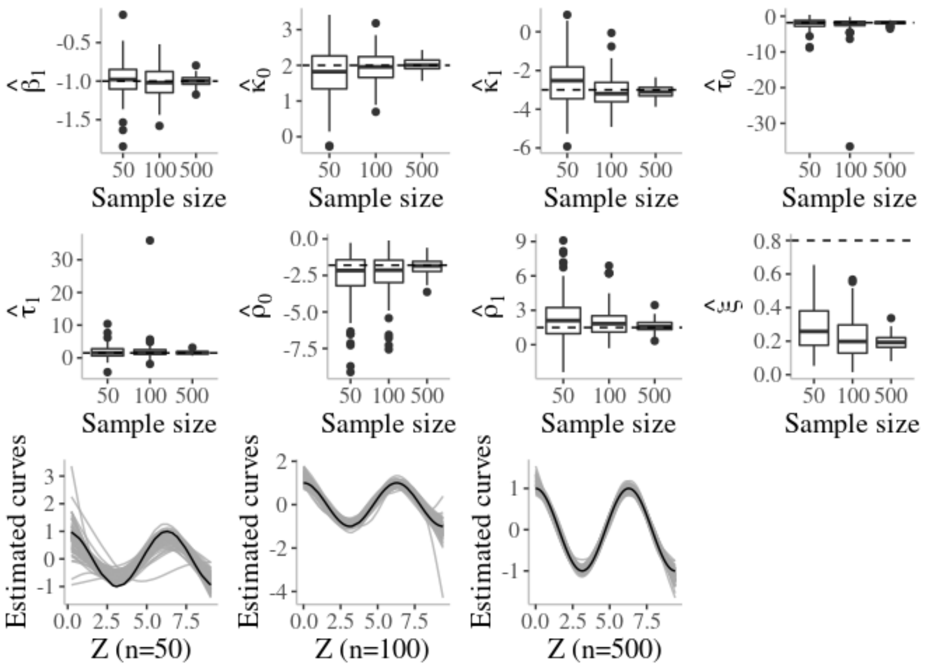


Figure 86 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAS-SPAM-GEE with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

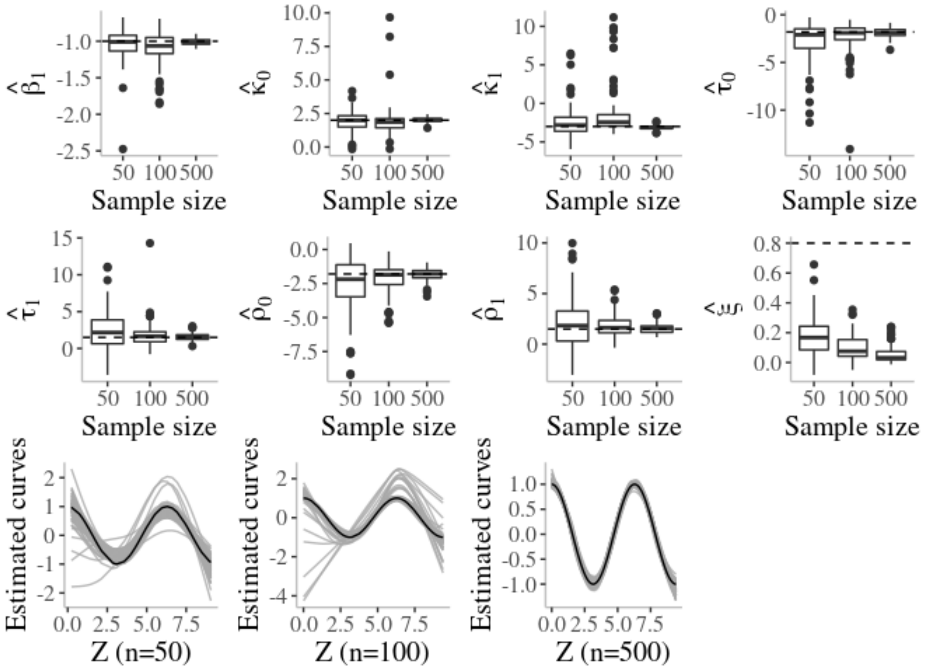


Figure 87 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAS-SPAM-GEE with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

