

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

Uma Classe de Modelos de Regressão Bivariados para Respostas Discreta e Contínua

Willian Luís de Oliveira

São Carlos - SP
2016

Willian Luís de Oliveira

Uma Classe de Modelos de Regressão Bivariados para Respostas Discreta e Contínua

Tese apresentada ao Departamento de Estatística da
Universidade Federal de São Carlos - DEs/UFSCar
como parte dos requisitos para obtenção do título de
doutor em Estatística.

Orientador: Prof. Dr. Carlos Alberto Ribeiro Diniz

São Carlos - SP
2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

Oliveira, Willian Luís de
048c Uma classe de modelos de regressão bivariados
para respostas discreta e contínua / Willian Luís de
Oliveira. -- São Carlos : UFSCar, 2016.
139 p.

Tese (Doutorado) -- Universidade Federal de São
Carlos, 2016.

1. Modelos de regressão bivariados. 2. Respostas
discreta e contínua. 3. Método da fatoração. 4.
Dependência entre respostas. I. Título.



Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Willian Luís de Oliveira, realizada em 28/01/2016:

Prof. Dr. Carlos Alberto Ribeiro Diniz
UFSCar

Prof. Dr. Enrico Antonio Colosimo
UFMG

Profa. Dra. Juliana Cobre
USP

Prof. Dr. Mário de Castro Andrade Filho
USP

Profa. Dra. Roseli Aparecida Leandro
USP

Agradecimentos

Agradeço a Deus por me dar saúde e sabedoria.

Aos meus pais Idemar e Cleusa que sempre me apoiaram e incentivaram.

À Aline Campos pela motivação, ajuda e companheirismo.

À minha tia Maria pelo carinho.

Ao meu irmão Wander, minha sobrinha Tainara e à Marli pelos momentos de distração.

Ao meu orientador Prof. Carlos Alberto Ribeiro Diniz pela confiança, paciência e ensinamentos.

Aos professores Enrico Colosimo, Juliana Cobre, Mário de Castro e Roseli Leandro, membros da banca examinadora, pelas contribuições dadas.

À professora Maria Durbán, pela ajuda e ensinamentos.

Aos meus amigos de pós-graduação pelos momentos de estudo e diversão.

Aos professores do ICMC-USP e DEs-UFSCar pela atenção dada.

Aos funcionários do DEs-UFSCar, em especial à secretária Isabel Araujo por toda ajuda.

À CAPES pelo apoio financeiro.

Resumo

Nesta tese é proposta uma classe ampla e geral de modelos bivariados para respostas mistas em que as distribuições conjuntas são construídas pelo método da fatoração (função densidade de probabilidade, (fdp), como o produto de uma fdp marginal e uma fdp condicional). É assumido que a distribuição da variável resposta discreta e a distribuição condicional da variável resposta contínua dada a variável discreta pertencem à família exponencial de distribuições uniparamétrica ou biparamétrica. Além disso, as médias marginais são relacionadas a covariáveis através de funções de ligação usando preditores linear e/ou não linear e/ou não paramétrico e uma estrutura de dependência entre as respostas é inserida no modelo via a média condicional. Métodos de estimação, análises de diagnóstico e técnicas de influência são apresentadas assim como um estudo de simulação considerando os modelos Bernoulli-exponencial e Poisson-normal semiparamétrico, dois casos particulares da classe proposta. Finalmente, um dos modelos propostos é usado em um conjunto de dados reais envolvendo gastos totais com cuidados para cada paciente durante a hospitalização, o uso ou não da unidade de tratamento intensivo e a idade do paciente.

Palavras-chave: Modelos de regressão bivariados. Respostas discreta e contínua. Método da fatoração. Dependência entre respostas.

Abstract

In this thesis, a wide general class of models for mixed responses is proposed in which joint distributions are constructed by the conditional approach (probability density functions, (pdf), as the product of a marginal pdf and a conditional pdf). It is assumed that the distribution of the discrete response and the conditional distribution of the continuous response given the discrete variable belong to one- or two-parameter exponential family of distributions. Furthermore, the marginal means are related to the covariates by link functions using linear and/or nonlinear and/or non-parametric predictors and a dependency structure between the responses is introduced into the model via the conditional mean. Estimation methods, diagnostic analysis and influence techniques are presented as well as a simulation study considering the Bernoulli-exponential and Poisson-normal semiparametric models, two particular cases of the proposed class. Finally, one of the proposed models is used in a real data set involving the total cost of care for each patient during hospitalization, the use or not of the intensive treatment units and the age of the patient.

Keywords: Bivariate regression models. Discrete and continuous responses. Conditional approach. Dependency between the responses.

Sumário

Resumo	i
Abstract	ii
1 Introdução	1
1.1 Revisão bibliográfica	3
1.2 Apresentação dos capítulos	9
2 Modelos bivariados mistos	12
2.1 Modelo paramétrico	17
2.1.1 Máxima verossimilhança	17
2.1.2 Intervalos de confiança assintóticos	21
2.1.3 Teste de hipóteses	22
2.2 Modelo semiparamétrico	24
2.2.1 Escore local	24
3 Modelos bivariados semiparamétricos	28

3.1	P-splines	28
3.1.1	Base B-spline	29
3.1.2	Matriz de penalização	31
3.1.3	Aplicação da técnica P-splines	32
3.2	Modelos aditivos generalizados com P-splines	36
3.2.1	Transformação da base	38
3.3	Modelos bivariados com P-splines	45
4	Casos particulares	53
4.1	Modelo Bernoulli-normal	53
4.1.1	Estimação	55
4.2	Modelo Poisson-normal	56
4.2.1	Estimação	57
4.3	Modelo Bernoulli-exponencial	57
4.3.1	Estimação	58
4.4	Modelo Poisson-normal semiparamétrico	60
4.4.1	Estimação	60
5	Técnicas de diagnóstico e influência	62
5.1	Resíduos	63
5.1.1	Resíduos ordinários padronizados	64
5.1.2	Resíduos quantis normalizados aleatorizados	65

5.1.3	Resíduos componentes da desviância	66
5.2	Influência local	67
5.3	Influência global ou total	71
6	Estudo de simulação	73
6.1	Modelo Bernoulli-exponencial	75
6.2	Modelo Poisson-normal semiparamétrico	89
7	Aplicação	111
8	Conclusão	120
9	Propostas de trabalhos futuros	123
A	Derivadas da função log-verossimilhança	125
	Bibliografia	132

Capítulo 1

Introdução

Em muitas áreas da ciência, tais como economia, medicina e psicologia, são comuns situações em que duas variáveis respostas associadas com o mesmo indivíduo são observadas simultaneamente. Por exemplo, em um hospital pode-se observar os gastos médios dos pacientes no período de internação e se ele utilizou ou não o centro cirúrgico neste período. Na universidade, pode-se observar as notas dos alunos em uma determinada disciplina e a quantidade de faltas do aluno. Nestes casos, uma possível análise poderia ser realizada considerando cada uma das variáveis resposta de forma independente, isto é, adotando um modelo marginal para cada uma das respostas consideradas. Entretanto, se a relação intrínseca entre as duas respostas não é considerada na análise, resultados irrealistas podem ser gerados. Neste contexto, modelos bivariados são vistos como uma boa alternativa, em que uma estrutura de dependência entre as respostas é adotada.

A inserção de uma estrutura de dependência na modelagem muitas vezes a torna melhor se comparada às análises marginais, levando a resultados mais confiáveis. A estrutura de dependência entre as variáveis resposta tem sido objeto de estudo de alguns trabalhos envolvendo modelos bivariados (Olkin & Tate, 1961).

Nos modelos bivariados em geral, as variáveis resposta podem ser ambas discretas, ambas contínuas ou uma discreta e a outra contínua. Modelos bivariados com ambas as respostas contínuas têm sido amplamente explorados na literatura (Scollnik, 2002; Song *et al.*, 2004), com destaque para respostas seguindo distribuição normal (ou aproximadamente normal), em que uma classe geral adequada de modelos lineares está disponível para a análise dos dados (Fitzmaurice & Laird, 1995; Lin *et al.*, 2010). Modelos bivariados para dados binários (Cox, 1972; McDonald, 1993; Shults *et al.*, 2009) e dados de contagem (Jung & Winkelmann, 1993; van Ophem, 1999; Berkhout & Plug, 2004; Khafri *et al.*, 2008) também têm ganhado destaque. O mesmo não ocorre quando ambas variáveis resposta são categóricas, por exemplo.

Para problemas envolvendo variáveis resposta discreta e contínua, a análise conjunta dos dados pode ser complexa devido à quantidade relativamente baixa de modelos bivariados com respostas mistas disponíveis, além da ausência de um modelo padrão e mais geral, englobando os casos já existentes. Esta dificuldade é causada principalmente pela falta de uma distribuição bivariada mista natural (Gueorguieva & Agresti, 2001).

As distribuições bivariadas do vetor de variáveis resposta podem ser construídas por meio de alguns métodos, tais como método da fatoração, método da transformação marginal, método de construção via cópulas, método de mistura e combinação (Balakrishnan & Lai, 2009). Uma técnica que tem sido utilizada em alguns trabalhos (Gueorguieva, 2001; Gueorguieva & Agresti, 2001; de Leon & Chough, 2013) é a da combinação, em que cada resposta é adotada como sendo um modelo linear generalizado misto e então, impondo uma distribuição multivariada normal para os efeitos aleatórios, combinam as respostas em um modelo único conjunto. Neste caso, a estrutura de dependência entre as respostas é inserida na distribuição conjunta dos efeitos aleatórios. Lin *et al.* (2010) também consideram efeitos aleatórios na construção do modelo bivariado para resposta binária e contínua, porém, consideram uma distribuição não normal para o efeito aleatório associado à variável binária. Modelos bivariados

para respostas mistas utilizando cópulas também têm sido explorados na literatura, embora a quantidade de trabalhos disponíveis seja pequena. Para mais detalhes, vide de Leon & Wu (2011) e Beilei (2013).

Uma maneira mais direta para construir modelos bivariados mistos, bastante aplicada na literatura, é a técnica da fatoração em que a distribuição conjunta das respostas discreta e contínua é escrita como produto entre uma distribuição marginal, de uma das respostas e uma distribuição condicional, da outra resposta dada a primeira. Desta forma, dois distintos modelos podem ser considerados: o modelo que considera a distribuição marginal da resposta discreta e condiciona a variável contínua à discreta e o modelo que considera a densidade marginal da variável contínua e condiciona a variável resposta discreta à contínua. No primeiro caso, a variável resposta discreta é tratada como resposta primária enquanto que a variável resposta contínua é considerada a variável intermediária. Pensamento análogo é realizado ao segundo modelo. A seguir, apresentamos uma revisão bibliográfica dos principais trabalhos que utilizam a técnica da fatoração, com destaque para os modelos com respostas mistas.

1.1 Revisão bibliográfica

Entre os modelos bivariados considerando duas variáveis resposta discretas, destaca-se na literatura o modelo de Poisson bivariado, muito utilizado em problemas envolvendo dados de contagem. Este modelo é adotado por Jung & Winkelmann (1993) tendo como motivação dados de mobilidade de emprego. van Ophem (1999), motivado por uma análise envolvendo o comportamento turístico e recreativo de pessoas em viagens, considera uma modelagem com ambas as variáveis resposta seguindo distribuição de Poisson, sendo as variáveis respostas correlacionadas. Khafri *et al.* (2008), assim como Jung & Winkelmann (1993), também adotam o modelo de Poisson bivariado sendo que as observações não são correlacionadas entre os indivíduos e sim, entre as

respostas. Uma análise bayesiana para o modelo proposto é abordada tendo como motivação dados de tratamentos clínicos de casais inférteis.

Trabalhos considerando ambas as variáveis resposta contínuas também são encontrados na literatura. Scollnik (2002) analisa dois modelos de regressão motivados por dados bivariados de sinistros em que a análise é realizada em relação a seguros contra acidentes. São observadas as variáveis resposta valor de perda e valor de despesas alocadas com ajuste de sinistro. Em ambos os modelos, as variáveis resposta são consideradas diretamente relacionadas uma a outra. Song *et al.* (2004) têm como principal objetivo estimar a correlação entre as variáveis resposta contagem de células CD_4+ e carga viral de HIV, de uma pesquisa sobre HIV. Ambas as variáveis são assumidas normalmente distribuídas sendo que carga viral de HIV é condicionada à contagem de células de CD_4+ . Os dados são obtidos de mulheres grávidas e a correlação, estimada por meio de equações de estimação generalizadas, faz-se necessária uma vez que o teste para carga viral de HIV fornece informação que é utilizada em conjunto com contagem de células CD_4+ .

Nos casos de variáveis resposta mistas, em que uma é discreta e a outra contínua, alguns trabalhos são propostos seguindo a técnica da fatoração. Um dos modelos mais simples contendo variáveis resposta mistas é aquele em que uma das respostas é discreta e segue distribuição de Bernoulli enquanto que a outra é contínua. Um dos primeiros trabalhos nesta linha é apresentado por Tate (1954) que, motivado por um problema no campo da psicologia, considera um modelo para dados bivariados em que a distribuição de uma das variáveis resposta é discreta seguindo distribuição de Bernoulli enquanto que a outra variável é contínua e, condicionada ao valor da variável discreta, segue distribuição normal. O objetivo principal do trabalho é encontrar uma medida de associação entre as respostas discreta e contínua. Olkin & Tate (1961), também motivados por estudos em psicologia, consideram o modelo adotado em Tate (1954) e o estendem para uma versão multivariada em que um vetor de variáveis resposta é

discreto e segue distribuição multinomial enquanto que outro vetor de variáveis resposta é contínuo e sua distribuição, condicionada ao vetor de variáveis resposta discreto, é considerada normal multivariada. Como em Tate (1954), o interesse é encontrar uma medida de associação entre os vetores de respostas discretos e contínuos. Para ilustrar o modelo proposto, é utilizado um conjunto de dados contendo características de membros de partidos políticos britânicos. Lauritzen & Wermuth (1989) desenvolvem uma classe de modelos estatísticos denominada modelos gráficos para variáveis resposta mistas discreta e contínua, baseada na distribuição normal condicional. Esta classe de modelos tem como principal finalidade descrever e investigar a associação entre as variáveis resposta. Uma extensão do modelo apresentado por Olkin & Tate (1961) pode ser vista em de Leon & Carrière (2007).

Diferente dos modelos de Tate (1954) e Olkin & Tate (1961) em que o condicionamento é realizado na resposta contínua, Cox (1972) considera um modelo para dados bivariados misto em que a distribuição da variável resposta contínua é normal e a distribuição da variável discreta, condicionada à variável contínua, é Bernoulli, em que a função de ligação logística é adotada. Analogamente, Catalano & Ryan (1992) consideram um modelo bivariado cuja variável resposta discreta também é condicionada à variável resposta contínua e assumida seguindo distribuição Bernoulli. Além disso, uma variável latente contínua não observada associada à variável resposta discreta é inserida no modelo a fim de obter a distribuição conjunta do modelo bivariado misto. As variáveis resposta contínua e latente não observada seguem distribuição conjuntamente normal bivariada, sendo o modelo parametrizado de tal forma que a distribuição conjunta das respostas seja expressa como um produto entre a distribuição marginal da resposta contínua e a distribuição condicional da resposta discreta, dada a resposta contínua.

No modelo de Catalano & Ryan (1992), os parâmetros de ambas as distribuições marginal e condicional são relacionadas a covariáveis por meio das funções de ligação

linear e probito, respectivamente, sendo utilizados como motivação dados de um estudo de toxicidade em ratos. Entretanto, dois problemas são encontrados nos modelos de Cox (1972) e Catalano & Ryan (1992): os parâmetros de regressão nos modelos logístico (Cox, 1972) e probito (Catalano & Ryan, 1992) não têm uma interpretação marginal e além disso, suas estimativas de máxima verossimilhança não são consistentes quando a associação entre as variáveis resposta discreta e contínua no modelo bivariado é mal especificada, o que não é o ideal em algumas aplicações. Cox & Wermuth (1992) comparam alguns diferentes modelos para dados bivariados discreto e contínuo, em que a distribuição conjunta é expressa como um produto entre uma distribuição marginal e uma condicional. Além disso, os autores notam que a interpretação marginal dos parâmetros de regressão é possível somente se a média condicional da variável resposta é relacionada a covariáveis por meio de uma função de ligação linear.

Procurando solucionar os problemas encontrados nos modelos de Cox (1972) e Catalano & Ryan (1992), Fitzmaurice & Laird (1995) apresentam um modelo para respostas bivariadas contínua e discreta em que os parâmetros de regressão de ambas as variáveis têm interpretação marginal e os respectivos estimadores de máxima verossimilhança são consistentes, mesmo se a associação entre as variáveis resposta discreta e contínua for mal especificada no modelo. Isto é possível simplesmente invertendo a variável que está condicionada, isto é, assumindo que a variável discreta tem distribuição de Bernoulli e que a variável contínua, condicionada à variável discreta, tem distribuição normal. O modelo proposto em Fitzmaurice & Laird (1995) pode ser visto como uma extensão do modelo apresentado em Olkin & Tate (1961). Uma extensão do modelo inicial para permitir agrupamento entre as observações também é apresentada utilizando como ferramenta o método de equações de estimação generalizadas.

Assim como Fitzmaurice & Laird (1995), Yang *et al.* (2007) e George *et al.* (2007) apresentam um modelo bivariado para respostas mistas construído pela técnica da fatoração sendo que o condicionamento é feito na variável resposta contínua. Em ambos

os casos é assumido que a variável resposta contínua, dada a discreta, segue distribuição normal. No modelo de Yang *et al.* (2007), a variável discreta é assumida seguindo distribuição de Poisson sendo que o modelo inicial é também estendido para dados longitudinais. Neste contexto, possíveis mudanças na variância e na correlação dentro de uma unidade sobre o tempo é considerada e soluções são apresentadas. No modelo de George *et al.* (2007), as suposições de que os dados são permutáveis e que a variável discreta é binária, são adotadas. Como em Fitzmaurice & Laird (1995), os autores apresentam o modelo inicialmente para dados sem agrupamento e posteriormente o estendem para permitir o agrupamento.

Fitzmaurice & Laird (1997) e Yang & Kang (2010) consideram os modelos bivariados apresentados em Fitzmaurice & Laird (1995) e Yang *et al.* (2007), respectivamente, tratando também o problema no contexto em que há valores faltantes. Novamente, a técnica da fatoração é utilizada. Recentemente, Samani & Ganjali (2014) propõem um modelo de regressão bivariado com respostas correlacionadas ordinária e binomial negativa utilizando a técnica da fatoração na construção da distribuição conjunta. Efeitos aleatórios associados às respostas são inseridos na construção do modelo, no contexto da classe de modelos apresentada (Gueorguieva, 2001), para acomodar a correlação entre as respostas longitudinais de um mesmo indivíduo. O modelo também considera o problema de dados faltantes.

Embora alguns trabalhos no contexto de modelos bivariados com respostas mistas, construídos a partir da técnica da fatoração tenham sido encontrados na literatura, observa-se a escassez de modelos mais flexíveis e gerais, principalmente modelos cuja densidade condicional da variável contínua dada a discreta seja não normal. Neste contexto, considerando os modelos apresentados por Fitzmaurice & Laird (1995) e Yang *et al.* (2007), é proposta nesta tese uma classe ampla e geral de modelos de regressão bivariados com respostas mistas, cuja distribuição conjunta é construída a partir do produto de uma distribuição marginal e de uma distribuição condicional.

Nesta nova classe é assumido que uma das variáveis resposta é discreta e pertencente à família exponencial uniparamétrica ou biparamétrica enquanto que a outra resposta é considerada contínua e sua distribuição, condicionada à variável aleatória discreta, pertencente à família exponencial uniparamétrica ou biparamétrica.