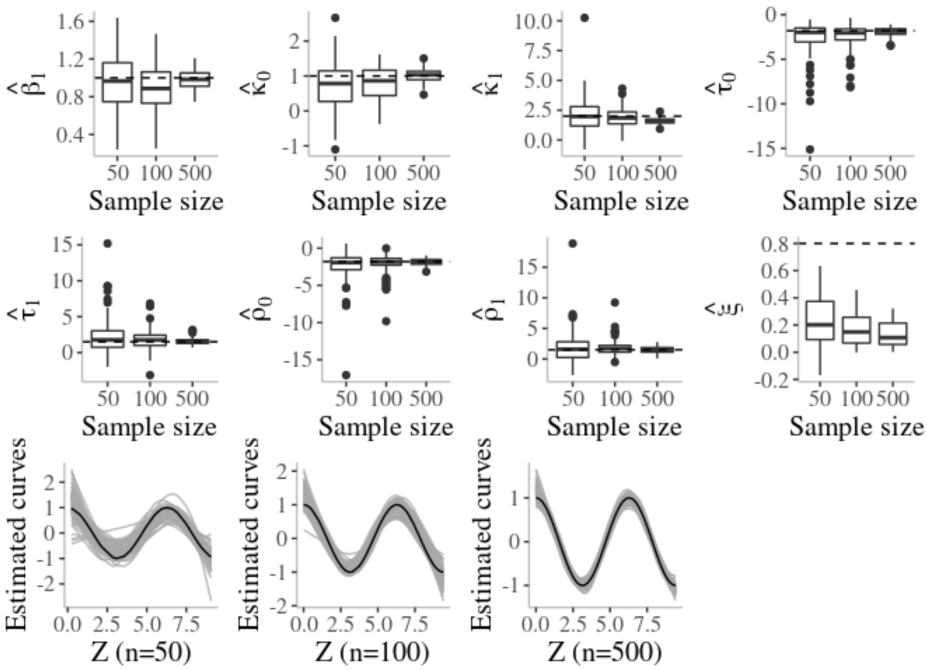


Figure 88 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1 and ξ for ZOAS-SPAM-GEE with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

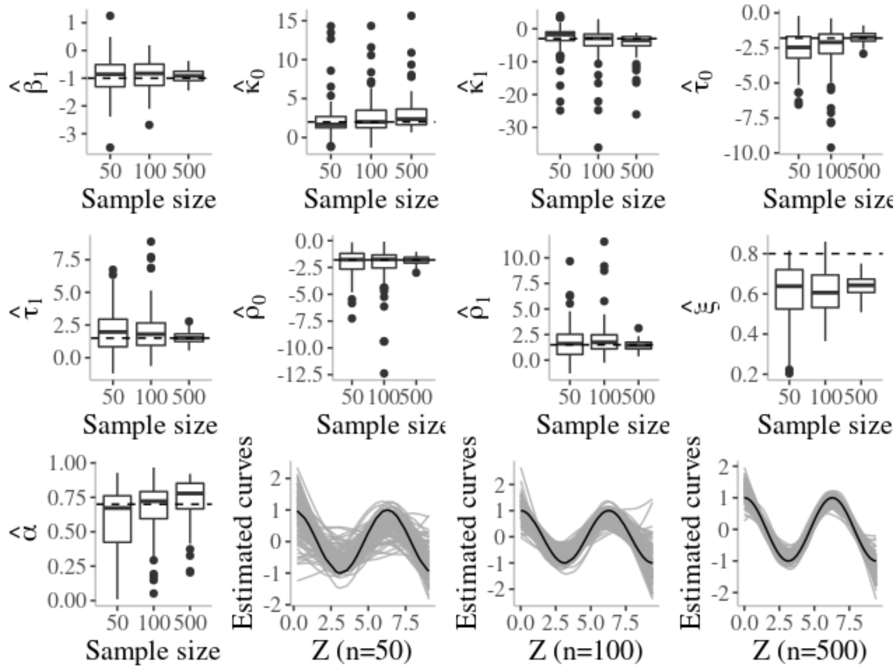


Figure 89 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1, α and ξ for ZOABR-SPAM-GEE with logit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