Covariáveis estão disponíveis e são relacionadas às médias das variáveis resposta discreta e contínua por meio de funções de ligação de acordo com a natureza dos dados. Entretanto, diferentemente do que ocorre nos modelos lineares generalizados, o componente sistemático relacionado à variável resposta discreta e o componente sistemático relacionado à variável resposta contínua são formados por termos paramétricos, compostos de funções lineares e não lineares e por termos não paramétricos no contexto dos modelos aditivos generalizados semiparamétrico (GAM) (Hastie & Tibshirani, 1990). Neste sentido, a classe proposta é mais flexível, uma vez que engloba uma variedade de modelos. A dependência existente entre as variáveis resposta discreta e contínua, relevante à modelagem, é inserida na classe de modelos por meio da distribuição condicional. Definindo uma estrutura de dependência, considera-se que uma transformação da média da variável resposta contínua, condicionada à resposta discreta, é uma função linear ou não linear conhecida da variável resposta discreta, da estrutura de dependência e das covariáveis através das médias marginais, fazendo com que a dependência entre as respostas seja expressada.

Inicialmente, é considerado o caso em que os preditores são formados apenas por termos lineares e não lineares paramétricos. Para esta classe de modelos, o método de máxima verossimilhança é utilizado para calcular os estimadores de máxima verossimilhança correspondentes aos parâmetros do modelo. O vetor escore e a matriz de informação observada utilizadas no processo de estimação são apresentados assim como intervalos de confiança assintótico e estatísticas de teste para testar alguma hipótese de interesse.

Na sequência, é apresentada uma técnica de estimação não paramétrica denominada P-spline, muito utilizada na estimação não paramétrica univariada. Essa técnica é então adaptada aos modelos bivariados da classe, cujos preditores contêm apenas funções suaves dos dados, em um contexto não paramétrico. Todo o processo de estimação é descrito em detalhes. Como consequência da técnica aplicada, pode-se ajustar modelos bivariados envolvendo preditores com termos lineares, não lineares e não paramétricos.

Uma vez apresentados métodos de estimação para os modelos da classe, quatro modelos pertencentes à classe bivariada proposta são exemplificados, sendo que dois deles não são encontrados na literatura. Em relação à parte de diagnóstico, propõem-se três resíduos para a classe proposta: padronizado, quantil e componentes da desviância. Além disso, consideram-se as medidas influência local e total, úteis para detecção de possíveis pontos influentes nos modelos paramétricos. Na sequência, um estudo com dados simulados é realizado considerando os modelos Bernoulli-exponencial e Poisson-normal semiparamétrico, ambos pertencentes à classe. Por fim, uma aplicação a um conjunto de dados reais contendo informações sobre a internação de pacientes em um hospital é proposta, considerado o modelo Bernoulli-exponencial. A organização da tese é descrita abaixo.

1.2 Apresentação dos capítulos

No Capítulo 1 foi apresentada uma breve introdução sobre modelos para dados com duas variáveis resposta, destacando a importância em se utilizar modelos que consideram a dependência entre as respostas. Casos em que ambas as variáveis resposta são discretas, ambas contínuas ou mesmo uma delas é discreta e a outra contínua, são abordados. Além disso, uma revisão bibliográfica acerca das principais referências em modelos bivariados é apresentada.

No Capítulo 2 uma classe geral de modelos de regressão bivariados com respostas mistas é introduzida, em que se considera que os componentes sistemáticos das médias marginais das variáveis respostas são compostos por termos lineares e não lineares paramétricos conhecidos e também por funções não paramétricas de covariáveis disponíveis. A classe particular, considerando apenas os termos paramétricos nos preditores, é explorada inicialmente em que um método de estimação clássico é apresentado. Depois, é apresentada a outra classe particular de modelos bivariados em que os interceptos são formados por termos lineares e não paramétricos. Um método padrão de estimação univariado é apresentado em que são destacados alguns problemas encontrados.

No Capítulo 3 é apresentada a técnica de suavização P-spline. As principais características da técnica são ilustradas assim como sua aplicação ao caso univariado com preditor formado por uma ou mais funções suaves dos dados. Uma extensão da técnica P-spline para estimar as funções suaves nos modelos bivariados é então proposta. Todo o processo de estimação e inferência é apresentado.

No Capítulo 4 quatro modelos pertencentes à classe proposta são ilustrados e seus respectivos vetores escore e matrizes de informação esperada ou observada são mostrados. Os modelos tratados são Bernoulli-normal, Poisson-normal, Bernoulli-exponencial e Poisson-normal semiparamétrico.

No Capítulo 5 técnicas de diagnóstico e influência são propostas e adaptadas para a classe geral de modelos bivariados com respostas discreta e contínua.

No Capítulo 6 um estudo com dados simulados, considerando o modelo Bernoulli-exponencial e o modelo Poisson-normal semiparamétrico é apresentado, em que são verificados as propriedades dos estimadores de máxima verossimilhança e as performances dos ajustes e resíduos.

No Capítulo 7 uma aplicação a um conjunto de dados reais fornecidos por uma operadora de planos de saúde é realizada em que se ajusta o modelo Bernoulli-exponencial.

Gráficos dos resíduos e diagnósticos são mostrados.

No Capítulo 8 uma conclusão desta tese é feita, enquanto que no Capítulo 9 tópicos para a sequência do trabalho, com propostas de continuidade, são listadas.

No Apêndice A encontram-se as principais derivadas necessárias para o processo de estimação clássico apresentado no Capítulo 2.

Capítulo 2

Modelos bivariados mistos

Modelos bivariados com respostas mistas discreta e contínua, em que a distribuição conjunta do vetor de variáveis resposta é dada pelo produto entre a distribuição marginal da variável resposta discreta e a distribuição condicional da variável resposta contínua, condicionada à discreta, têm sido utilizados em alguns trabalhos (Fitzmaurice & Laird, 1995; Yang *et al.*, 2007). Entretanto, poucos modelos nesta linha são encontrados na literatura, assim como a ausência de um modelo bivariado com respostas discreta e contínua mais geral, englobando outros modelos bivariados já existentes na literatura.

O interesse em uma classe mais geral de modelos bivariados com respostas mistas se deve à grande variedade de conjuntos de dados disponíveis, tendo assim, a necessidade de modelos mais flexíveis e alternativos. Desta forma, uma ampla classe de modelos de regressão bivariados com respostas mistas discreta e contínua é definida neste capítulo. Para isto, é considerado que a distribuição da variável resposta discreta e a distribuição da variável resposta contínua, condicionada à discreta, pertencem à família exponencial. Além disso, covariáveis estão disponíveis e são relacionadas às médias marginais por meio de funções de ligação, em que os componentes sistemáticos são compostos por

termos lineares, não lineares e não paramétricos. Já a dependência entre as respostas é inserida na classe de modelos por meio da média condicional da variável resposta contínua, condicionada à discreta.

Sejam Y_1, Y_2, \dots, Y_n n variáveis aleatórias discretas e independentes tais que a função de probabilidade de Y_i , $i = 1, \dots, n$, pertence à família exponencial na forma canônica. Assim,

$$f_{Y_i}(y_i | \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}, \quad (2.1)$$

em que θ_i é o parâmetro natural e $b(\cdot)$ e $c(\cdot)$ são funções reais específicas. A função $a(\cdot)$ geralmente é da forma $a(\phi) = \phi$, sendo ϕ dito parâmetro de escala ou parâmetro de dispersão. Para algumas distribuições discretas tais como as distribuições de Poisson e binomial, ϕ é conhecido. Assim, a distribuição é dita pertencer à família exponencial uniparamétrica. Por outro lado, ϕ pode ser desconhecido. Nestes casos, a família não necessariamente será exponencial biparamétrica. Por simplicidade, será assumido que a distribuição de Y_i pertence à família exponencial uniparamétrica ou biparamétrica.

Denotando por μ_{iY} a média marginal da variável aleatória discreta Y_i , isto é, $\mu_{iY} = E(Y_i)$, considere que um conjunto de covariáveis $z_{i1}, z_{i2}, \dots, z_{ip}$, esteja disponível para predizer Y_i , para cada unidade i , $i = 1, \dots, n$. No contexto dos modelos aditivos generalizados semiparamétricos, considerando também um termo não linear no preditor, relaciona-se μ_{iY} às covariáveis $z_{i1}, z_{i2}, \dots, z_{ip}$ por $g_1(\mu_{iY}) = \eta_{iY}$ em que $g_1(\cdot)$ é uma função monótona e diferenciável denotada função de ligação e η_{iY} é o componente sistemático, dado por

$$\begin{aligned} \eta_{iY} &= \beta_0 + \beta_1 z_{i1} + \dots + \beta_k z_{ik} + h_1(z_{ik+1}, \dots, z_{im}, \beta_{k+1}, \dots, \beta_m) + h_2(z_{im+1}, \dots, z_{ip}) \\ &= \mathbf{Z}_{i1} \mathbf{B}_1 + h_1(\mathbf{Z}_{i2}, \mathbf{B}_2) + h_2(\mathbf{Z}_{i3}), \end{aligned} \quad (2.2)$$

em que $\mathbf{B}_1 = (\beta_0, \beta_1, \dots, \beta_k)^\top$ e $\mathbf{B}_2 = (\beta_{k+1}, \dots, \beta_m)^\top$ são vetores de coeficientes de regressão desconhecidos, $h_1(\cdot)$ é uma função não linear conhecida, duas vezes continu-

amente diferenciável com respeito aos elementos de \mathbf{B}_2 , $h_2(\cdot)$ é uma soma de funções univariadas desconhecidas, isto é, $h_2(z_{im+1}, \dots, z_{ip}) = \sum_{j_Y=m+1}^p s_{j_Y}(z_{ij_Y})$ sendo $s_{j_Y}(\cdot)$ funções suaves não especificadas de uma forma paramétrica, $j_Y = m+1, m+2, \dots, p$, e assim sendo, estimadas no contexto não paramétrico e $\mathbf{Z}_{i1} = (1, z_{i1}, \dots, z_{ik})$, $\mathbf{Z}_{i2} = (z_{ik+1}, \dots, z_{im})$ e $\mathbf{Z}_{i3} = (z_{im+1}, \dots, z_{ip})$ são vetores com os valores das covariáveis disponíveis para prever Y_i , $i = 1, 2, \dots, n$. Logo, o componente sistemático do modelo é mais geral no sentido em que é uma soma de termos lineares, de termos não lineares paramétricos e de termos não paramétricos (Hastie & Tibshirani, 1990).

A média μ_{iY} é relacionada ao parâmetro natural θ_i por

$$\mu_{iY} = b'(\theta_i) = \frac{d[b(\theta_i)]}{d\theta_i}.$$

Se $g_1(\cdot)$ for uma função de ligação canônica, segue que $\eta_{iY} = \theta_i$ e neste caso, é fácil expressar (2.1) em função de μ_{iY} ou de η_{iY} . A função

$$b''(\theta_i) = \frac{d^2[b(\theta_i)]}{d\theta_i^2}$$

é a função de variância e pode ser expressa em termos de μ_{iY} por $V(\mu_{iY})$, em que $Var(Y_i) = a(\phi)V(\mu_{iY})$, $i = 1, 2, \dots, n$.

Analogamente, sejam X_1, X_2, \dots, X_n n variáveis aleatórias contínuas e independentes tais que a função densidade condicional de X_i , dado $Y_i = y_i$, $i = 1, \dots, n$, pertence à família exponencial na forma canônica. Segue que

$$f_{X_i|Y_i}(x_i | y_i, \vartheta_i, \varphi) = \exp \left\{ \frac{x_i \vartheta_i - b^*(\vartheta_i)}{a^*(\varphi)} + c^*(x_i, \varphi) \right\}, \quad (2.3)$$

em que $b^*(\cdot)$ e $c^*(\cdot)$ são funções reais específicas, $\vartheta_i = \vartheta_i(y_i)$ é o parâmetro natural o qual depende do valor observado y_i , $i = 1, \dots, n$ e $a^*(\cdot)$ é uma função da forma $a^*(\varphi) = \varphi$, com φ sendo o parâmetro de escala conhecido ou desconhecido. Se φ é conhecido, a distribuição é dita pertencer a família exponencial uniparamétrica. Para φ desconhecido, a família não necessariamente será exponencial biparamétrica. Como

em (2.1), será considerado que a função densidade de $X_i | Y_i = y_i$ pertence à família exponencial uniparamétrica ou biparamétrica.

Assim como ocorre com a variável resposta discreta Y_i , assume-se que um conjunto de covariáveis $t_{i1}, t_{i2}, \dots, t_{iq}$ esteja disponível para prever a variável resposta contínua X_i , $i = 1, 2, \dots, n$. Denotando por $\mu_{iX} = E(X_i)$ a média marginal da variável resposta contínua X_i , relaciona-se μ_{iX} às covariáveis disponíveis por $g_2(\mu_{iX}) = \eta_{iX}$, em que $g_2(\cdot)$ é uma função de ligação monótona e diferenciável e η_{iX} é o componente sistemático, dado no contexto dos modelos aditivos generalizados semiparamétricos com um termo não linear por

$$\begin{aligned} \eta_{iX} &= \delta_0 + \delta_1 t_{i1} + \dots + \delta_d t_{id} + h_3(t_{id+1}, \dots, t_{io}, \delta_{d+1}, \dots, \delta_o) + h_4(t_{io+1}, \dots, t_{iq}) \\ &= \mathbf{T}_{i1} \mathbf{\Delta}_1 + h_3(\mathbf{T}_{i2}, \mathbf{\Delta}_2) + h_4(\mathbf{T}_{i3}), \end{aligned} \quad (2.4)$$

em que $\mathbf{\Delta}_1 = (\delta_0, \delta_1, \dots, \delta_d)^\top$ e $\mathbf{\Delta}_2 = (\delta_{d+1}, \dots, \delta_o)^\top$ são vetores de coeficientes de regressão desconhecidos, $h_3(\cdot)$ é uma função não linear conhecida, duas vezes continuamente diferenciável com respeito aos elementos do vetor $\mathbf{\Delta}_2$, $h_4(t_{io+1}, \dots, t_{iq}) = \sum_{j_X=o+1}^q u_{j_X}(t_{ij_X})$ sendo $u_{j_X}(\cdot)$ funções suaves não especificadas de uma forma paramétrica, $j_X = o+1, o+2, \dots, q$, e assim sendo, estimadas no contexto não paramétrico e $\mathbf{T}_{i1} = (1, t_{i1}, \dots, t_{id})$, $\mathbf{T}_{i2} = (t_{id+1}, \dots, t_{io})$ e $\mathbf{T}_{i3} = (t_{io+1}, \dots, t_{iq})$ são vetores com os valores das covariáveis disponíveis para prever X_i , $i = 1, 2, \dots, n$.

Em alguns trabalhos (Tate, 1954; Olkin & Tate, 1961), o objetivo é encontrar uma medida de associação entre as respostas discreta e contínua. Entretanto, esse não será o foco deste trabalho. Nós estamos interessados na inserção de uma estrutura de dependência entre as variáveis respostas discreta e contínua, necessária, uma vez que as variáveis podem ser observadas de um mesmo indivíduo ou de um mesmo elemento da população. Esta dependência é determinada no modelo pela média condicional da variável resposta contínua X_i dada a variável resposta discreta Y_i , denotada por $\alpha_i = E(X_i | Y_i = y_i)$.

Assume-se que α_i é relacionada à média marginal de X_i , à variável resposta discreta Y_i , à média marginal de Y_i e a um parâmetro γ incluído no modelo, que pode ser o coeficiente de correlação ou outra medida de associação entre as variáveis resposta X_i e Y_i , dependendo da natureza dos dados, por $g_3(\alpha_i) = \xi_i$, em que $g_3(\cdot)$ é uma função de ligação monótona e diferenciável e ξ_i é o componente sistemático, dado por

$$\xi_i = h_5(y_i, \mu_{iY}, \mu_{iX}, \gamma), \quad (2.5)$$

em que $h_5(\cdot)$ é uma função linear ou não linear conhecida, de classe C^2 , do valor observado da variável aleatória y_i , das médias marginais μ_{iY} e μ_{iX} e do parâmetro desconhecido γ . As escolhas das funções $g_3(\cdot)$ e $h_5(\cdot)$ devem ser tais que a igualdade $E(\alpha_i) = \mu_{iX}$ seja verificada, já que $E[E(X_i | Y_i = y_i)] = E(X_i)$ e por construção, $E(X_i) = \mu_{iX}$, em que

$$\alpha_i = g_3^{-1}(h_5(y_i, \mu_{iY}, \mu_{iX}, \gamma)). \quad (2.6)$$

A média condicional α_i é relacionada ao parâmetro natural ϑ_i por

$$\alpha_i = b^*(\vartheta_i) = \frac{d[b^*(\vartheta_i)]}{d\vartheta_i}.$$

Se $g_3(\cdot)$ for uma função de ligação canônica, segue que $\xi_i = \vartheta_i$ e assim, é fácil expressar (2.3) em função de α_i ou de ξ_i . A função

$$b^{*''}(\vartheta_i) = \frac{d^2[b^*(\vartheta_i)]}{d\vartheta_i^2}$$

é a função de variância e pode ser expressa em termos de α_i por $Q(\alpha_i)$, em que $Var(X_i | Y_i = y_i) = a^*(\varphi)Q(\alpha_i)$.

Utilizando as funções (2.1) e (2.3), é possível expressar a função densidade de probabilidade conjunta de (X_i, Y_i) como

$$\begin{aligned} f_{(X_i, Y_i)}(x_i, y_i | \theta_i, \vartheta_i, \phi, \varphi) &= f_{Y_i}(y_i | \theta_i, \phi) f_{X_i|Y_i}(x_i | y_i, \vartheta_i, \varphi) \\ &= \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \exp \left\{ \frac{x_i \vartheta_i - b^*(\vartheta_i)}{a^*(\varphi)} + c^*(x_i, \varphi) \right\} \\ &= \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \frac{x_i \vartheta_i - b^*(\vartheta_i)}{a^*(\varphi)} + c(y_i, \phi) + c^*(x_i, \varphi) \right\}. \end{aligned} \quad (2.7)$$

2.1 Modelo paramétrico

Um caso particular da classe geral de modelos bivariados com respostas discreta e contínua apresentada acima ocorre quando $h_2(\cdot) \equiv h_4(\cdot) \equiv 0$ nas equações (2.2) e (2.4), respectivamente, isto é, quando os componentes sistemáticos η_{iY} e η_{iX} são compostos por termos lineares, como ocorre nos modelos lineares generalizados (McCullagh & Nelder, 1989) e por funções não lineares conhecidas. Os componentes sistemáticos são então, formados apenas por termos paramétricos. Considerando que p covariáveis estejam disponíveis para prever Y_i , $i = 1, 2, \dots, n$, o preditor η_{iY} dado pela equação (2.2) se reduz a

$$\begin{aligned}\eta_{iY} &= \beta_0 + \beta_1 z_{i1} + \dots + \beta_k z_{ik} + h_1(z_{ik+1}, \dots, z_{ip}, \beta_{k+1}, \dots, \beta_p) \\ &= \mathbf{Z}_{i1} \mathbf{B}_1 + h_1(\mathbf{Z}_{i2}, \mathbf{B}_2),\end{aligned}\tag{2.8}$$

com $\mathbf{B}_1 = (\beta_0, \beta_1, \dots, \beta_k)^\top$, $\mathbf{B}_2 = (\beta_{k+1}, \dots, \beta_p)^\top$, $\mathbf{Z}_{i1} = (1, z_{i1}, \dots, z_{ik})$ e $\mathbf{Z}_{i2} = (z_{ik+1}, \dots, z_{ip})$. Analogamente, supondo que q covariáveis estejam disponíveis para prever a variável resposta contínua X_i , $i = 1, 2, \dots, n$, o preditor η_{iX} dado pela equação (2.4) se reduz a

$$\begin{aligned}\eta_{iX} &= \delta_0 + \delta_1 t_{i1} + \dots + \delta_d t_{id} + h_3(t_{id+1}, \dots, t_{iq}, \delta_{d+1}, \dots, \delta_q) \\ &= \mathbf{T}_{i1} \mathbf{\Delta}_1 + h_3(\mathbf{T}_{i2}, \mathbf{\Delta}_2),\end{aligned}\tag{2.9}$$

com $\mathbf{\Delta}_1 = (\delta_0, \delta_1, \dots, \delta_d)^\top$, $\mathbf{\Delta}_2 = (\delta_{d+1}, \dots, \delta_q)^\top$, $\mathbf{T}_{i1} = (1, t_{i1}, \dots, t_{id})$ e $\mathbf{T}_{i2} = (t_{id+1}, \dots, t_{iq})$. Para estimar os parâmetros do modelo apresentado nesta seção, utilizamos o método de máxima verossimilhança.

2.1.1 Máxima verossimilhança

Sejam $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ um conjunto de valores observados de $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, com (X_i, Y_i) independente de (X_j, Y_j) , para $i, j = 1, \dots, n$, $i \neq j$ e função densidade de probabilidade conjunta dada pela expressão (2.7). A função de

verossimilhança de $\boldsymbol{\beta} = (\mathbf{B}_1^\top, \mathbf{B}_2^\top)^\top$, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1^\top, \boldsymbol{\Delta}_2^\top)^\top$, γ , ϕ e φ pode ser escrita como

$$\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T}) &= \prod_{i=1}^n L_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid x_i, y_i, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \mathbf{T}_{i1}, \mathbf{T}_{i2}) \\
&= \prod_{i=1}^n f_{(X_i, Y_i)}(x_i, y_i \mid \boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \mathbf{T}_{i1}, \mathbf{T}_{i2}) \\
&= \prod_{i=1}^n \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + \frac{x_i \vartheta_i - b^*(\vartheta_i)}{a^*(\varphi)} + c(y_i, \phi) + c^*(x_i, \varphi) \right\} \\
&= \exp \left\{ \frac{1}{a(\phi)} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \frac{1}{a^*(\varphi)} \sum_{i=1}^n [x_i \vartheta_i - b^*(\vartheta_i)] \right\} \\
&\quad \times \exp \left\{ \sum_{i=1}^n c(y_i, \phi) + \sum_{i=1}^n c^*(x_i, \varphi) \right\}, \tag{2.10}
\end{aligned}$$

em que $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$, $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$, $\mathbf{Z}_1 = (\mathbf{Z}_{11}^\top, \mathbf{Z}_{21}^\top, \dots, \mathbf{Z}_{n1}^\top)^\top$, $\mathbf{Z}_2 = (\mathbf{Z}_{12}^\top, \mathbf{Z}_{22}^\top, \dots, \mathbf{Z}_{n2}^\top)^\top$, $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, $\mathbf{T}_1 = (\mathbf{T}_{11}^\top, \mathbf{T}_{21}^\top, \dots, \mathbf{T}_{n1}^\top)^\top$, $\mathbf{T}_2 = (\mathbf{T}_{12}^\top, \mathbf{T}_{22}^\top, \dots, \mathbf{T}_{n2}^\top)^\top$ e $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2)$.

A função log-verossimilhança de $\boldsymbol{\beta} = (\mathbf{B}_1^\top, \mathbf{B}_2^\top)^\top$, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1^\top, \boldsymbol{\Delta}_2^\top)^\top$, γ , ϕ e φ , dada uma amostra de tamanho n , pode ser escrita como

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T}) &= \sum_{i=1}^n \ell_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid x_i, y_i, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \mathbf{T}_{i1}, \mathbf{T}_{i2}) \\
&= \frac{1}{a(\phi)} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \frac{1}{a^*(\varphi)} \sum_{i=1}^n [x_i \vartheta_i - b^*(\vartheta_i)] \\
&\quad + \sum_{i=1}^n c(y_i, \phi) + \sum_{i=1}^n c^*(x_i, \varphi). \tag{2.11}
\end{aligned}$$

As equações de máxima verossimilhança são facilmente obtidas e formam um conjunto de $p + q + 5$ equações correspondentes aos parâmetros $\beta_0, \beta_1, \dots, \beta_p, \delta_0, \delta_1, \dots, \delta_q$, γ , ϕ e φ . Seja $\mathbf{U} = (\mathbf{U}(\boldsymbol{\beta})^\top, \mathbf{U}(\boldsymbol{\Delta})^\top, \mathbf{U}(\gamma), \mathbf{U}(\phi), \mathbf{U}(\varphi))^\top$ o vetor escore da função log-verossimilhança dada na equação (2.11). Os elementos dos vetores $\mathbf{U}(\boldsymbol{\beta})$ e $\mathbf{U}(\boldsymbol{\Delta})$ e os componentes $\mathbf{U}(\gamma)$, $\mathbf{U}(\phi)$ e $\mathbf{U}(\varphi)$ são dados em detalhes no Apêndice A. Em forma matricial, os componentes de \mathbf{U} são dados por

$$\mathbf{U}(\beta) = \mathbf{S}_1^\top \left(\frac{1}{a(\phi)} \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu}_Y) + \frac{1}{a^*(\phi)} \mathbf{R}^{\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} \mathbf{J}_Y (\mathbf{x} - \boldsymbol{\alpha}) \right), \quad (2.12)$$

$$\mathbf{U}(\boldsymbol{\Delta}) = \frac{1}{a^*(\phi)} \mathbf{S}_2^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} \mathbf{J}_X \mathbf{O} (\mathbf{x} - \boldsymbol{\alpha}), \quad (2.13)$$

$$\mathbf{U}(\gamma) = \frac{1}{a^*(\phi)} \mathbf{J}_\gamma^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} (\mathbf{x} - \boldsymbol{\alpha}), \quad (2.14)$$

$$\mathbf{U}(\phi) = -\frac{a'(\phi)}{a(\phi)^2} (\mathbf{y}^\top \boldsymbol{\theta} - \mathbf{b}^\top \mathbf{1}) + \mathbf{c}^\top \mathbf{1}, \quad (2.15)$$

$$\text{e } \mathbf{U}(\varphi) = -\frac{a^{*\prime}(\varphi)}{a^*(\varphi)^2} (\mathbf{x}^\top \boldsymbol{\vartheta} - \mathbf{b}^{*\top} \mathbf{1}) + \mathbf{c}^{*\top} \mathbf{1}, \quad (2.16)$$

em que

$$\boldsymbol{\mu}_Y = (\mu_{1Y}, \mu_{2Y}, \dots, \mu_{nY})^\top, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^\top, \quad \mathbf{J}_\gamma = \left(\frac{d\xi_1}{d\gamma}, \frac{d\xi_2}{d\gamma}, \dots, \frac{d\xi_n}{d\gamma} \right)^\top,$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)^\top, \quad \mathbf{b} = (b(\theta_1), b(\theta_2), \dots, b(\theta_n))^\top, \quad \boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_n)^\top,$$

$$\mathbf{b}^* = (b^*(\vartheta_1), b^*(\vartheta_2), \dots, b^*(\vartheta_n))^\top, \quad \mathbf{c} = (c'(y_1, \phi), c'(y_2, \phi), \dots, c'(y_n, \phi))^\top,$$

$$\mathbf{c}^* = (c^*(x_1, \varphi), c^*(x_2, \varphi), \dots, c^*(x_n, \varphi))^\top, \quad \text{e } \mathbf{1} = (1, 1, \dots, 1)^\top,$$

são vetores n -dimensionais, com $c'(y_i, \phi) = \frac{d[c(y_i, \phi)]}{d\phi}$ e $c^*(x_i, \varphi) = \frac{d[c^*(x_i, \varphi)]}{d\varphi}$,

$$\mathbf{S}_1 = \begin{pmatrix} \frac{\partial \eta_{1Y}}{\partial \beta_0} & \frac{\partial \eta_{2Y}}{\partial \beta_0} & \cdots & \frac{\partial \eta_{nY}}{\partial \beta_0} \\ \frac{\partial \eta_{1Y}}{\partial \beta_1} & \frac{\partial \eta_{2Y}}{\partial \beta_1} & \cdots & \frac{\partial \eta_{nY}}{\partial \beta_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_{1Y}}{\partial \beta_p} & \frac{\partial \eta_{2Y}}{\partial \beta_p} & \cdots & \frac{\partial \eta_{nY}}{\partial \beta_p} \end{pmatrix}^\top \quad \text{e } \mathbf{S}_2 = \begin{pmatrix} \frac{\partial \eta_{1X}}{\partial \delta_0} & \frac{\partial \eta_{2X}}{\partial \delta_0} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_0} \\ \frac{\partial \eta_{1X}}{\partial \delta_1} & \frac{\partial \eta_{2X}}{\partial \delta_1} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_{1X}}{\partial \delta_q} & \frac{\partial \eta_{2X}}{\partial \delta_q} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_q} \end{pmatrix}^\top,$$

são matrizes $n \times (p+1)$ e $n \times (q+1)$, respectivamente,

$$\mathbf{W} = \text{diag}(w_1, w_2, \dots, w_n), \quad \mathbf{V} = \text{diag}(V_1, V_2, \dots, V_n), \quad \mathbf{R} = \text{diag}(r_1, r_2, \dots, r_n),$$

$$\mathbf{Q} = \text{diag}(Q_1, Q_2, \dots, Q_n), \quad \mathbf{O} = \text{diag} \left(\frac{d\mu_{1X}}{d\eta_{1X}}, \frac{d\mu_{2X}}{d\eta_{2X}}, \dots, \frac{d\mu_{nX}}{d\eta_{nX}} \right),$$

$$\mathbf{J}_Y = \text{diag} \left(\frac{d\xi_1}{d\mu_{1Y}}, \frac{d\xi_2}{d\mu_{2Y}}, \dots, \frac{d\xi_n}{d\mu_{nY}} \right) \quad \text{e } \mathbf{J}_X = \text{diag} \left(\frac{d\xi_1}{d\mu_{1X}}, \frac{d\xi_2}{d\mu_{2X}}, \dots, \frac{d\xi_n}{d\mu_{nX}} \right),$$

são matrizes diagonais, em que $V_i = V(\mu_{iY}) = b''(\theta_i) = \frac{d\mu_{iY}}{d\theta_i}$, é a função de variância

de Y_i e $Q_i = Q(\alpha_i) = b^{*\prime\prime}(\vartheta_i) = \frac{d\alpha_i}{d\vartheta_i}$ é a função de variância $X_i | Y_i$, w_i e r_i são pesos

definidos por

$$w_i = \frac{1}{V_i} \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \quad \text{e} \quad r_i = \frac{1}{Q_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2, \quad i = 1, \dots, n,$$

$$a'(\phi) = \frac{d[a(\phi)]}{d\phi} \quad \text{e} \quad a^{*\prime}(\varphi) = \frac{d[a^*(\varphi)]}{d\varphi}.$$

Observação:

Se $g_1(\cdot)$ e $g_3(\cdot)$ são as funções de ligação canônicas para os modelos associados a Y_i e $X_i \mid Y_i = y_i$, respectivamente, $i = 1, \dots, n$, segue que $\mathbf{W} = \mathbf{V}$ e $\mathbf{R} = \mathbf{Q}$ e, desta forma, as equações (2.12), (2.13), e (2.14) podem ser reescritas como

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{S}_1^\top \left(\frac{1}{a(\phi)} (\mathbf{y} - \boldsymbol{\mu}_Y) + \frac{1}{a^*(\varphi)} \mathbf{V} \mathbf{J}_Y (\mathbf{x} - \boldsymbol{\alpha}) \right), \quad (2.17)$$

$$\mathbf{U}(\boldsymbol{\Delta}) = \frac{1}{a^*(\varphi)} \mathbf{S}_2^\top \mathbf{J}_X \mathbf{O} (\mathbf{x} - \boldsymbol{\alpha}), \quad (2.18)$$

$$\text{e } \mathbf{U}(\boldsymbol{\gamma}) = \frac{1}{a^*(\varphi)} \mathbf{J}_\gamma^\top (\mathbf{x} - \boldsymbol{\alpha}). \quad (2.19)$$

Os componentes $\mathbf{U}(\phi)$ e $\mathbf{U}(\varphi)$ são os mesmos apresentados nas equações (2.15) e (2.16), respectivamente.

O estimador de máxima verossimilhança para $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, $\boldsymbol{\gamma}$, ϕ e φ pode ser obtido da solução do sistema de equações $\mathbf{U} = \mathbf{0}$. Geralmente, esse sistema é não linear e as soluções podem ser obtidas numericamente, usando métodos iterativos tais como Newton-Raphson e escore de Fisher. Entretanto, tais métodos requerem que as matrizes de informação esperada e/ou observadas, necessárias nos algoritmos, sejam definidas positivas.

A matriz de covariâncias assintótica do estimador de máxima verossimilhança para os parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, $\boldsymbol{\gamma}$, ϕ e φ é obtida pela inversão da matriz de informação esperada. Entretanto, para algumas escolhas de distribuições e funções de ligação, a obtenção desta se torna complexa. Nestes casos, a matriz de covariâncias assintótica pode ser obtida pela inversão da matriz de informação observada, desde que esta matriz seja

definida positiva. No Apêndice A encontram-se todas as derivadas necessárias. Considerando a matriz de informação observada, denotada por \mathbf{M} , tem-se que a matriz de covariâncias assintótica de $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{\phi}}$ e $\widehat{\boldsymbol{\varphi}}$ é dada por

$$\text{cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Delta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\varphi}}) = \mathbf{M}^{-1} = \begin{bmatrix} \mathbf{M}_{\boldsymbol{\beta}} & \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\Delta}} & \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\gamma}} & \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\phi}} & \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\varphi}} \\ \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\Delta}}^{\top} & \mathbf{M}_{\boldsymbol{\Delta}} & \mathbf{M}_{\boldsymbol{\Delta}\boldsymbol{\gamma}} & \mathbf{0}_{(q+1)\times 1} & \mathbf{M}_{\boldsymbol{\Delta}\boldsymbol{\varphi}} \\ \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\gamma}}^{\top} & \mathbf{M}_{\boldsymbol{\Delta}\boldsymbol{\gamma}}^{\top} & \mathbf{M}_{\boldsymbol{\gamma}} & 0 & \mathbf{M}_{\boldsymbol{\gamma}\boldsymbol{\varphi}} \\ \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\phi}}^{\top} & \mathbf{0}_{1\times(q+1)} & 0 & \mathbf{M}_{\boldsymbol{\phi}} & 0 \\ \mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\varphi}}^{\top} & \mathbf{M}_{\boldsymbol{\Delta}\boldsymbol{\varphi}}^{\top} & \mathbf{M}_{\boldsymbol{\gamma}\boldsymbol{\varphi}}^{\top} & 0 & \mathbf{M}_{\boldsymbol{\varphi}} \end{bmatrix}^{-1}, \quad (2.20)$$

com as submatrizes $\mathbf{M}_{\boldsymbol{\beta}}$, $\mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\Delta}}$, $\mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\gamma}}$, $\mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\phi}}$, $\mathbf{M}_{\boldsymbol{\beta}\boldsymbol{\varphi}}$, $\mathbf{M}_{\boldsymbol{\Delta}}$, $\mathbf{M}_{\boldsymbol{\Delta}\boldsymbol{\gamma}}$, $\mathbf{M}_{\boldsymbol{\Delta}\boldsymbol{\varphi}}$, $\mathbf{M}_{\boldsymbol{\gamma}}$, $\mathbf{M}_{\boldsymbol{\gamma}\boldsymbol{\varphi}}$, $\mathbf{M}_{\boldsymbol{\phi}}$ e $\mathbf{M}_{\boldsymbol{\varphi}}$, dadas pela equação (A.14). A matriz de informação esperada, $\mathbf{M}_{\mathbf{E}}$, é obtida substituindo os elementos da matriz \mathbf{M} por seus respectivos valores esperados com respeito às distribuições das variáveis respostas X_i e Y_i . Neste caso, a matriz de covariâncias assintótica na equação (2.20) é obtida substituindo a matriz \mathbf{M} pela matriz $\mathbf{M}_{\mathbf{E}}$. A matriz de covariâncias assintótica é estimada substituindo os parâmetros por suas respectivas estimativas de máxima verossimilhança.

2.1.2 Intervalos de confiança assintóticos

Nesta subseção apresentamos a construção dos intervalos de confiança assintóticos dos estimadores de máxima verossimilhança dos parâmetros dos modelos pertencentes à classe bivariada proposta no Capítulo 2, considerando o caso em que ambos os preditores são formados apenas por termos paramétricos, como detalhado na Subseção 2.1.

É importante destacar que este intervalo é baseado na distribuição assintótica dos estimadores de máxima verossimilhança dos parâmetros. Desta forma, para determinados valores de amostra, os resultados podem não ser adequados. Para a classe de

modelos proposta, as condições de regularidade estão satisfeitas. De fato, as distribuições das variáveis resposta discreta Y_i e da variável resposta contínua X_i , dado a variável discreta Y_i , pertencem à família exponencial de distribuições e como tal, o suporte de ambas é independente dos parâmetros do modelo. Além disso, pode-se notar que $\int_{\Omega_{X_i}} \sum_{\Omega_{Y_i}} f_{(Y_i, X_i)(y_i, x_i)} = 1$, com $f_{(Y_i, X_i)(\cdot)}$ dada pela equação (2.7), em que Ω_{Y_i} e Ω_{X_i} são os domínios das variáveis Y_i e X_i , respectivamente. Por fim, observa-se que a troca das operações de derivação e integração sob a distribuição conjunta do par (Y_i, X_i) é possível. Assim, sob tais condições segue que para amostras grandes,

$$\sqrt{n}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) \sim N_{dm}(\mathbf{0}, \mathbf{M}^{-1}),$$

em que $\boldsymbol{\nu}$ é o vetor formado pelos parâmetros do modelo, isto é, por $\beta_0, \dots, \beta_p, \delta_0, \dots, \delta_q, \gamma, \phi$ e φ , e $\mathbf{M} = \mathbf{M}(\boldsymbol{\nu})$ é a matriz de informação esperada (ou observada), de dimensão $dm \times dm$, dada como na equação (2.20). Os intervalos de confiança com nível de confiança $100 \times (1 - \tilde{\alpha})\%$ são então dados por

$$\hat{\nu}_j \pm \mathbb{Z}_{\tilde{\alpha}/2} \sqrt{\mathbf{M}_j^{-1}(\hat{\boldsymbol{\nu}})},$$

em que $\hat{\nu}_j$ é a estimativa de máxima verossimilhança do j -ésimo parâmetro do vetor de parâmetros $\boldsymbol{\nu}$, $\mathbb{Z}_{\tilde{\alpha}/2}$ é o $(\tilde{\alpha}/2)$ -ésimo quantil superior da distribuição normal padrão e $\mathbf{M}_j^{-1}(\hat{\boldsymbol{\nu}})$ é o j -ésimo elemento da diagonal principal da inversa da matriz de informação \mathbf{M} avaliada nas estimativas de máxima verossimilhança dos parâmetros do modelo, $j = 1, 2, \dots, p + q + 5$.