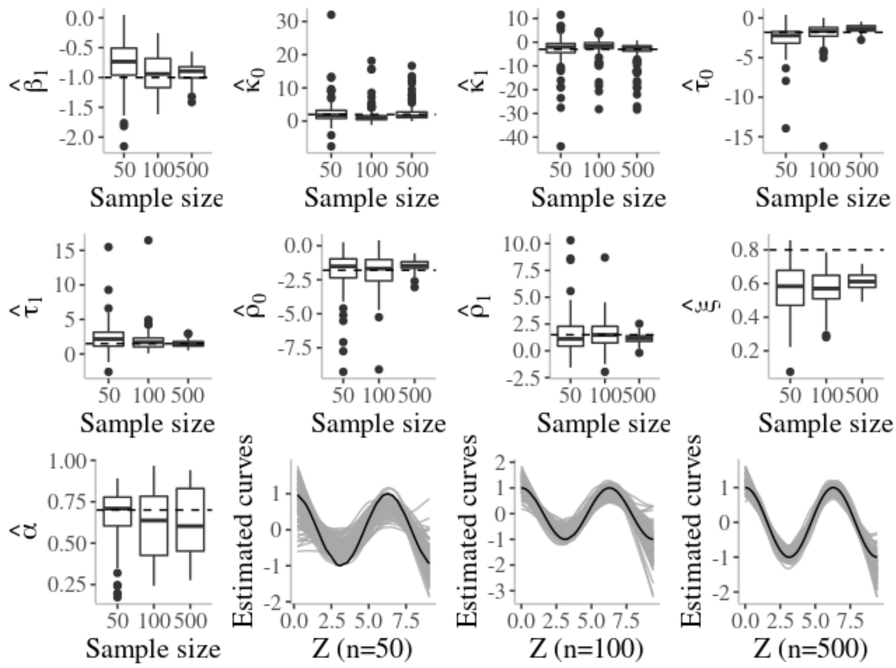


Figure 90 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1, α and ξ for ZOABR-SPAM-GEE with probit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

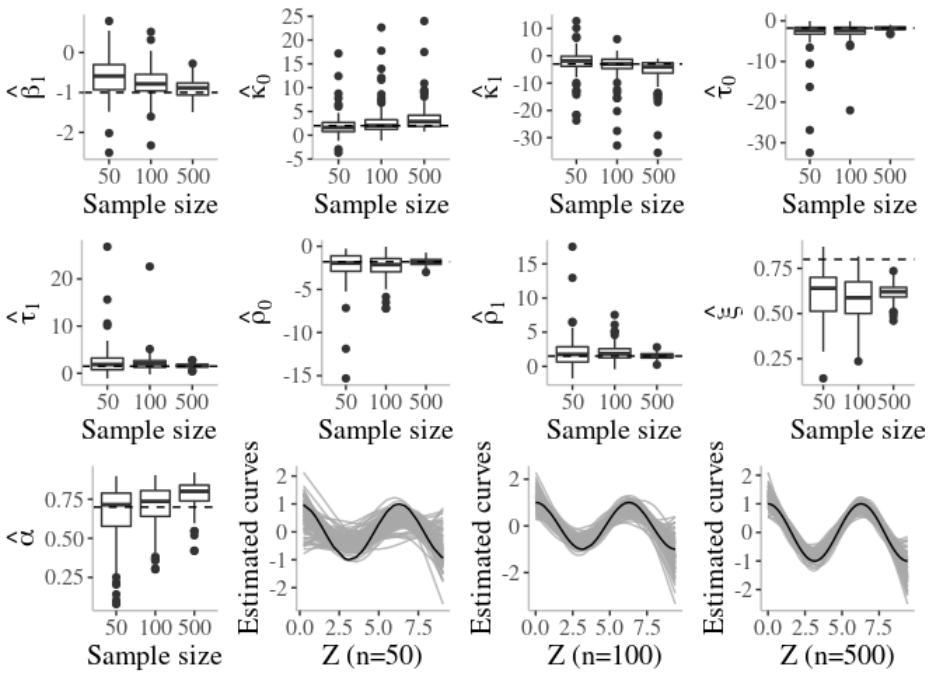


Figure 91 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1, α and ξ for ZOABR-SPAM-GEE with cauchit as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