2.1.3 Teste de hipóteses

Nesta subseção apresentamos o teste da razão de verossimilhanças para testar algumas hipóteses de interesse. Por exemplo, pode-se estar interessado em testar se a parte linear do preditor associado a uma das variáveis resposta é significativa ou, se o parâmetro γ , associado a alguma medida de associação entre as respostas, é igual a um determinado valor. A não linearidade dos preditores pode ser testada pelas hipóteses

$$\text{Hipóteses 1} = \begin{cases} H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0 & \text{contra} \\ H_1 : \beta_{j_Y} \neq 0 & \text{para pelo menos um } j_Y \in \{1, \dots, k\}, \end{cases}$$

no caso do preditor associado à variável resposta discreta Y_i , e

$$\text{Hipóteses 2} = \begin{cases} H_0 : \delta_0 = \delta_1 = \dots = \delta_d = 0 & \text{contra} \\ H_1 : \delta_{j_X} \neq 0 & \text{para pelo menos um } j_X \in \{1, \dots, d\}, \end{cases}$$

no caso do preditor associado à resposta contínua X_j . Para testar se o parâmetro de associação γ é igual a um determinado valor, tem-se as hipóteses

$$\text{Hipóteses 3} = \begin{cases} H_0 : \gamma = \gamma_0 & \text{contra} \\ H_1 : \gamma \neq \gamma_0 . \end{cases}$$

Seja ν_j o j -ésimo parâmetro do vetor de parâmetros $\boldsymbol{\nu}$, como definido na Subseção 2.1.2, e suponha que estamos interessados em testar as hipóteses

$$\text{Hipóteses 4} = \begin{cases} H_0 : \nu_j = \nu_j^0 & \text{contra} \\ H_1 : \nu_j \neq \nu_j^0 , \end{cases}$$

em que ν_j^0 é uma constante conhecida. O teste da razão de verossimilhanças para testar as Hipóteses 4 é obtido a partir da estatística

$$\text{RV} = -2\{\ell(\tilde{\boldsymbol{\nu}} \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T}) - \ell(\hat{\boldsymbol{\nu}} \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T})\},$$

em que $\ell(\tilde{\boldsymbol{\nu}} \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T})$ e $\ell(\hat{\boldsymbol{\nu}} \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T})$, são as funções de log-verossimilhança calculadas nos estimadores de máxima verossimilhança sob H_0 e nos estimadores de máxima verossimilhança irrestritos, respectivamente. Sob H_0 , RV tem distribuição assintótica χ^2 com o número de graus de liberdade igual ao número de restrições da hipótese testada. Assim, rejeitamos H_0 com um nível de significância $\tilde{\alpha}$ se o valor de RV é maior do que o quantil $(1 - \tilde{\alpha})$ da distribuição χ^2 .

2.2 Modelo semiparamétrico

Considere agora que os preditores η_{iY} e η_{iX} são formados apenas pelo intercepto e pelas equações $h_2(\cdot)$ e $h_4(\cdot)$, respectivamente, como nos modelos aditivos generalizados (Hastie & Tibshirani, 1990). Neste caso, ambos os componentes sistemáticos são formados por uma soma de funções suaves não especificadas das covariáveis.

Supondo que p covariáveis estejam disponíveis para prever Y_i , $i = 1, 2, \dots, n$, o preditor η_{iY} dado pela equação (2.2) se reduz a

$$\eta_{iY} = \beta_0 + h_2(z_{i1}, \dots, z_{ip}) = \beta_0 + \sum_{j_Y=1}^p s_{j_Y}(z_{ij_Y}), \quad (2.21)$$

em que $s_{j_Y}(\cdot)$, $j_Y = 1, 2, \dots, p$, são funções univariadas desconhecidas não especificadas de uma forma paramétrica. Analogamente, supondo que q covariáveis estejam disponíveis para prever a variável resposta contínua X_i , $i = 1, 2, \dots, n$, o preditor η_{iX} dado pela equação (2.4) se reduz a

$$\eta_{iX} = \delta_0 + h_4(t_{i1}, \dots, t_{iq}) = \delta_0 + \sum_{j_X=1}^q u_{j_X}(t_{ij_X}), \quad (2.22)$$

em que $u_{j_X}(\cdot)$, $j_X = 1, 2, \dots, q$, são funções suaves não especificadas de uma forma paramétrica.

A inserção do intercepto se faz necessária no contexto dos modelos aditivos generalizados para se evitar o problema de identificabilidade (Hastie & Tibshirani, 1990). Além disso, assume-se que $\sum_{i=1}^n s_{j_Y}(z_{ij_Y}) = 0$, para todo $j_Y = 1, 2, \dots, p$, e $\sum_{i=1}^n u_{j_X}(t_{ij_X}) = 0$, para todo $j_X = 1, 2, \dots, q$, a fim de garantir a unicidade das curvas ajustadas.

2.2.1 Escore local

Considerando os modelos associados às variáveis respostas discreta e contínua individualmente, pode-se estimar as funções $s_{j_Y}(\cdot)$, $j_Y = 1, 2, \dots, p$, e $u_{j_X}(\cdot)$, $j_X = 1, 2, \dots, q$, não especificadas de uma forma paramétrica, através de um suavizador

(*smoother*), uma ferramenta que representa a tendência das variáveis respostas Y_i e X_i como funções das covariáveis disponíveis.

No caso em que apenas uma covariável está disponível para prever cada uma das variáveis respostas, por exemplo, z_{i1} e t_{i1} com $\eta_{iY} = s_1(z_{i1})$ e $\eta_{iX} = u_1(t_{i1})$, $i = 1, 2, \dots, n$, um suavizador do gráfico de dispersão é frequentemente utilizado. Neste caso, as estimativas obtidas (curvas ajustada) $\hat{s}_{j_Y}(\cdot)$, $j_Y = 1, 2, \dots, p$, e $\hat{u}_{j_X}(\cdot)$, $j_X = 1, 2, \dots, q$, são ditas suavização do gráfico de dispersão e são o grande interesse da regressão não paramétrica uma vez que não se tem parâmetros estimados. É importante destacar que as estimativas das tendências terão menos variabilidade que as variáveis respostas observadas, o que explica o nome de suavizador para a técnica aplicada (Hastie & Tibshirani, 1990).

Dentre os principais suavizadores do gráfico de dispersão destaca-se a técnica de regressão polinomial local, conhecida como *loess* (Cleveland, 1979), suavização com núcleos ou suavizadores tipo *kernel*, splines de regressão (*regression splines*) com destaque para os splines cúbicos e splines de suavização (*smoothing splines*). Para mais detalhes destas técnicas de suavização e de outras, vide Hastie & Tibshirani (1990) e Green & Silverman (1994).

Quando mais de uma covariável está disponível para prever as respostas Y_i e X_i , com preditores dados pelas equações (2.21) e (2.22), isto é, por uma soma de termos univariados no contexto dos modelos aditivos generalizados, frequentemente utiliza-se o algoritmo *escore local* para estimar cada função suave $s_{j_Y}(\cdot)$, $j_Y = 1, 2, \dots, p$, e $u_{j_X}(\cdot)$, $j_X = 1, 2, \dots, q$, em um cenário não paramétrico, além dos parâmetros β_0 e δ_0 referentes aos interceptos.

O algoritmo *escore local* (Hastie & Tibshirani, 1990) mescla o algoritmo *retroajuste*, utilizado em modelos aditivos (Hastie & Tibshirani, 1987; Buja *et al.*, 1989), com o método iterativo *escore de Fisher*, utilizado em modelos lineares generalizados

(McCullagh & Nelder, 1989). O que difere do ajuste de um modelo linear generalizado é a inserção de um ciclo retroajuste em cada iteração do algoritmo escore. Desta forma, o algoritmo escore local para o ajuste do modelo associado à variável resposta Y_i , $i = 1, 2, \dots, n$, é dado por

1. Adotam-se os valores iniciais $\beta_0^{(0)} = g_1(\sum_{i=1}^n y_i/n)$ e $s_1^{(0)}(\cdot) = s_2^{(0)}(\cdot) = \dots = s_p^{(0)}(\cdot) = 0$. Consequentemente, tem-se $\eta_{iY}^{(0)}$, o valor inicial do preditor dado pela equação (2.21) e $\mu_{iY}^{(0)} = g_1^{-1}(\eta_{iY}^{(0)})$;
2. Constrói-se uma variável resposta modificada, denotada por \mathbf{y}^* , a partir dos valores de $\eta_{iY}^{(0)}$ e $\mu_{iY}^{(0)}$ e pesos \mathbf{W} , análogos aos apresentados na Subseção 2.1.1, como nos modelos lineares generalizados. Note que a variável resposta modificada deve ser atualizada em cada iteração assim como os pesos;
3. Aplica-se um ciclo retroajuste, isto é, para cada $j_Y = 1, 2, \dots, p$, atualizam-se as funções $s_{j_Y}(\cdot)$ suavizando $\mathbf{y}^* - \beta_0 - \sum_{jj \neq j_Y} s_{jj}^{(0)} \mid (z_{jj}, \widehat{\mathbf{W}})$ por meio de algum suavizador do gráfico de dispersão, resultando em novas funções suaves $s_1^{(1)}(\cdot), s_2^{(1)}(\cdot), \dots, s_p^{(1)}(\cdot)$. Para acelerar a convergência, podem-se utilizar as funções suaves atualizadas, por exemplo, $s_1^{(1)}(\cdot)$ ao invés de $s_1^{(0)}(\cdot)$ no cálculo de $s_2^{(1)}(\cdot)$. Consequentemente, um novo valor de $\eta_{iY}^{(1)}$ e $\mu_{iY}^{(1)}$ é obtido;
4. Repetem-se os passos (ii)-(iii) substituindo $\eta_{iY}^{(0)}$ por $\eta_{iY}^{(1)}$ e $\mu_{iY}^{(0)}$ por $\mu_{iY}^{(1)}$, até que se obtenha a convergência.

Para o critério de convergência, pode-se adotar por exemplo a razão

$$\frac{\sum_{j_Y=1}^p \left\| s_{j_Y}^{(1)}(\cdot) - s_{j_Y}^{(0)}(\cdot) \right\|}{\sum_{j_Y=1}^p \left\| s_{j_Y}^{(0)}(\cdot) \right\|},$$

em que $\|s_{j_Y}(\cdot)\|$ é a norma euclidiana do vetor resultante do cálculo de $s_{j_Y}(\cdot)$ nos n pontos amostrais. Assim, é dito que o algoritmo convergiu quando o valor da razão é

menor do que um determinado valor de erro adotado.

Para o modelo associado à variável resposta contínua X_i , $i = 1, 2, \dots, n$, o procedimento de ajuste é análogo ao apresentado acima. Note porém que, quando a variável resposta tem distribuição normal, o algoritmo escore local se reduz ao método retroajuste.

Existe uma grande eficiência computacional em ambos os algoritmos. Entretanto, por se tratar de métodos iterativos, as expressões explícitas das curvas suaves não estão disponíveis. Além disso, as técnicas de suavização do gráfico de dispersão utilizadas no processo de estimação podem não ser viáveis em alguns problemas. A técnica splines de suavização, por exemplo, aplicada para estimar cada uma das curvas no ciclo retroajuste, reduz o problema para o contexto dos modelos lineares generalizados com uma penalização. Entretanto, a quantidade de parâmetros (associados à base) é proporcional ao tamanho da amostra, não sendo indicado em problemas com grandes quantidades de dados (Green & Silverman, 1994). Por outro lado, a técnica de splines de regressão também reduz o problema para o contexto dos MLG, porém, determinar a quantidade e a posição dos nós pode ser uma tarefa difícil.

Procurando solucionar tais problemas, apresentamos no próximo capítulo uma técnica de suavização bastante eficiente e muito utilizada em regressão não paramétrica e propomos o seu uso na classe de modelos bivariados semiparamétricos com respostas discreta e contínua apresentada nesta seção.

Capítulo 3

Modelos bivariados semiparamétricos

Neste capítulo, apresentamos em detalhes a técnica de suavização splines com penalização ou simplesmente P-splines, proposta por Eilers & Marx (1996) e bastante utilizada na regressão não paramétrica (Marx & Eilers, 1998; Eilers & Marx, 2002; Currie & Durban, 2002; Currie *et al.*, 2006). Esta técnica é capaz de solucionar alguns dos problemas apresentados por outras técnicas de suavização, como mencionado na Subseção 2.2.1. Sua aplicação no contexto dos modelos aditivos generalizados, proposta em Marx & Eilers (1998) e denotada por *P-GAM*, também é apresentada. Por fim, propomos a aplicação da técnica P-splines à nossa classe de modelos bivariados semiparamétricos com respostas discreta e contínua, considerada na Seção 2.2.

3.1 P-splines

A técnica de suavização P-splines é caracterizada pelo uso de uma base B-spline para a regressão, construída a partir das covariáveis disponíveis e uma penalização discreta inserida na função log-verossimilhança e necessária para controlar a suavidade do ajuste, reduzindo a importância da escolha da quantidade e localização dos nós.

Nesta técnica, o cenário não paramétrico é reduzido a um problema no contexto dos modelos lineares generalizados em que os coeficientes da regressão são penalizados.

3.1.1 Base B-spline

Sucintamente, um spline é uma função polinomial por partes (combinação linear) em que os polinômios são conectados por um conjunto de pontos denotados por nós. Cada polinômio que compõe um spline é dito função base spline e um conjunto com tais funções forma uma base spline (Ruppert *et al.*, 2003). Os nós dividem o domínio da covariável em regiões e quanto maior a quantidade de nós, maior é a quantidade de splines correspondente.

Um B-spline é um spline construído a partir de funções bases B-spline. Em de Boor (1977) e Dierckx (1993) é apresentada a construção da base B-spline em detalhes assim como os algoritmos necessários.

Como destacado por Eilers & Marx (1996), um B-spline de grau $bdeg$:

- consiste de $bdeg + 1$ polinômios, cada um de grau $bdeg$;
- unem-se em $bdeg$ nós internos;
- as derivadas de ordem até $bdeg - 1$ são contínuas nos pontos de união;
- o B-spline é positivo no domínio expandido por $bdeg + 2$ nós e é 0 nos demais;
- exceto nas fronteiras, se sobrepõe com $2 \times bdeg$ polinômios de seus vizinhos;
- para cada valor da covariável utilizada na construção da base B-spline, $bdeg + 1$ B-splines são não nulos.

Os nós dividem o domínio da covariável utilizada no cálculo da base B-spline, em nd intervalos de mesmo tamanho sendo que cada intervalo será coberto por $bdeg + 1$

B-splines de grau $bdeg$. Já o número total de nós para a construção dos B-splines é $nk = nd + 2 \times bdeg + 1$, que em geral é menor do que o tamanho amostral n . Entretanto, em muitos trabalhos é considerado que a quantidade de nós adequada é dada pelo mínimo entre 40 e $1/4$ da quantidade de valores únicos da covariável utilizada para a construção da base B-spline, sendo os nós igualmente espaçados no domínio da covariável. O número de B-splines na base de regressão, isto é, o número de colunas da base B-spline é dado por $c_B = bdeg + nd$ enquanto que o número de linhas é n .

Na Figura 3.1, retirada de Eilers & Marx (1996), são apresentados alguns B-splines de ordem 1 e 2, em que é possível verificar as propriedades acima descritas. Neste exemplo, x_1, x_2, \dots, x_{10} , são os nós.

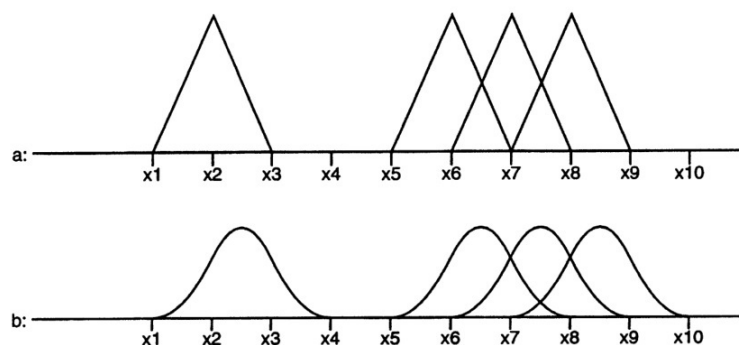


Figura 3.1: Exemplo de apenas um B-spline e de alguns B-splines sobrepostos, considerando polinômios de grau $bdeg = 1$ e $bdeg = 2$ em (a) e (b), respectivamente.

Uma característica que difere os B-splines de outros métodos de suavização é que ele não "sofre" os efeitos de fronteira em que ao estender a curva ajustada fora do domínio dos dados, esta tende para zero rapidamente. Além disso, se a curva é um polinômio, o método P-splines consegue recuperar a forma exata. É importante destacar que a técnica conserva os momentos dos dados.

3.1.2 Matriz de penalização

Em O’Sullivan (1986) e Green & Silverman (1994), uma penalização sob a derivada de segunda ordem da curva é considerada na função log-verossimilhança com o objetivo de controlar a suavidade da curva. Entretanto, no método P-splines (Eilers & Marx, 1996) a penalização é baseada na diferença finita dos coeficientes de regressão adjacentes à base B-spline. Desta forma, é uma penalização discreta e que penaliza diretamente os coeficientes ao invés de penalizar a curva, como ocorre na técnica splines de suavização, por exemplo, reduzindo assim a dimensão do problema.

A penalização \mathbf{P}_B é uma matriz de dimensão $c_B \times c_B$ definida por

$$\mathbf{P}_B = \mathbf{D}_{d_B}^\top \mathbf{D}_{d_B},$$

em que \mathbf{D}_{d_B} é a forma matricial do operador diferenças de ordem d_B do vetor de coeficientes de regressão $\mathbf{a}_B = (a_B^{(1)}, a_B^{(2)}, \dots, a_B^{(c_B)})^\top$, correspondente a base B-spline, sendo c_B o número de colunas da base B-spline. O operador diferenças é calculado recursivamente, isto é,

$$\begin{aligned} \nabla^1 a_B^{(j)} &= a_B^{(j)} - a_B^{(j-1)}, \\ \nabla^2 a_B^{(j)} &= \nabla^1(\nabla^1 a_B^{(j)}) = a_B^{(j)} - 2a_B^{(j-1)} + a_B^{(j-2)}, \\ &\vdots \\ \nabla^{d_B} a_B^{(j)} &= \nabla^1(\nabla^{d_B-1} a_B^{(j)}), \end{aligned}$$

$j = 1, 2, \dots, c_B$, e desta forma, tem-se que \mathbf{D}_{d_B} é tal que $\mathbf{D}_{d_B} \mathbf{a}_B = \nabla^{d_B} \mathbf{a}_B$. A matriz \mathbf{D}_{d_B} e conseqüentemente a matriz de penalização \mathbf{P}_B é facilmente obtida nos softwares estatísticos. Por exemplo, considerando uma base B-spline com $c_B = 5$, a matriz diferenças de ordem 1 e 2 são dadas respectivamente por,

$$\mathbf{D}_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}.$$

A ordem d_B controla a mudança entre os coeficientes adjacentes. Em geral, adota-se uma diferença de ordem 2. Esta penalização é uma boa aproximação discreta à integral da d_B -ésima derivada ao quadrado, da curva, que é a penalização geralmente adotada (Green & Silverman, 1994).

Como ocorre em outros métodos de suavização (Green & Silverman, 1994), um parâmetro que controla a suavidade da curva também é considerado no método P-splines e inserido no termo de penalização da função log-verossimilhança. Este parâmetro de suavidade penaliza os coeficientes que estão muito separados entre si. Quanto maior o valor deste parâmetro, mais se aproxima a curva ajustada à um polinômio já que neste caso os coeficientes se aproximarão a zero. Como veremos posteriormente neste capítulo, o parâmetro de suavidade não é estimado como os demais parâmetros do modelo. Em geral, seu valor é encontrado utilizando critérios tais como AIC, BIC e validação cruzada generalizada.

3.1.3 Aplicação da técnica P-splines

Considere inicialmente o modelo univariado associado à variável resposta Y_i e que uma única covariável z_1 esteja disponível para predizer Y_i , $i = 1, 2, \dots, n$, em um contexto não paramétrico tal que $\eta_{iY} = s_1(z_{i1})$ com z_{i1} sendo o valor da covariável z_1 correspondente à i -ésima observação. Utilizando a técnica P-splines, é possível reescrever o preditor como sendo

$$\boldsymbol{\eta}_{\mathbf{Y}} = s_1(\mathbf{z}_1) = \mathbf{B}_{\mathbf{Y}}\mathbf{a}_{\mathbf{Y}}, \quad (3.1)$$

em que $\boldsymbol{\eta}_{\mathbf{Y}} = (\eta_{1Y}, \eta_{2Y}, \dots, \eta_{mY})^\top$, $\mathbf{B}_{\mathbf{Y}} = \mathbf{B}_{\mathbf{Y}}(\mathbf{z}_1)$ é uma base de regressão B-spline construída a partir da covariável $\mathbf{z}_1 = (z_{11}, z_{21}, \dots, z_{n1})^\top$ e $\mathbf{a}_{\mathbf{Y}}$ é um vetor de coeficientes da regressão. A base B-spline $\mathbf{B}_{\mathbf{Y}}$ tem dimensão $n \times c_Y$ e é formada por c_Y B-splines. Sendo y_1, y_2, \dots, y_n os valores observados de Y_1, Y_2, \dots, Y_n , respectivamente, o vetor de coeficientes $\mathbf{a}_{\mathbf{Y}}$ é estimado através da maximização da função log-verossimilhança penalizada, isto é,

$$\ell^*(\mathbf{a}_{\mathbf{Y}}, \phi \mid \mathbf{y}, \mathbf{Z}) = \ell(\mathbf{a}_{\mathbf{Y}}, \phi \mid \mathbf{y}, \mathbf{Z}) - \frac{1}{2}\lambda_Y \mathbf{a}_{\mathbf{Y}}^\top \mathbf{P}_{\mathbf{Y}} \mathbf{a}_{\mathbf{Y}}, \quad (3.2)$$

em que $\ell(\cdot)$ é a função log-verossimilhança associada ao modelo da variável resposta Y_i , $i = 1, 2, \dots, n$, $\mathbf{P}_{\mathbf{Y}}$ é a matriz de penalização de dimensão $c_Y \times c_Y$ e de ordem d_Y e $\lambda_Y \geq 0$ é o parâmetro de suavidade da curva. O vetor de médias $\boldsymbol{\mu}_{\mathbf{Y}}$ é dado por $\boldsymbol{\mu}_{\mathbf{Y}} = g_1^{-1}(\mathbf{B}_{\mathbf{Y}}\mathbf{a}_{\mathbf{Y}})$ em que $g_1(\cdot)$ é a função de ligação. O parâmetros ϕ , caso seja desconhecido, pode ser estimado através da maximização direta da função log-verossimilhança penalizada. Note porém, que tal parâmetro não é penalizado. O valor do parâmetro λ_Y é calculado, utilizando validação cruzada generalizada, o qual seleciona o valor de λ_Y dentre uma malha de valores, que minimiza

$$GCV(\lambda_Y) = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - \text{traço}(\mathbf{L}_{\mathbf{Y}})}. \quad (3.3)$$

A matriz $\mathbf{L}_{\mathbf{Y}}$ é obtida através da versão penalizada do algoritmo score, utilizado para resolver as equações de estimação penalizadas associadas à equação (3.2). De fato, seja $\hat{\mathbf{a}}_{\mathbf{Y}}^{(N+1)} = (\mathbf{B}_{\mathbf{Y}}^\top \widehat{\mathbf{W}} \mathbf{B}_{\mathbf{Y}} + \lambda_Y \mathbf{P}_{\mathbf{Y}})^{-1} \mathbf{B}_{\mathbf{Y}}^\top \widehat{\mathbf{W}} \mathbf{y}^*$, a estimativa de $\mathbf{a}_{\mathbf{Y}}$ no passo $(N + 1)$, em que $\widehat{\mathbf{W}}$ é a matriz de pesos associada à Y_i , definida na Subseção 2.1.1 e calculada em $\hat{\mathbf{a}}_{\mathbf{Y}}^{(N)}$ e $\mathbf{Z}_1 = \mathbf{B}_{\mathbf{Y}}$, e \mathbf{y}^* é uma variável modificada dada por $\mathbf{y}^* = \mathbf{B}_{\mathbf{Y}} \hat{\mathbf{a}}_{\mathbf{Y}}^{(N)} + \widehat{\mathbf{W}}^{-1/2} \widehat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}}_{\mathbf{Y}})$ e que também depende de $\hat{\mathbf{a}}_{\mathbf{Y}}^{(N)}$, com $\widehat{\mathbf{V}}$ dado como na Subseção 2.1.1. Assim, fazendo uma analogia com a teoria dos modelos lineares em que se pode obter uma matriz $\mathbf{L}_{\mathbf{Y}}$ de projeção tal que $\hat{\mathbf{y}} = \mathbf{L}_{\mathbf{Y}} \mathbf{y}$, adota-se que $\mathbf{L}_{\mathbf{Y}} = \mathbf{B}_{\mathbf{Y}} (\mathbf{B}_{\mathbf{Y}}^\top \widehat{\mathbf{W}} \mathbf{B}_{\mathbf{Y}} + \lambda_Y \mathbf{P}_{\mathbf{Y}})^{-1} \mathbf{B}_{\mathbf{Y}}^\top \widehat{\mathbf{W}}$, calculada após a convergência. Neste caso,

$\mathbf{y} = \mathbf{y}^*$, com \mathbf{y}^* avaliada após a convergência. Note porém, que \mathbf{L}_Y não é uma matriz de projeção já que não é idempotente. Além disso, para um determinado valor de λ_Y fixo, o suavizador é linear.

Quando $\lambda_Y = 0$, o processo de estimação pode ser realizado no contexto dos modelos lineares generalizados em que a base \mathbf{B}_Y é a matriz de planejamento. Desta forma, a matriz de covariâncias assintótica dos estimadores de máxima verossimilhança é dada pelo inverso da matriz de informação esperada (ou observada), desde que esta matriz seja definida positiva. Note porém, que λ_Y pode ser maior do que 0. Conseqüentemente, a matriz de covariâncias assintótica pode subestimar as variâncias, dependendo do valor de λ_Y .

Em modelos não paramétricos em geral, o interesse está na curva ajustada. Desta forma, definir a matriz de covariâncias da curva ajustada $\mathbf{B}_Y \widehat{\mathbf{a}}_Y$ é de grande interesse, principalmente para a construção das bandas de confiança. Fazendo uma analogia à teoria dos modelos lineares, tem-se que

$$\begin{aligned} \text{Cov}(\mathbf{B}_Y \widehat{\mathbf{a}}_Y) &= \text{Cov}(\mathbf{L}_Y \mathbf{y}^*) \\ &\approx a(\phi) \mathbf{B}_Y (\mathbf{B}_Y^\top \widehat{\mathbf{W}} \mathbf{B}_Y + \lambda_Y \mathbf{P}_Y)^{-1} \mathbf{B}_Y^\top \widehat{\mathbf{W}} \mathbf{B}_Y (\mathbf{B}_Y^\top \widehat{\mathbf{W}} \mathbf{B}_Y + \lambda_Y \mathbf{P}_Y)^{-1} \mathbf{B}_Y^\top. \end{aligned} \quad (3.4)$$

Para o parâmetro ϕ , pode-se utilizar a variância assintótica habitual, quando necessária.

Analogamente, considerando apenas o modelo associado à variável resposta contínua X_i , em que a função densidade de $X_i \mid Y_i = y_i$ é dada pela equação (2.3), $i = 1, 2, \dots, n$, e supondo que as médias marginais $\boldsymbol{\mu}_Y = (\mu_{1Y}, \mu_{2Y}, \dots, \mu_{nY})^\top$ sejam conhecidas e que uma única covariável \mathbf{t}_1 esteja disponível para predizer X_i , $i = 1, 2, \dots, n$, em um contexto não paramétrico, segue aplicando a técnica P-splines que

$$\boldsymbol{\eta}_X = u_1(\mathbf{t}_1) = \mathbf{B}_X \mathbf{a}_X, \quad (3.5)$$

em que $\boldsymbol{\eta}_X = (\eta_{1X}, \eta_{2X}, \dots, \eta_{nX})^\top$, $\mathbf{B}_X = \mathbf{B}_X(\mathbf{t}_1)$ é uma base de regressão B-spline, de dimensão $n \times c_X$, construída a partir da covariável $\mathbf{t}_1 = (t_{11}, t_{21}, \dots, t_{n1})^\top$ e \mathbf{a}_X é um vetor de coeficientes da regressão. Dada uma amostra de valores observa-

dos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ dos pares $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, respectivamente, com (X_i, Y_i) independente de (X_k, Y_k) , $i, k = 1, 2, \dots, n$, o vetor de coeficientes \mathbf{a}_X é estimado através da maximização direta da função log-verossimilhança penalizada, associada ao modelo de $X_i | Y_i = y_i$, isto é,

$$\ell^*(\mathbf{a}_X, \gamma, \varphi | \mathbf{x}, \mathbf{y}, \boldsymbol{\mu}_Y, \mathbf{T}) = \ell(\mathbf{a}_X, \gamma, \varphi | \mathbf{x}, \mathbf{y}, \boldsymbol{\mu}_Y, \mathbf{T}) - \frac{1}{2} \lambda_X \mathbf{a}_X^\top \mathbf{P}_X \mathbf{a}_X, \quad (3.6)$$

em que $\ell(\cdot)$ é a função log-verossimilhança associada ao modelo da variável resposta X_i , \mathbf{P}_X é a matriz de penalização dos coeficientes da base B-spline, de dimensão $c_X \times c_X$ e $\lambda_X \geq 0$ é o parâmetro de suavidade. O vetor de médias $\boldsymbol{\mu}_X$ é dado por $\boldsymbol{\mu}_X = g_2^{-1}(\mathbf{B}_X \mathbf{a}_X)$ em que $g_2(\cdot)$ é a função de ligação. Os parâmetros γ e φ também são estimados através da maximização direta da função log-verossimilhança penalizada dada na equação (3.6), sendo que ambos os parâmetros não são penalizados. De maneira análoga à λ_Y , o parâmetro de suavidade λ_X também é calculado através da validação cruzada generalizada, dada por

$$GCV(\lambda_X) = \sum_{i=1}^n \frac{(x_i - \hat{x}_i)^2}{n - \text{traço}(\mathbf{L}_X)}, \quad (3.7)$$

em que $\mathbf{L}_X = \mathbf{B}_X (\mathbf{B}_X^\top \hat{\mathbf{R}} \mathbf{B}_X + \lambda_X \mathbf{P}_X)^{-1} \mathbf{B}_X^\top \hat{\mathbf{R}}$, com $\hat{\mathbf{R}}$ sendo a matriz de pesos associada à X_i , definida na Subseção 2.1.1 e calculada em $\hat{\mathbf{a}}_X^{(N)}$ e $\mathbf{T}_1 = \mathbf{B}_X$.

Como na equação (3.4), define-se a matriz de covariâncias da curva ajustada $\mathbf{B}_X \hat{\mathbf{a}}_X$ por

$$\begin{aligned} \text{Cov}(\mathbf{B}_X \hat{\mathbf{a}}_X) &= \text{Cov}(\mathbf{L}_X \mathbf{x}^*) \\ &\approx a^*(\varphi) \mathbf{B}_X (\mathbf{B}_X^\top \hat{\mathbf{R}} \mathbf{B}_X + \lambda_X \mathbf{P}_X)^{-1} \mathbf{B}_X^\top \hat{\mathbf{R}} \mathbf{B}_X (\mathbf{B}_X^\top \hat{\mathbf{R}} \mathbf{B}_X + \lambda_X \mathbf{P}_X)^{-1} \mathbf{B}_X^\top. \end{aligned} \quad (3.8)$$

Para os parâmetros γ e φ , pode-se utilizar a variância assintótica habitual, quando necessária.

Entretanto, o principal objetivo deste capítulo é apresentar um método eficiente para o ajuste do modelo bivariado geral adotado, considerando preditores como os apresentados nas equações (2.21) e (2.22), ou seja, para os casos em que mais de uma

covariável esteja disponível para prever as variáveis respostas discretas e contínuas. Para este objetivo, Marx & Eilers (1998) estendem a aplicação da técnica P-splines para o contexto dos modelos aditivos generalizados, denotado-a por *P-GAMs*.

3.2 Modelos aditivos generalizados com P-splines

O uso da técnica P-splines no contexto dos modelos aditivos generalizados (*P-GAMs*) é de grande importância na regressão não paramétrica. Diferentemente dos métodos *escore local* e *retroajuste*, a técnica é capaz de estimar simultaneamente todos os termos suaves, evitando assim, os ciclos *retroajuste* muitas vezes indesejáveis. Nos *P-GAMs*, o problema de estimação não paramétrica é reduzido a um problema de estimação no contexto dos modelos lineares generalizados, com uma penalização. Desta forma, resultados de diagnósticos e o cálculo de erros padrão e outras medidas tratadas em MLG podem ser aplicadas. Outra característica importante da técnica é que o sistema de equações é relativamente baixo e conseqüentemente, o custo computacional é baixo. Para mais detalhes, Marx & Eilers (1998).

Suponha novamente o modelo marginal associado à Y_i e preditor dado pela equação (2.21), $i = 1, 2, \dots, n$. Cada uma das funções $s_{j_Y}(\mathbf{z}_{j_Y})$ pode ser escrita como na equação (3.1) em que denota-se por $\mathbf{B}_{j_Y \mathbf{Y}}$ a base B-spline construída a partir da covariável \mathbf{z}_{j_Y} e $\mathbf{a}_{j_Y \mathbf{Y}} = (a_{j_Y \mathbf{Y}}^{(1)}, a_{j_Y \mathbf{Y}}^{(2)}, \dots, a_{j_Y \mathbf{Y}}^{(c_{j_Y \mathbf{Y}})})^\top$ o vetor de coeficientes de tamanho $c_{j_Y \mathbf{Y}}$ associado à $\mathbf{B}_{j_Y \mathbf{Y}}$, $j_Y = 1, \dots, p$. Assim, reescreve-se o vetor $\boldsymbol{\eta}_{\mathbf{Y}} = (\eta_{1\mathbf{Y}}, \eta_{2\mathbf{Y}}, \dots, \eta_{n\mathbf{Y}})^\top$ como sendo

$$\begin{aligned} \boldsymbol{\eta}_{\mathbf{Y}} = \mathbf{1}_n \beta_0 + \sum_{j_Y=1}^p s_{j_Y}(\mathbf{z}_{j_Y}) &= \mathbf{1}_n \beta_0 + \mathbf{B}_{1\mathbf{Y}} \mathbf{a}_{1\mathbf{Y}} + \mathbf{B}_{2\mathbf{Y}} \mathbf{a}_{2\mathbf{Y}} + \dots + \mathbf{B}_{p\mathbf{Y}} \mathbf{a}_{p\mathbf{Y}} \\ &= \mathbf{B}_{\mathbf{Y}} \mathbf{a}_{\mathbf{Y}}, \end{aligned} \quad (3.9)$$

em que $\mathbf{B}_{\mathbf{Y}} = (\mathbf{1}_n : \mathbf{B}_{1\mathbf{Y}} : \mathbf{B}_{2\mathbf{Y}} : \dots : \mathbf{B}_{p\mathbf{Y}})$ é a base de regressão B-spline, de dimensão $n \times (1 + \sum_{j_Y=1}^p c_{j_Y \mathbf{Y}})$, com $\mathbf{1}_n$ sendo um vetor coluna de uns de tamanho n e

$\mathbf{a}_Y = (\beta_0, \mathbf{a}_{1Y}^\top, \mathbf{a}_{2Y}^\top, \dots, \mathbf{a}_{pY}^\top)^\top$ é o vetor dos coeficientes de regressão. O vetor \mathbf{a}_Y tem $(1 + \sum_{j_Y=1}^p c_{j_Y Y})$ elementos. Como pode ser observado na equação (3.9), o preditor é reduzido ao caso linear. Sendo $\mathbf{P}_{j_Y Y}$ a matriz de penalização de dimensão $c_{j_Y Y} \times c_{j_Y Y}$ e ordem $d_{j_Y Y}$, associada ao j_Y -ésimo vetor de coeficientes $\mathbf{a}_{j_Y Y}$, e $\lambda_{j_Y Y} \geq 0$ o respectivo parâmetro de suavidade, $j_Y = 1, \dots, p$, segue que a matriz de penalização associada à \mathbf{a}_Y é uma matriz bloco diagonal dada por

$$\mathbf{P}_Y = \text{bloco diag}(0, \lambda_{1Y} \mathbf{P}_{1Y}, \lambda_{2Y} \mathbf{P}_{2Y}, \dots, \lambda_{pY} \mathbf{P}_{pY}). \quad (3.10)$$

O valor 0 em \mathbf{P}_Y corresponde ao termo do intercepto, que não é penalizado.

Analogamente, para o modelo associado à resposta contínua X_i com preditor dado pela equação (2.22), $i = 1, 2, \dots, n$, pode-se aplicar a técnica de Marx & Eilers (1998) e reescrever o vetor $\boldsymbol{\eta}_X = (\eta_{1X}, \eta_{2X}, \dots, \eta_{nX})^\top$ como sendo

$$\begin{aligned} \boldsymbol{\eta}_X &= \mathbf{1}_n \delta_0 + \sum_{j_X=1}^q u_{j_X}(\mathbf{t}_{j_X}) = \mathbf{1}_n \delta_0 + \mathbf{B}_{1X} \mathbf{a}_{1X} + \mathbf{B}_{2X} \mathbf{a}_{2X} + \dots + \mathbf{B}_{qX} \mathbf{a}_{qX} \\ &= \mathbf{B}_X \mathbf{a}_X, \end{aligned} \quad (3.11)$$

em que $\mathbf{B}_{j_X X}$ é a base B-spline associada à \mathbf{t}_{j_X} e $\mathbf{a}_{j_X X} = (a_{j_X X}^{(1)}, a_{j_X X}^{(2)}, \dots, a_{j_X X}^{(c_{j_X X})})^\top$ é o respectivo vetor de coeficientes, de tamanho $c_{j_X X}$, $j_X = 1, \dots, q$. A matriz composta $\mathbf{B}_X = (\mathbf{1}_n : \mathbf{B}_{1X} : \mathbf{B}_{2X} : \dots : \mathbf{B}_{qX})$ é a base de regressão B-spline, de dimensão $n \times (1 + \sum_{j_X=1}^q c_{j_X X})$, com $\mathbf{1}_n$ sendo um vetor coluna de uns de tamanho n e $\mathbf{a}_X = (\delta_0, \mathbf{a}_{1X}^\top, \mathbf{a}_{2X}^\top, \dots, \mathbf{a}_{qX}^\top)^\top$ é o vetor de coeficientes de regressão, de tamanho $1 + \sum_{j_X=1}^q c_{j_X X}$. Novamente, o preditor é reduzido ao caso linear. Neste caso, a matriz de penalização associada à \mathbf{a}_X é uma matriz bloco diagonal dada por

$$\mathbf{P}_X = \text{bloco diag}(0, \lambda_{1X} \mathbf{P}_{1X}, \lambda_{2X} \mathbf{P}_{2X}, \dots, \lambda_{qX} \mathbf{P}_{qX}), \quad (3.12)$$

em que $\mathbf{P}_{j_X X}$ é a matriz de penalização de dimensão $c_{j_X X} \times c_{j_X X}$ e ordem $d_{j_X X}$, associada à $\mathbf{a}_{j_X X}$ e $\lambda_{j_X X} \geq 0$ é o respectivo parâmetro de suavidade, $j_X = 1, \dots, p$. O primeiro elemento de \mathbf{P}_X é 0 e corresponde ao intercepto, não penalizado.