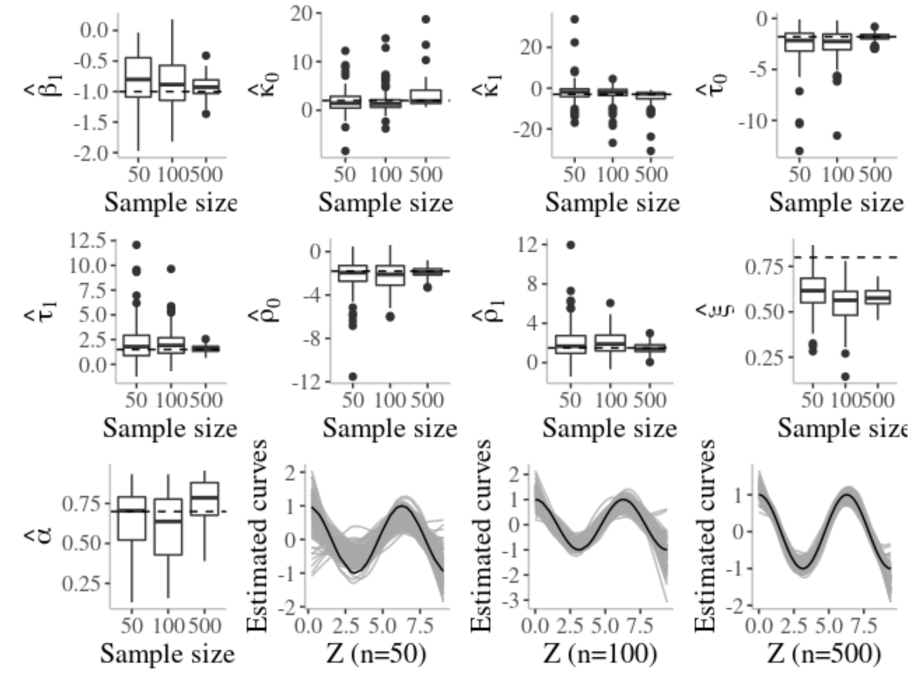


Figure 92 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1, α and ξ for ZOABR-SPAM-GEE with cloglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).

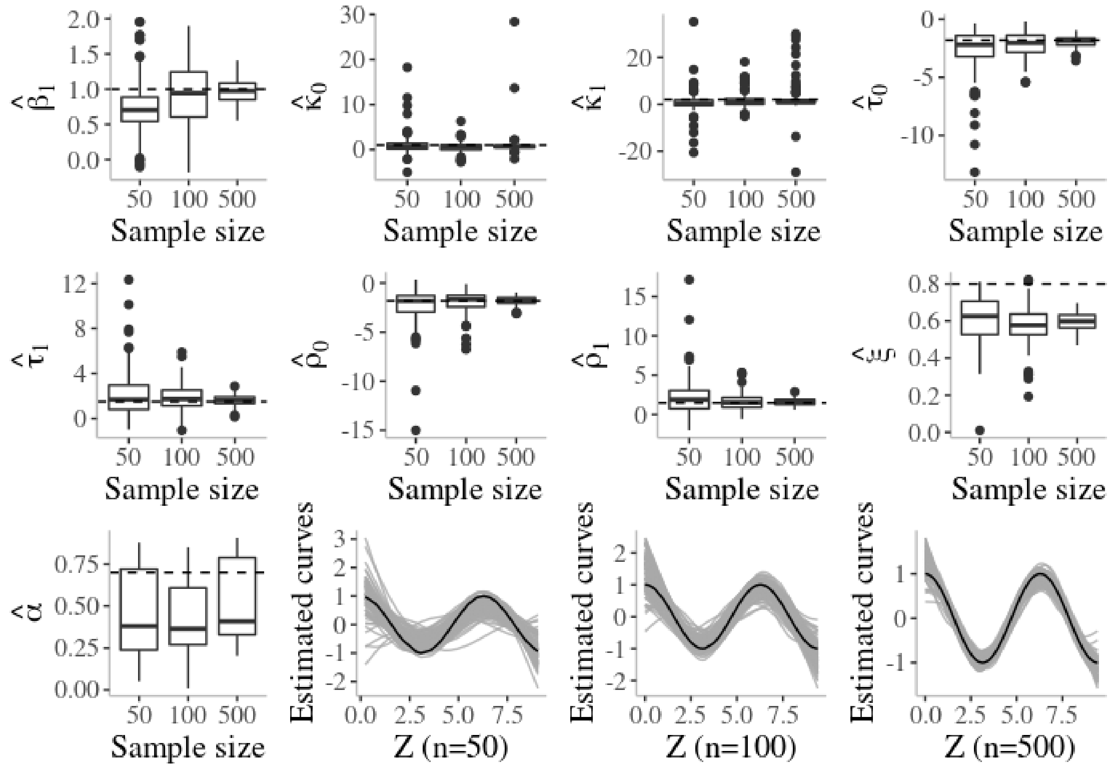


Figure 93 – The first plots are the estimates of parameters β_1, κ_0, κ_1, ρ_0, ρ_1, τ_0, τ_1, α and ξ for ZOABR-SPAM-GEE with loglog as link function by the sample size. The plots of the last line are the estimated curves (grey lines) and real curve (black line) for the same model versus the covariate Z varying the sample size (n).