As estimativas dos vetores de coeficientes de regressão \mathbf{a}_Y e \mathbf{a}_X são obtidas maximizando as funções log-verossimilhança penalizadas dadas pelas equações (3.2) e (3.6),

respectivamente, considerando as novas composições de \mathbf{a}_Y , \mathbf{B}_Y , \mathbf{P}_Y , \mathbf{a}_X , \mathbf{B}_X e \mathbf{P}_X e as relações $\boldsymbol{\mu}_Y = g_1^{-1}(\mathbf{B}_Y \mathbf{a}_Y)$ e $\boldsymbol{\mu}_X = g_2^{-1}(\mathbf{B}_X \mathbf{a}_X)$, notando que em \mathbf{P}_Y e \mathbf{P}_X , cada submatriz de penalização $\mathbf{P}_{j_Y Y}$ e $\mathbf{P}_{j_X X}$ já é multiplicada por um parâmetro de suavidade $\lambda_{j_Y Y}$ e $\lambda_{j_X X}$, respectivamente, $j_Y = 1, 2, \dots, p$, e $j_X = 1, 2, \dots, q$, não sendo assim necessário considerá-los novamente em (3.2) e em (3.6). Entretanto, \mathbf{B}_Y e \mathbf{B}_X não são de posto completo (Marx & Eilers, 1998). Conseqüentemente, $\mathbf{B}_Y^\top \mathbf{B}_Y$ e $\mathbf{B}_X^\top \mathbf{B}_X$ necessárias no processo de estimação, são singulares, comprometendo a resolução dos sistemas de equações resultantes.

Para resolver este problema, Marx & Eilers (1998) adotam uma pequena "ridge penalty" nos sistemas de equações. Outra solução é transformar as bases B-splines \mathbf{B}_Y e \mathbf{B}_X originais em novas bases não singulares. Isto é possível reescrevendo os modelos P-splines associados às variáveis respostas Y_i e X_i , $i = 1, 2, \dots, n$, na configuração dos modelos mistos lineares (generalizados), em que um efeito aleatório é incorporado. Entende-se por modelos P-splines, os modelos não paramétricos ou semiparamétricos em que uma base B-spline é utilizada na regressão e uma penalização discreta, que penaliza os coeficientes da regressão, é adicionada a função de log-verossimilhança. Em alguns casos, a inversa generalizada poderia resolver o problema da singularidade.

3.2.1 Transformação da base

O preditor de um modelo P-splines pode ser decomposto em uma soma de duas componentes: uma fixa polinomial cujo grau está associado à ordem da matriz de penalização e uma aleatória formada por um termo suave não linear. Nesta nova composição, a parte fixa não depende do parâmetro de suavização e não é penalizada enquanto que a parte aleatória é suavizada utilizando um suavizador P-splines. Desta forma, o modelo P-splines original pode ser visto como um modelo misto linear (generalizado). Em Currie & Durban (2002), Durbán & Currie (2003), Durbán *et al.* (2006)

e Currie *et al.* (2006), os autores tratam desta abordagem.

Considere o modelo marginal P-splines associado à variável resposta Y_i , $i = 1, 2, \dots, n$, com preditor nas mesmas condições apresentadas na equação (3.1). Neste caso, o preditor $\boldsymbol{\eta}_Y = \mathbf{B}_Y \mathbf{a}_Y$ pode ser decomposto em

$$\boldsymbol{\eta}_Y = \mathbf{B}_Y \mathbf{a}_Y = \mathbf{G}_Y \mathbf{e}_Y + \mathbf{H}_Y \mathbf{l}_Y,$$

em que $\mathbf{G}_Y \mathbf{e}_Y$ é a parte fixa polinomial de grau $d_Y - 1$, com d_Y sendo a ordem da matriz de penalização \mathbf{P}_Y associada ao vetor \mathbf{a}_Y , e $\mathbf{H}_Y \mathbf{l}_Y$ é a parte aleatória formada por um termo suave não linear.

Para calcular as matrizes \mathbf{G}_Y e \mathbf{H}_Y , é necessário definir uma matriz ortogonal \mathbf{E}_Y de transformação da base B-spline \mathbf{B}_Y na nova base $[\mathbf{G}_Y : \mathbf{H}_Y]$, isto é, uma matriz \mathbf{E}_Y tal que $\mathbf{B}_Y \mathbf{E}_Y = [\mathbf{G}_Y : \mathbf{H}_Y]$. Isto é realizado utilizando-se a decomposição em valores singulares (SVD) da matriz de penalização \mathbf{P}_Y (Good, 1969). Seja

$$\mathbf{P}_Y = \mathbf{D}_Y^\top \mathbf{D}_Y = \mathbf{K}_Y \boldsymbol{\Gamma} \mathbf{K}_Y^\top,$$

a SVD da matriz simétrica $\mathbf{D}_Y^\top \mathbf{D}_Y$, em que $\mathbf{K}_Y = [\mathbf{K}_{nY} : \mathbf{K}_{sY}]$ é uma matriz ortogonal de autovetores da SVD de $\mathbf{D}_Y^\top \mathbf{D}_Y$ com \mathbf{K}_{nY} e \mathbf{K}_{sY} correspondendo as partes nulas e não nulas da decomposição, respectivamente, e $\boldsymbol{\Gamma}$ é uma matriz diagonal cujos elementos são os autovalores da SVD de $\mathbf{D}_Y^\top \mathbf{D}_Y$. Pode-se representar $\boldsymbol{\Gamma}$ por

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{0}_{d_Y} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_Y \end{bmatrix}, \quad (3.13)$$

em que $\mathbf{0}_{d_Y}$ é uma matriz quadrada de zeros, de dimensão $d_Y \times d_Y$, indicando que d_Y autovalores são nulos, e $\boldsymbol{\Sigma}_Y$ é uma matriz diagonal, de dimensão $(c_Y - d_Y) \times (c_Y - d_Y)$, cujos elementos são os autovalores positivos. Note que d_Y é a ordem da penalização de \mathbf{P}_Y enquanto c_Y é o número de colunas de \mathbf{B}_Y .

Adotando $\mathbf{E}_Y = \mathbf{K}_Y$, tem-se que os componentes da nova base $[\mathbf{G}_Y : \mathbf{H}_Y]$ são

dados por

$$\mathbf{G}_Y = \mathbf{B}_Y \mathbf{K}_{nY} \text{ e } \mathbf{H}_Y = \mathbf{B}_Y \mathbf{K}_{sY}, \quad (3.14)$$

enquanto que os vetores de coeficientes \mathbf{e}_Y e \mathbf{l}_Y em função de \mathbf{a}_Y são dados por

$$\mathbf{e}_Y = \mathbf{K}_{nY}^\top \mathbf{a}_Y \text{ e } \mathbf{l}_Y = \mathbf{K}_{sY}^\top \mathbf{a}_Y.$$

Nesta nova configuração, uma nova matriz de penalização associada aos novos vetores de coeficientes é definida por

$$\mathbf{P}_Y^* = \mathbf{K}_Y^\top \mathbf{P}_Y \mathbf{K}_Y = \mathbf{K}_Y^\top \mathbf{D}_Y^\top \mathbf{D}_Y \mathbf{K}_Y = \mathbf{\Gamma},$$

com $\mathbf{\Gamma}$ dada pela equação (3.13). Note que a parte fixa do modelo não é penalizada enquanto que a parte aleatória é penalizada. Logo, a matriz \mathbf{G}_Y dada na equação (3.14) pode ser substituída por outra matriz tal que a nova matriz composta $[\mathbf{G}_Y : \mathbf{H}_Y]$ seja de posto completo.

O número de colunas de \mathbf{G}_Y é igual a ordem de penalização d_Y da matriz \mathbf{P}_Y já que o número de autovalores zeros e consequentemente a dimensão da parte fixa está diretamente relacionada ao valor de d_Y . Logo, uma escolha natural para \mathbf{G}_Y é

$$\mathbf{G}_Y = [\mathbf{1}_n : \mathbf{z}_1 : \mathbf{z}_1^2 : \cdots : \mathbf{z}_1^{d_Y-1}] = \begin{bmatrix} 1 & z_{11} & z_{11}^2 & z_{11}^3 & \cdots & z_{11}^{d_Y-1} \\ 1 & z_{21} & z_{21}^2 & z_{21}^3 & \cdots & z_{21}^{d_Y-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n1}^2 & z_{n1}^3 & \cdots & z_{n1}^{d_Y-1} \end{bmatrix},$$

de dimensão $n \times d_Y$. A matriz \mathbf{G}_Y neste caso é de posto coluna completo. Em geral a ordem da penalização adotada é $d_Y = 2$.

O mesmo procedimento pode ser aplicado quando p covariáveis estão disponíveis para prever Y_i , $i = 1, 2, \dots, n$, com preditor nas condições da equação (3.9). Neste caso, o preditor $\boldsymbol{\eta}_Y$ pode ser decomposto em

$$\boldsymbol{\eta}_Y = \mathbf{B}_Y \mathbf{a}_Y = \mathbf{G}_Y \mathbf{e}_Y + \mathbf{H}_Y \mathbf{l}_Y,$$

em que \mathbf{G}_Y e \mathbf{H}_Y são matrizes associadas as partes fixas e aleatórias da decomposição,

respectivamente, e \mathbf{e}_Y e \mathbf{l}_Y são os respectivos vetores de coeficientes, ambos desconhecidos. A matriz de transformação da base B-spline \mathbf{B}_Y na nova base $[\mathbf{G}_Y : \mathbf{H}_Y]$, obtida da decomposição em valores singulares da matriz de penalização \mathbf{P}_Y definida na equação (3.10), é dada por

$$\begin{aligned} \mathbf{E}_Y^* &= [\mathbf{E}_{nY}^* : \mathbf{E}_{sY}^*] \\ &= \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{1nY} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{K}_{1sY} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{2nY} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{K}_{2sY} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_{qnY} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_{qsY} \end{bmatrix}, \end{aligned} \quad (3.15)$$

em que as submatrizes $\mathbf{K}_{j_Y n_Y}$ e $\mathbf{K}_{j_Y s_Y}$ são os autovetores correspondentes aos autovalores zeros e não zeros, respectivamente, da SVD da submatriz de penalização $\mathbf{P}_{j_Y Y}$, $j_Y = 1, 2, \dots, p$, de \mathbf{P}_Y . A matriz \mathbf{E}_Y^* tem dimensão $(1 + \sum_{j_Y=1}^p c_{j_Y Y}) \times (1 + \sum_{j_Y=1}^p c_{j_Y Y})$, em que $c_{j_Y Y}$ é o número de colunas de $\mathbf{B}_{j_Y Y}$, $j_Y = 1, 2, \dots, p$. Segue que

$$\begin{aligned} \mathbf{G}_Y &= \mathbf{B}_Y \mathbf{E}_{nY}^* = [\mathbf{1}_n : \mathbf{B}_{1Y} : \dots : \mathbf{B}_{pY}] \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{1nY} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{2nY} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_{pnY} \end{bmatrix} \\ &= [\mathbf{1}_n : \mathbf{B}_{1Y} \mathbf{K}_{1nY} : \dots : \mathbf{B}_{pY} \mathbf{K}_{pnY}] = [\mathbf{1}_n : \mathbf{G}_{1Y} : \dots : \mathbf{G}_{pY}], \end{aligned} \quad (3.16)$$

em que $\mathbf{G}_{j_Y Y}$ corresponde a parte fixa da decomposição da base B-spline $\mathbf{B}_{j_Y Y}$, $j_Y = 1, 2, \dots, p$. Como não são penalizadas, $\mathbf{G}_{j_Y Y}$ pode ser substituída por outra matriz que seja de posto completo. Uma escolha natural seria

$$\mathbf{G}_{j_Y Y} = [\mathbf{1}_n : \mathbf{z}_{j_Y} : \mathbf{z}_{j_Y}^2 : \cdots : \mathbf{z}_{j_Y}^{d_{j_Y Y}-1}] = \begin{bmatrix} 1 & z_{1j_Y} & z_{1j_Y}^2 & z_{1j_Y}^3 & \cdots & z_{1j_Y}^{d_{j_Y Y}-1} \\ 1 & z_{2j_Y} & z_{2j_Y}^2 & z_{2j_Y}^3 & \cdots & z_{2j_Y}^{d_{j_Y Y}-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{nj_Y} & z_{nj_Y}^2 & z_{nj_Y}^3 & \cdots & z_{nj_Y}^{d_{j_Y Y}-1} \end{bmatrix},$$

de dimensão $n \times d_{j_Y Y}$, em que $\mathbf{z}_{j_Y} = (z_{1j_Y}, z_{2j_Y}, \dots, z_{nj_Y})^\top$ e $d_{j_Y Y}$ é a ordem de penalização de $\mathbf{P}_{j_Y Y}$, $j_Y = 1, 2, \dots, p$. Para uma ordem de penalização $d_{j_Y Y} = 2$, $\forall j_Y = 1, 2, \dots, p$, tem-se que $\mathbf{G}_{j_Y Y} = [\mathbf{1}_n : \mathbf{z}_{j_Y}]$. Logo, da equação (3.16) segue que

$$\mathbf{G}_Y = [\mathbf{1}_n : \mathbf{1}_n : \mathbf{z}_1 : \mathbf{1}_n : \mathbf{z}_2 : \cdots : \mathbf{1}_n : \mathbf{z}_p] = \begin{bmatrix} 1 & 1 & z_{11} & 1 & z_{12} & \cdots & 1 & z_{1p} \\ 1 & 1 & z_{21} & 1 & z_{22} & \cdots & 1 & z_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & z_{n1} & 1 & z_{n2} & \cdots & 1 & z_{np} \end{bmatrix}.$$

Entretanto, \mathbf{G}_Y não é de posto completo. Logo, removem-se os vetores colunas de uns repetidos, deixando apenas o primeiro, referente ao termo do intercepto. Segue que

$$\mathbf{G}_Y = [\mathbf{1}_n : \mathbf{z}_1 : \mathbf{z}_2 : \cdots : \mathbf{z}_p] = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1p} \\ 1 & z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix},$$

de posto completo.

A matriz de transformação, desconsiderando os vetores colunas associados aos vetores de uns removidos, é dada por

$$\begin{aligned} \mathbf{E}_Y &= [\mathbf{E}_{nY} : \mathbf{E}_{sY}] \\ &= \begin{bmatrix} 1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{1nY}^{(-1)} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{K}_{1sY} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_{2nY}^{(-1)} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{K}_{2sY} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_{pnY}^{(-1)} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_{psY} \end{bmatrix}, \end{aligned}$$

em que $\mathbf{K}_{j_Y n Y}^{(-1)}$ é a submatriz $\mathbf{K}_{j_Y n Y}$ de \mathbf{E}_Y^* , dada na equação (3.15), sem a primeira coluna e $\mathbf{K}_{j_Y s Y}$ é como definida na equação (3.15), $j_Y = 1, 2, \dots, p$. A matriz \mathbf{E}_Y tem dimensão $(1 + \sum_{j_Y=1}^p c_{j_Y Y}) \times (1 + \sum_{j_Y=1}^p c_{j_Y Y} - p)$.

A matriz \mathbf{H}_Y é dada por

$$\begin{aligned} \mathbf{H}_Y &= \mathbf{B}_Y \mathbf{E}_{sY} = [\mathbf{1}_n : \mathbf{B}_{1Y} : \dots : \mathbf{B}_{pY}] \begin{bmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{K}_{1sY} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{2sY} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_{psY} \end{bmatrix} \\ &= [\mathbf{B}_{1Y} \mathbf{K}_{1sY} : \dots : \mathbf{B}_{pY} \mathbf{K}_{psY}] = [\mathbf{H}_{1Y} : \dots : \mathbf{H}_{pY}], \end{aligned}$$

em que $\mathbf{H}_{1Y}, \mathbf{H}_{2Y}, \dots, \mathbf{H}_{pY}$, correspondem a parte aleatória da decomposição das bases B-splines $\mathbf{B}_{1Y}, \mathbf{B}_{2Y}, \dots, \mathbf{B}_{pY}$, respectivamente. Os vetores \mathbf{e}_Y e \mathbf{l}_Y , em função de \mathbf{a}_Y do modelo P-splines inicial, são dados por

$$\mathbf{e}_Y = \mathbf{E}_{nY}^\top \mathbf{a}_Y \quad \text{e} \quad \mathbf{l}_Y = \mathbf{E}_{sY}^\top \mathbf{a}_Y.$$

A nova matriz de penalização, associada à nova base não singular, é dada por

$$\mathbf{P}_Y^* = \mathbf{E}_Y^\top \mathbf{P}_Y \mathbf{E}_Y = \text{blockdiag}(\mathbf{0}_Y, \lambda_{1Y} \boldsymbol{\Sigma}_{1Y}, \lambda_{2Y} \boldsymbol{\Sigma}_{2Y}, \dots, \lambda_{pY} \boldsymbol{\Sigma}_{pY}),$$

em que $\mathbf{0}_Y$ é uma matriz quadrada de zeros, de dimensão $(1 + \sum_{j_Y=1}^p d_{j_Y Y} - p) \times (1 + \sum_{j_Y=1}^p d_{j_Y Y} - p)$, $\boldsymbol{\Sigma}_{j_Y Y}$ é uma matriz diagonal, de dimensão $(c_{j_Y Y} - d_{j_Y Y}) \times (c_{j_Y Y} - d_{j_Y Y})$ cujos elementos são os autovalores positivos da SVD da matriz de penalização $\mathbf{P}_{j_Y Y}$ e $\lambda_{j_Y Y}$ é o parâmetro de suavidade associado, com $d_{j_Y Y}$ sendo a ordem da penalização

da submatriz $\mathbf{P}_{j_Y Y}$ e $c_{j_Y Y}$ sendo o número de colunas de $\mathbf{B}_{j_Y Y}$, $j_Y = 1, 2, \dots, p$. A parte fixa correspondente a matriz \mathbf{G}_Y não é penalizada enquanto que a parte aleatória correspondente a matriz \mathbf{H}_Y é penalizada. Denotaremos a nova base não singular por $\mathbf{B}_Y^* = [\mathbf{G}_Y : \mathbf{H}_Y]$ e o novo vetor de coeficientes de regressão por $\mathbf{a}_Y^* = (\mathbf{e}_Y^\top, \mathbf{l}_Y^\top)^\top$.

De maneira análoga, é possível reescrever o modelo P-splines associado à variável resposta X_i , $i = 1, 2, \dots, n$, nas configurações dos modelos mistos e assim, encontrar uma nova base não singular o qual denotaremos por \mathbf{B}_X^* . A correspondente nova matriz de penalização e novo vetor de coeficientes de regressão serão denotados \mathbf{P}_X^* e \mathbf{a}_X^* , respectivamente.

Uma vez transformados os modelos P-splines originais, associados às variáveis respostas Y_i e X_i , em modelos nas características dos modelos mistos generalizados, seria possível estimar \mathbf{a}_Y^* e \mathbf{a}_X^* utilizando técnicas de estimação para modelos mistos lineares generalizados, por exemplo, o método de quase-verossimilhança penalizada (Breslow & Clayton, 1993). Entretanto, isto não é necessário. Pode-se estimar \mathbf{a}_Y^* e \mathbf{a}_X^* maximizando individualmente as funções log-verossimilhança penalizadas dadas pelas equações (3.2) e (3.6), respectivamente, considerando as novas bases não singulares \mathbf{B}_Y^* e \mathbf{B}_X^* e as correspondentes novas matrizes de penalização \mathbf{P}_Y^* e \mathbf{P}_X^* . Os vetores $\boldsymbol{\mu}_Y$ e $\boldsymbol{\mu}_X$ são dados pelas relações $\boldsymbol{\mu}_Y = g_1^{-1}(\mathbf{B}_Y^* \mathbf{a}_Y^*)$ e $\boldsymbol{\mu}_X = g_2^{-1}(\mathbf{B}_X^* \mathbf{a}_X^*)$, respectivamente.

Por construção, \mathbf{P}_Y^* e \mathbf{P}_X^* já contêm os parâmetros de suavidade $\lambda_{j_Y Y}$ e $\lambda_{j_X X}$, respectivamente, $j_Y = 1, 2, \dots, p$, e $j_X = 1, 2, \dots, q$, não sendo assim necessário considerá-los novamente em (3.2) e em (3.6). O parâmetro ϕ é estimado maximizando a função log-verossimilhança penalizada dada na equação (3.2). Já os parâmetros γ e φ são estimados maximizando a função log-verossimilhança penalizada dada na equação (3.6). Em ambos os casos, os parâmetros não são penalizados.

Note que os valor dos parâmetros $\lambda_{j_Y Y}$, $j_Y = 1, 2, \dots, p$, são encontrados utilizando o critério validação cruzada generalizada, como dado na equação (3.3), considerando a

nova base não singular e a respectiva matriz de penalização. Deve-se considerar uma malha de valores para cada um dos $\lambda_{j_Y Y}$, $j_Y = 1, 2, \dots, p$, e então, verificar todas as possíveis combinações. Este processo pode gerar um alto custo computacional. Desta forma, as escolhas dos valores da malha devem ser realizadas com cuidado. Analogamente, determinam-se os valores de $\lambda_{j_X X}$, $j_X = 1, 2, \dots, q$, através da validação cruzada generalizada, como dada pela equação (3.7), considerando as novas base e matriz de penalização. Novamente, todas as combinações possíveis devem ser analisadas.

As matrizes de covariâncias das curvas ajustadas $\mathbf{B}_Y^* \hat{\mathbf{a}}_Y^*$ e $\mathbf{B}_X^* \hat{\mathbf{a}}_X^*$, são dadas como nas equações (3.4) e (3.8), respectivamente, notando que os parâmetros $\lambda_{j_Y Y}$, $j_Y = 1, 2, \dots, p$, e $\lambda_{j_X X}$, $j_X = 1, 2, \dots, q$, já estão inseridos nas matrizes de penalização e portanto, não devem ser consideradas novamente.

3.3 Modelos bivariados com P-splines

Na Seção 3.2, a metodologia P-splines é utilizada individualmente no ajuste de cada um dos modelos marginais associados às variáveis respostas discreta e contínua. Entretanto, os modelos da classe proposta são bivariados em que uma estrutura de dependência entre as variáveis resposta discreta e contínua é inserida na formulação da classe. Ajustar os modelos individuais neste caso, poderia gerar resultados não realísticos. Desta forma, a estimação deve ser realizada conjuntamente.

Propomos nesta seção, estender o uso da técnica P-splines para os modelos bivariados semiparamétricos, em que ambos os preditores são dados no contexto dos modelos aditivos generalizados. Considere o modelo bivariado apresentado no Capítulo 2 com preditores dados pelas equações (2.21) e (2.22), ou seja, formados por uma soma de funções suaves dos dados e um termo referente ao intercepto. É possível representar cada uma das funções suaves dos dados como uma regressão linear, utilizando ba-

ses B-splines. Suponha que $\boldsymbol{\eta}_Y$ e $\boldsymbol{\eta}_X$ sejam representados como nas equações (3.9) e (3.11), respectivamente. Além disso, sem perda de generalidade, suponha que \mathbf{B}_Y e \mathbf{B}_X sejam bases não singulares (se não forem, utilizar procedimento dado na Subseção 3.2.1), \mathbf{a}_Y e \mathbf{a}_X , sejam os vetores de coeficientes de regressão associados à \mathbf{B}_Y e \mathbf{B}_X , respectivamente, e \mathbf{P}_Y e \mathbf{P}_X sejam as respectivas matrizes de penalização. Utilizando a função log-verossimilhança conjunta, dada pela equação (2.11), propomos estimar os coeficientes \mathbf{a}_Y e \mathbf{a}_X e os parâmetros γ , ϕ e φ do modelo conjuntamente, maximizando a função log-verossimilhança conjunta penalizada, dada por

$$\begin{aligned} \ell^*(\mathbf{a}_Y, \mathbf{a}_X, \gamma, \phi, \varphi \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T}) &= \frac{1}{a(\phi)} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \frac{1}{a^*(\varphi)} \sum_{i=1}^n [x_i \vartheta_i - b^*(\vartheta_i)] \\ &+ \sum_{i=1}^n c(y_i, \phi) + \sum_{i=1}^n c^*(x_i, \varphi) \\ &- \frac{1}{2} \mathbf{a}_Y^\top \mathbf{P}_Y \mathbf{a}_Y - \frac{1}{2} \mathbf{a}_X^\top \mathbf{P}_X \mathbf{a}_X, \end{aligned} \quad (3.17)$$

em que $\boldsymbol{\mu}_Y = g_1^{-1}(\mathbf{B}_Y \mathbf{a}_Y)$ e $\boldsymbol{\mu}_X = g_2^{-1}(\mathbf{B}_X \mathbf{a}_X)$. Os parâmetros de suavização $\lambda_{1Y}, \lambda_{2Y}, \dots, \lambda_{pY}, \lambda_{1X}, \lambda_{2X}, \dots, \lambda_{qX}$, já estão inseridos nas matrizes de penalização \mathbf{P}_Y e \mathbf{P}_X , e por isso, são desconsiderados na equação (3.17). Note que a função log-verossimilhança conjunta penalizada dada pela equação (3.17) penaliza os coeficientes \mathbf{a}_Y , como na equação (3.2) e os coeficientes \mathbf{a}_X , como na equação (3.6). Os demais parâmetros não são penalizados.

O estimador de máxima verossimilhança para os parâmetros do modelo pode ser obtido utilizando a metodologia apresentada na Subseção 2.1.1 já que o problema de estimação semiparamétrico é transformado em um problema paramétrico, através da técnica P-splines. Note porém, que neste caso a função log-verossimilhança conjunta é penalizada nos coeficientes de regressão \mathbf{a}_Y e \mathbf{a}_X associados às bases B-splines. Assim, os componentes do vetor escore correspondente aos vetores \mathbf{a}_Y e \mathbf{a}_X são dados, respectivamente, por

$$\begin{aligned}
U_p(\mathbf{a}_Y) &= U(\boldsymbol{\beta})|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X} - \mathbf{P}_Y \mathbf{a}_Y \\
\text{e } U_p(\mathbf{a}_X) &= U(\boldsymbol{\Delta})|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X} - \mathbf{P}_X \mathbf{a}_X,
\end{aligned}$$

em que $U(\boldsymbol{\beta})|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X}$ e $U(\boldsymbol{\Delta})|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X}$, são os componentes dados pelas equações (2.12) e (2.13), respectivamente, considerando que $\boldsymbol{\beta} = \mathbf{B}_1 = \mathbf{a}_Y$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}_1 = \mathbf{a}_X$, $\mathbf{S}_1 = \mathbf{Z}_1 = \mathbf{B}_Y$ e $\mathbf{S}_2 = \mathbf{T}_1 = \mathbf{B}_X$, e \mathbf{P}_Y e \mathbf{P}_X são as matrizes de penalização associadas à \mathbf{a}_Y e \mathbf{a}_X , respectivamente. Os demais componentes da função escore correspondentes aos parâmetros γ , ϕ e φ , são dados como nas equações (2.14), (2.15) e (2.16), respectivamente, considerando também que $\boldsymbol{\beta} = \mathbf{B}_1 = \mathbf{a}_Y$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}_1 = \mathbf{a}_X$, $\mathbf{S}_1 = \mathbf{Z}_1 = \mathbf{B}_Y$ e $\mathbf{S}_2 = \mathbf{T}_1 = \mathbf{B}_X$. Denotaremos tais componentes por $U_p(\gamma)$, $U_p(\phi)$ e $U_p(\varphi)$, respectivamente, e referiremos ao vetor escore $U_p = (U_p(\mathbf{a}_Y)^\top, U_p(\mathbf{a}_X)^\top, U_p(\gamma), U_p(\phi), U_p(\varphi))^\top$, associado à função log-verossimilhança conjunta penalizada dada na equação (3.17), por vetor escore penalizado.

A matriz de informação observada \mathbf{M} utilizada no processo de estimação é dada como na equação (2.20) porém, com as submatrizes \mathbf{M}_β e \mathbf{M}_Δ sendo substituídas pelas submatrizes $\mathbf{M}_{\mathbf{a}_Y}^*$ e $\mathbf{M}_{\mathbf{a}_X}^*$, respectivamente. Tais submatrizes são dadas por

$$\begin{aligned}
\mathbf{M}_{\mathbf{a}_Y}^* &= \mathbf{M}_\beta|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X} + \mathbf{P}_Y \\
\mathbf{M}_{\mathbf{a}_X}^* &= \mathbf{M}_\Delta|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X} + \mathbf{P}_X,
\end{aligned}$$

em que $\mathbf{M}_\beta|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X}$ e $\mathbf{M}_\Delta|_{\mathbf{a}_Y, \mathbf{a}_X, \mathbf{B}_Y, \mathbf{B}_X}$ são as submatrizes \mathbf{M}_β e \mathbf{M}_Δ de \mathbf{M} considerando $\boldsymbol{\beta} = \mathbf{B}_1 = \mathbf{a}_Y$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}_1 = \mathbf{a}_X$, $\mathbf{S}_1 = \mathbf{Z}_1 = \mathbf{B}_Y$ e $\mathbf{S}_2 = \mathbf{T}_1 = \mathbf{B}_X$, e \mathbf{P}_Y e \mathbf{P}_X são as matrizes de penalização associadas à \mathbf{a}_Y e \mathbf{a}_X , respectivamente. As demais submatrizes de \mathbf{M} não sofrem alterações porém, também são calculadas considerando $\boldsymbol{\beta} = \mathbf{B}_1 = \mathbf{a}_Y$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}_1 = \mathbf{a}_X$, $\mathbf{S}_1 = \mathbf{Z}_1 = \mathbf{B}_Y$ e $\mathbf{S}_2 = \mathbf{T}_1 = \mathbf{B}_X$. Denotaremos a matriz resultante por \mathbf{M}_p e a referiremos por matriz de informação penalizada, o qual é associado à função log-verossimilhança penalizada dada na equação (3.17).

Quando os parâmetros de suavização $\lambda_{1Y}, \lambda_{2Y}, \dots, \lambda_{pY}, \lambda_{1X}, \lambda_{2X}, \dots, \lambda_{qX}$ são todos iguais a 0, o modelo bivariado semiparamétrico é reduzido a um modelo bivariado

paramétrico, como apresentado na Seção 2.1, com preditores formados por termos lineares (bases B-splines). Desta forma, a matriz de covariâncias assintótica é dada pelo inverso da matriz de informação esperada ou observada, como na equação (2.20). Entretanto, os valores dos parâmetros de suavização em geral são maiores do que 0. Consequentemente, a matriz de covariâncias assintótica tende a subestimar as variâncias, dependendo dos valores dos parâmetros de suavização. Como o interesse está na curva ajustada e consequentemente, na matriz de covariâncias associada à tal curva, propomos um procedimento análogo ao realizado na Subseção 3.1.3, isto é, encontrar uma matriz \mathbf{L}_Y tal que $\hat{\mathbf{y}} = \mathbf{L}_Y \mathbf{y}^*$, para alguma variável modificada \mathbf{y}^* e então, calcular a covariância da curva ajustada $\mathbf{B}_Y \hat{\mathbf{a}}_Y$ pela relação $\text{Cov}(\mathbf{B}_Y \hat{\mathbf{a}}_Y) = \text{Cov}(\mathbf{L}_Y \mathbf{y}^*)$.

Utilizando o componente $\mathbf{U}_p(\mathbf{a}_Y)$ do vetor escore penalizado, e a submatriz $\mathbf{M}_{\mathbf{a}_Y}$ da matriz de informação penalizada, ambos associados ao vetor \mathbf{a}_Y , é possível escrever um processo iterativo associado à \mathbf{a}_Y , isto é,

$$\mathbf{a}_Y^1 = \mathbf{a}_Y^0 + [(\mathbf{M}_{\mathbf{a}_Y})^{-1}]^0 [\mathbf{U}_p(\mathbf{a}_Y)]^0,$$

com \mathbf{a}_Y^0 e \mathbf{a}_Y^1 , sendo os valores do vetor \mathbf{a}_Y avaliados no passo 0 e 1, respectivamente, $[(\mathbf{M}_{\mathbf{a}_Y})^{-1}]^0$ sendo a submatriz $[(\mathbf{M}_{\mathbf{a}_Y})^{-1}]$ avaliada no passo 0 e $[\mathbf{U}_p(\mathbf{a}_Y)]^0$ sendo o vetor $[\mathbf{U}_p(\mathbf{a}_Y)]$ avaliado no passo 0. Substituindo $\mathbf{U}_p(\mathbf{a}_Y)$ e $\mathbf{M}_{\mathbf{a}_Y}$ por suas respectivas expressões, e após algumas álgebras, é possível escrever $\hat{\boldsymbol{\eta}}_Y = \mathbf{B}_Y \hat{\mathbf{a}}_Y = \mathbf{L}_Y \mathbf{y}^*$. Segue que a matriz de covariâncias da curva ajustada $\mathbf{B}_Y \hat{\mathbf{a}}_Y$, utilizada para a construção das bandas de confiança, é dada por

$$\text{Cov}(\mathbf{B}_Y \hat{\mathbf{a}}_Y) = \text{Cov}(\mathbf{L}_Y \mathbf{y}^*) \approx \mathbf{L}_Y \text{Cov}(\mathbf{y}^*) \mathbf{L}_Y^\top. \quad (3.18)$$

A matriz \mathbf{L}_Y calculada em (3.18) pode ser utilizada também para encontrar os valores dos parâmetros de suavização $\lambda_{1Y}, \lambda_{2Y}, \dots, \lambda_{pY}$, através da validação cruzada generalizada, em procedimento análogo ao desenvolvido na Subseção 3.1.3.

De maneira análoga, é possível escrever $\hat{\boldsymbol{\eta}}_X = \mathbf{B}_X \hat{\mathbf{a}}_X = \mathbf{L}_X \mathbf{x}^*$. Segue que a matriz de covariâncias da curva ajustada $\mathbf{B}_X \hat{\mathbf{a}}_X$, utilizada para a construção das bandas de confiança, é dada por

$$\text{Cov}(\mathbf{B}_X \hat{\mathbf{a}}_X) = \text{Cov}(\mathbf{L}_X \mathbf{x}^*) \approx \mathbf{L}_X \text{Cov}(\mathbf{x}^*) \mathbf{L}_X^\top. \quad (3.19)$$

A matriz \mathbf{L}_X em (3.19) pode ser utilizada também para encontrar os valores dos parâmetros de suavização $\lambda_{1X}, \lambda_{2X}, \dots, \lambda_{qX}$, através da validação cruzada generalizada, em procedimento análogo ao desenvolvido na Subseção 3.1.3. Para os parâmetros γ, ϕ e φ , pode-se utilizar a variância assintótica habitual, quando necessária.

Como as funções suaves de ambos os preditores são representadas por termos lineares, é natural utilizar o mesmo procedimento para estimar o modelo bivariado com preditores formados por termos lineares e não paramétricos. De fato, considere

$$\eta_{iY} = \beta_0 + \beta_1 z_{i1} + \dots + \beta_k z_{ik} + h_2(z_{ik+1}, \dots, z_{ip}) = \mathbf{Z}_{i1} \mathbf{B}_1 + h_2(\mathbf{Z}_{i3}) \quad (3.20)$$

$$\text{e } \eta_{iX} = \delta_0 + \delta_1 t_{i1} + \dots + \delta_d t_{id} + h_4(t_{id+1}, \dots, t_{iq}) = \mathbf{T}_{i1} \mathbf{\Delta}_1 + h_4(\mathbf{T}_{i3}), \quad (3.21)$$

em que $\mathbf{B}_1 = (\beta_0, \beta_1, \dots, \beta_k)^\top$ e $\mathbf{\Delta}_1 = (\delta_0, \delta_1, \dots, \delta_d)^\top$ são os coeficientes de regressão associados à parte linear dos preditores, $h_2(z_{ik+1}, \dots, z_{ip}) = \sum_{j_Y=k+1}^p s_{j_Y}(z_{ij_Y})$ e $h_4(t_{id+1}, \dots, t_{iq}) = \sum_{j_X=d+1}^q u_{j_X}(t_{ij_X})$ são soma de funções suaves dos dados, $j_Y = k+1, k+2, \dots, p$, e $j_X = d+1, d+2, \dots, q$, e $\mathbf{Z}_{i1} = (1, z_{i1}, \dots, z_{ik})$, $\mathbf{Z}_{i3} = (z_{ik+1}, \dots, z_{ip})$, $\mathbf{T}_{i1} = (1, t_{i1}, \dots, t_{id})$ e $\mathbf{T}_{i3} = (t_{id+1}, \dots, t_{iq})$ são vetores com os valores das covariáveis na i -ésima observação, $i = 1, 2, \dots, n$.

Os preditores dados pelas equações (3.20) e (3.21) podem ser representados como nas equações (3.9) e (3.11), respectivamente. Neste caso, sem perda de generalidade, $s_1(\mathbf{z}_1) = \mathbf{Z}_1 \mathbf{B}_1$ e $u_1(\mathbf{t}_1) = \mathbf{T}_1 \mathbf{\Delta}_1$, isto é, as funções suaves $s_1(\cdot)$ e $u_1(\cdot)$ são reescritas pelos termos lineares dos preditores, com \mathbf{Z}_1 e \mathbf{T}_1 como na Subseção 2.1.1. Assim, $\mathbf{B}_{1Y} = \mathbf{Z}_1$ e $\mathbf{B}_{1X} = \mathbf{T}_1$ são as bases associadas a $s_1(\cdot)$ e $u_1(\cdot)$, respectivamente, $\mathbf{a}_{1Y} = \mathbf{B}_1$ e $\mathbf{a}_{1X} = \mathbf{\Delta}_1$ são os coeficientes de regressão e $\mathbf{P}_{1Y} = \mathbf{0}$ e $\mathbf{P}_{1X} = \mathbf{0}$ são as matrizes de penalização associadas a \mathbf{a}_{1Y} e \mathbf{a}_{1X} , respectivamente. A dimensão de \mathbf{P}_{1Y} é $(k+1) \times (k+1)$ e de \mathbf{P}_{1X} , $(d+1) \times (d+1)$. As demais funções suaves são representadas como nas equações (3.9) e (3.11).

Sendo \mathbf{B}_Y e \mathbf{B}_X as bases não singulares (vide Subseção 3.2.1), \mathbf{a}_Y e \mathbf{a}_X os respec-

tivos vetores de coeficientes de regressão em que $\boldsymbol{\eta}_Y = \mathbf{B}_Y \mathbf{a}_Y$ e $\boldsymbol{\eta}_X = \mathbf{B}_X \mathbf{a}_X$, e \mathbf{P}_Y e \mathbf{P}_X as matrizes de penalização associadas à \mathbf{a}_Y e \mathbf{a}_X , respectivamente, pode-se ajustar o modelo utilizando o vetor escore penalizado e a matriz de informação observada penalizada, como realizado no início desta seção. Note que as matrizes de penalização utilizadas no processo de estimação penalizam apenas os coeficientes associados aos termos não paramétricos.

Quando termos não lineares também são considerados, não é possível representar os preditores somente por termos lineares, como visto anteriormente. Entretanto, a metodologia de estimação apresentada na Subseção 2.1.1 pode ser utilizada. De fato, considere o caso geral em que os preditores são dados como nas equações (2.2) e (2.4). A parte linear e a parte não paramétrica dos preditores $\boldsymbol{\eta}_Y$ e $\boldsymbol{\eta}_X$ podem ser representadas por $\mathbf{B}_Y \mathbf{a}_Y$ e $\mathbf{B}_X \mathbf{a}_X$, respectivamente, em que \mathbf{B}_Y e \mathbf{B}_X são bases não singulares e \mathbf{a}_Y e \mathbf{a}_X são coeficientes de regressão. Desta forma, têm-se os preditores formados por termos lineares e não lineares paramétricos. Conseqüentemente, utilizando a metodologia apresentada na Subseção 2.1, é possível obter o estimador de máxima verossimilhança para os parâmetros do modelo. Considerando a função log-verossimilhança conjunta penalizada dada pela equação (3.17), têm-se que os componentes do vetor escore correspondente aos vetores $\boldsymbol{\beta} = (\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top$ e $\boldsymbol{\Delta} = (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top$, são dados respectivamente por

$$U_p(\boldsymbol{\beta}) = \mathbf{U}(\boldsymbol{\beta})|_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{s}_1, \mathbf{s}_2} - \mathbf{P}_Y \boldsymbol{\beta} \quad (3.22)$$

$$\text{e } U_p(\boldsymbol{\Delta}) = \mathbf{U}(\boldsymbol{\Delta})|_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{s}_1, \mathbf{s}_2} - \mathbf{P}_X \boldsymbol{\Delta}, \quad (3.23)$$

em que $\mathbf{U}(\boldsymbol{\beta})|_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{s}_1, \mathbf{s}_2}$ e $\mathbf{U}(\boldsymbol{\Delta})|_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{s}_1, \mathbf{s}_2}$, são os componentes dados pelas equações (2.12) e (2.13), respectivamente, considerando que $\boldsymbol{\beta} = (\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top$, $\boldsymbol{\Delta} = (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top$, com \mathbf{B}_2 e $\boldsymbol{\Delta}_2$ sendo os vetores de coeficientes de regressão associados à parte não linear dos preditores $\boldsymbol{\eta}_Y$ e $\boldsymbol{\eta}_X$, respectivamente,

$$\mathbf{S}_1 = \begin{pmatrix} & \mathbf{B}_Y^\top & & \\ \frac{\partial \eta_{1Y}}{\partial \beta_{k+1}} & \frac{\partial \eta_{2Y}}{\partial \beta_{k+1}} & \cdots & \frac{\partial \eta_{mY}}{\partial \beta_{k+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_{1Y}}{\partial \beta_m} & \frac{\partial \eta_{2Y}}{\partial \beta_m} & \cdots & \frac{\partial \eta_{mY}}{\partial \beta_m} \end{pmatrix}^\top \text{ e } \mathbf{S}_2 = \begin{pmatrix} & \mathbf{B}_X^\top & & \\ \frac{\partial \eta_{1X}}{\partial \delta_{d+1}} & \frac{\partial \eta_{2X}}{\partial \delta_{d+1}} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_{d+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_{1X}}{\partial \delta_o} & \frac{\partial \eta_{2X}}{\partial \delta_o} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_o} \end{pmatrix}^\top, \quad (3.24)$$

e \mathbf{P}_Y e \mathbf{P}_X são as matrizes de penalização associadas à $\boldsymbol{\beta}$ e $\boldsymbol{\Delta}$, respectivamente, sendo que apenas os coeficientes de \mathbf{a}_Y e \mathbf{a}_X associados a $s_{j_Y}(\cdot)$ e $u_{j_X}(\cdot)$ são penalizados.

Os demais componentes da função escore correspondentes aos parâmetros γ , ϕ e φ , são dados como nas equações (2.14), (2.15) e (2.16), respectivamente, considerando $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, \mathbf{S}_1 e \mathbf{S}_2 como nas equações (3.22) e (3.23). Denotaremos tais componentes por $\mathbf{U}_p(\gamma)$, $\mathbf{U}_p(\phi)$ e $\mathbf{U}_p(\varphi)$, respectivamente. O vetor escore penalizado é dado por $\mathbf{U}_p = (\mathbf{U}_p(\boldsymbol{\beta}))^\top, \mathbf{U}_p(\boldsymbol{\Delta})^\top, \mathbf{U}_p(\gamma), \mathbf{U}_p(\phi), \mathbf{U}_p(\varphi))^\top$.

A matriz de informação observada penalizada \mathbf{M}_p associada à função log-verosimilhança conjunta penalizada dada pela equação (3.17) e utilizada no processo de estimação, é dada por

$$\mathbf{M}_p = \begin{bmatrix} \mathbf{M}_\beta^p & \mathbf{M}_{\beta\Delta}^p & \mathbf{M}_{\beta\gamma}^p & \mathbf{M}_{\beta\phi}^p & \mathbf{M}_{\beta\varphi}^p \\ \mathbf{M}_{\beta\Delta}^{p\top} & \mathbf{M}_\Delta^p & \mathbf{M}_{\Delta\gamma}^p & \mathbf{0}_{(q+1)\times 1} & \mathbf{M}_{\Delta\varphi}^p \\ \mathbf{M}_{\beta\gamma}^{p\top} & \mathbf{M}_{\Delta\gamma}^{p\top} & \mathbf{M}_\gamma^p & 0 & \mathbf{M}_{\gamma\varphi}^p \\ \mathbf{M}_{\beta\phi}^{p\top} & \mathbf{0}_{1\times(q+1)} & 0 & \mathbf{M}_\phi^p & 0 \\ \mathbf{M}_{\beta\varphi}^{p\top} & \mathbf{M}_{\Delta\varphi}^{p\top} & \mathbf{M}_{\gamma\varphi}^{p\top} & 0 & \mathbf{M}_\varphi^p \end{bmatrix}, \quad (3.25)$$

com as submatrizes \mathbf{M}_β^p e \mathbf{M}_Δ^p dadas por

$$\begin{aligned} \mathbf{M}_\beta^p &= \mathbf{M}_\beta |_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{S}_1, \mathbf{S}_2} + \mathbf{P}_Y \\ \text{e } \mathbf{M}_\Delta^p &= \mathbf{M}_\Delta |_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{S}_1, \mathbf{S}_2} + \mathbf{P}_X, \end{aligned}$$

em que $\mathbf{M}_\beta |_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{S}_1, \mathbf{S}_2}$ e $\mathbf{M}_\Delta |_{(\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top, (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top, \mathbf{S}_1, \mathbf{S}_2}$ são as submatrizes \mathbf{M}_β e \mathbf{M}_Δ dadas na equação (2.20), considerando $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, \mathbf{S}_1 e \mathbf{S}_2 como nas equações (3.22) e (3.23). Já as matrizes de penalização \mathbf{P}_Y e \mathbf{P}_X são análogas as consideradas nas

equações (3.22) e (3.23), respectivamente.

As submatrizes $M_{\beta\Delta}^p$, $M_{\beta\gamma}^p$, $M_{\beta\phi}^p$, $M_{\beta\varphi}^p$, $M_{\Delta\gamma}^p$, $M_{\Delta\varphi}^p$, M_{γ}^p , $M_{\gamma\varphi}^p$, M_{ϕ}^p e M_{φ}^p são as submatrizes $M_{\beta\Delta}$, $M_{\beta\gamma}$, $M_{\beta\phi}$, $M_{\beta\varphi}$, $M_{\Delta\gamma}$, $M_{\Delta\varphi}$, M_{γ} , $M_{\gamma\varphi}$, M_{ϕ} e M_{φ} , respectivamente, dadas na equação (2.20), calculadas adotando β , Δ , S_1 , S_2 , P_Y e P_X como nas equações (3.22) e (3.23).

Desta forma, tem-se que a matriz de informação esperada penalizada M_{E_p} associada à função log-verossimilhança conjunta penalizada dada pela equação (3.17), é obtida substituindo os elementos da matriz M_p por seus respectivos valores esperados com respeito às distribuições das variáveis respostas Y_i e X_i .

Note que a inserção de termos paramétricos não lineares nos preditores pode gerar dificuldades na obtenção da inversa da matriz de informação observada, necessária no método de estimação. Uma alternativa a esse problema é substituir o termo não linear da covariável por uma função suave não paramétrica da mesma. A vantagem desta substituição é que o preditor se reduz a uma soma de termos lineares (parte linear e bases B-splines). Entretanto, em problemas cujo interesse está na interpretação, a substituição deve ser evitada.

Capítulo 4

Casos particulares

A classe de modelos bivariados com respostas discreta e contínua, apresentada no Capítulo 2, é mais geral no sentido que engloba uma variedade de modelos. Para determinadas escolhas de distribuições e funções, é possível chegar a modelos já conhecidos na literatura, e bem explorados nos dias atuais, assim como a novos modelos.

Neste capítulo, aplicamos os resultados anteriores a alguns modelos pertencentes à classe geral proposta. Consideramos o modelo de regressão bivariado Bernoulli-normal proposto por Fitzmaurice & Laird (1995) e o modelo de regressão bivariado Poisson-normal proposto por Yang *et al.* (2007). Além disso, construímos o modelo de regressão bivariado Bernoulli-exponencial e o modelo de regressão bivariado Poisson-normal semiparamétrico, uma extensão do modelo bivariado Poisson-normal.

4.1 Modelo Bernoulli-normal

Introduzido por Fitzmaurice & Laird (1995), o modelo de regressão bivariado Bernoulli-normal pertence à classe geral proposta no Capítulo 2. Suponha que Y_i é uma variável aleatória discreta tal que $Y_i \sim \text{Bernoulli}(\mu_{iY})$, em que μ_{iY} é a média de

Y_i , $i = 1, \dots, n$. A distribuição de Y_i pode ser escrita na forma da equação (2.1) com

$$\theta_i = \log\left(\frac{\mu_{iY}}{1 - \mu_{iY}}\right), \quad b(\theta_i) = \log[1 + \exp(\theta_i)], \quad a(\phi) = 1 \quad \text{e} \quad c(y_i, \phi) = 0.$$

Neste caso, a função de variância é dada por $V(\mu_{iY}) = \mu_{iY}(1 - \mu_{iY})$.

Assuma que um conjunto de p covariáveis $z_{i1}, z_{i2}, \dots, z_{ip}$ esteja disponível para prever Y_i e considere a relação dada na equação (2.2), com $h_1(\cdot) \equiv h_2(\cdot) \equiv 0$ e $k = p$, isto é, o preditor é linear. Assim,

$$\eta_{iY} = \beta_0 + \beta_1 z_{i1} + \dots + \beta_p z_{ip} = \mathbf{Z}_{i1} \mathbf{B}_1. \quad (4.1)$$

Assuma também que $g_1(\cdot)$ é a função de ligação canônica para o modelo Bernoulli, ou seja,

$$g_1(\mu_{iY}) = \log\left(\frac{\mu_{iY}}{1 - \mu_{iY}}\right) = \eta_{iY}.$$

Logo,

$$\mu_{iY} = g_1^{-1}(\eta_{iY}) = \frac{\exp(\eta_{iY})}{1 + \exp(\eta_{iY})}.$$

Por outro lado, suponha que X_i é uma variável aleatória contínua tal que $X_i | Y_i = y_i \sim N(\alpha_i, \sigma^2)$, em que α_i e σ^2 são a média e a variância, respectivamente, de $X_i | Y_i = y_i$, $i = 1, \dots, n$. A função densidade de $X_i | Y_i = y_i$ pode ser escrita na forma da equação (2.3) com

$$\vartheta_i = \alpha_i, \quad b^*(\vartheta_i) = \frac{\vartheta_i^2}{2}, \quad \frac{1}{a^*(\varphi)} = \frac{1}{\varphi} = \frac{1}{\sigma^2}, \quad \text{e} \quad c^*(x_i, \varphi) = -\frac{1}{2} \log(2\pi\varphi) - \frac{1}{2} \frac{x_i^2}{\varphi}.$$

A função de variância é dada por $Q(\alpha_i) = 1$.

Considerando que o mesmo conjunto de covariáveis $z_{i1}, z_{i2}, \dots, z_{ip}$ esteja disponível para prever X_i e a relação dada na equação (2.4) com $t_{ij_X} = z_{ij_X}$ para $i = 1, \dots, n$, e $j_X = 1, \dots, p$, $h_3(\cdot) \equiv h_4(\cdot) \equiv 0$ e $d = p$, tem-se que

$$\eta_{iX} = \delta_0 + \delta_1 z_{i1} + \dots + \delta_p z_{ip} = \mathbf{Z}_{i1} \mathbf{\Delta}_1,$$

com \mathbf{Z}_{i1} dado como na equação (4.1).

Assuma que $g_2(\cdot)$ e $g_3(\cdot)$ são funções de ligação identidade e $h_5(\cdot)$ é dada por

$$h_5(y_i, \mu_{iY}, \mu_{iX}, \gamma) = \mu_{iX} + \gamma(y_i - \mu_{iY}) = \xi_i.$$

Da relação dada na equação (2.6),

$$\alpha_i = \mu_{iX} + \gamma(y_i - \mu_{iY}).$$

Note que $E[\alpha_i] = E[\mu_{iX} + \gamma(y_i - \mu_{iY})] = E[\mu_{iX}] = \mu_{iX}$, indicando que as escolhas de $g_3(\cdot)$ e $h_5(\cdot)$ estão coerentes com a construção do modelo.

4.1.1 Estimação

Os vetores $\boldsymbol{\beta}$ e $\boldsymbol{\Delta}$ são compostos apenas pelos coeficientes associados à parte linear dos preditores, ou seja, \mathbf{B}_1 e $\boldsymbol{\Delta}_1$, respectivamente. Os componentes do vetor escore em termos de \mathbf{B}_1 , $\boldsymbol{\Delta}_1$ e γ , são dados pelas equações (2.17), (2.18) e (2.19), são respectivamente

$$\begin{aligned} \mathbf{U}(\mathbf{B}_1) &= \mathbf{S}_1^\top [(y - \boldsymbol{\mu}_Y) + \sigma^{-2} \mathbf{V} \mathbf{J}_Y (\mathbf{x} - \boldsymbol{\alpha})] \\ &= \sum_{i=1}^n \left[\mathbf{Z}_{i1}^\top (y_i - \mu_{iY}) - \mathbf{Z}_{i1}^\top (\mu_{iY} (1 - \mu_{iY})) (x_i - \alpha_i) \gamma \sigma^{-2} \right], \end{aligned} \quad (4.2)$$

$$\mathbf{U}(\boldsymbol{\Delta}_1) = \sigma^{-2} \mathbf{S}_2^\top \mathbf{J}_X \mathbf{O} (\mathbf{x} - \boldsymbol{\alpha}) = \sum_{i=1}^n \mathbf{Z}_{i1}^\top (x_i - \alpha_i) \sigma^{-2} \quad (4.3)$$

$$\text{e } \mathbf{U}(\gamma) = \sigma^{-2} \mathbf{J}_\gamma^\top (\mathbf{x} - \boldsymbol{\alpha}) = \sum_{i=1}^n (y_i - \mu_{iY}) (x_i - \alpha_i) \sigma^{-2}, \quad (4.4)$$

em que $\mathbf{V} = \text{diag} [\mu_{1Y}(1 - \mu_{1Y}), \dots, \mu_{nY}(1 - \mu_{nY})]$, $\mathbf{J}_Y = -\gamma \mathbf{I}$, \mathbf{I} uma matriz identidade n -dimensional, $\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{Z}_1$ com $\mathbf{Z}_1 = (\mathbf{Z}_{11}^\top, \dots, \mathbf{Z}_{n1}^\top)^\top$, $\mathbf{J}_\gamma = [(y_1 - \mu_{1Y}), \dots, (y_n - \mu_{nY})]^\top$ e $\mathbf{J}_X = \mathbf{O} = \mathbf{I}$.

O componente do vetor escore para σ^2 , em que $a^*(\varphi) = \varphi = \sigma^2$ na equação (2.16), é dado por

$$\mathbf{U}(\sigma^2) = -\frac{1}{(\sigma^2)^2} (\mathbf{x}^\top \boldsymbol{\vartheta} - \mathbf{b}^* \mathbf{1}) + \mathbf{c}^* \mathbf{1} = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n [(x_i - \alpha_i)^2 - \sigma^2], \quad (4.5)$$

$$\boldsymbol{\vartheta} = (\alpha_1, \dots, \alpha_n)^\top, \mathbf{b}^* = \left(\frac{\alpha_1^2}{2}, \dots, \frac{\alpha_n^2}{2} \right) \text{ e } \mathbf{c}^* = \left(-\frac{1}{2\sigma^2} + \frac{x_1^2}{2(\sigma^2)^2}, \dots, -\frac{1}{2\sigma^2} + \frac{x_n^2}{2(\sigma^2)^2} \right).$$

Como $X_i | Y_i = y_i \sim N(\alpha_i, \sigma^2)$, segue que $\left(\frac{X_i - \alpha_i}{\sigma}\right)^2 \sim \chi_1^2$. Se $C_i = (X_i - \alpha_i)^2$, então $\frac{C_i}{\sigma^2} \sim \chi_{(1)}^2$ com $Var(C_i) = 2(\sigma^2)^2$. Assim, definindo $\Psi = Var(C_i)$, a equação (4.5) pode ser reescrita como

$$U(\sigma^2) = \Psi^{-1} \sum_{i=1}^n (c_i - \sigma^2), \quad (4.6)$$

em que $c_i = (x_i - \alpha_i)^2$.

A solução do sistema $\mathbf{U} = \mathbf{0}$ pode ser calculada iterativamente, usando o algoritmo escore de Fisher para a equação (4.2) e o método de mínimos quadrados ponderados para as equações (4.3), (4.4), e (4.6) (Fitzmaurice & Laird, 1995). Neste caso, a matriz de covariâncias assintótica de $\widehat{\mathbf{B}}_1, \widehat{\Delta}_1, \widehat{\gamma}$ e $\widehat{\sigma}^2$ é obtida pelo inverso da matriz de informação esperada \mathbf{M}_E , apresentada na equação (2.20). Esta matriz é dada por

$$\text{cov}(\widehat{\mathbf{B}}_1, \widehat{\Delta}_1, \widehat{\gamma}, \widehat{\sigma}^2) = \sum_{i=n}^n \begin{bmatrix} (\mu_{iY}(1 - \mu_{iY}) + \mu_{iY}^2(1 - \mu_{iY})^2\gamma^2\sigma^{-2})\mathbf{Z}_{i1}^\top\mathbf{Z}_{i1} & -(\sigma^{-2}\gamma\mu_{iY}(1 - \mu_{iY}))\mathbf{Z}_{i1}^\top\mathbf{Z}_{i1} & \mathbf{0} & \mathbf{0} \\ -(\sigma^{-2}\gamma\mu_{iY}(1 - \mu_{iY}))\mathbf{Z}_{i1}^\top\mathbf{Z}_{i1} & \sigma^{-2}\mathbf{Z}_{i1}^\top\mathbf{Z}_{i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu_{iY}(1 - \mu_{iY})\sigma^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2\sigma^4} \end{bmatrix}^{-1}.$$

4.2 Modelo Poisson-normal

Outro modelo pertencente à classe geral apresentada no Capítulo 2 é o modelo de regressão bivariado Poisson-normal (Yang *et al.*, 2007). Neste modelo, a variável resposta discreta, Y_i , segue uma distribuição de Poisson com média μ_{iY} , $i = 1, \dots, n$, e a variável resposta contínua, dada a variável resposta discreta, $X_i | Y_i = y_i$, segue uma distribuição normal com parâmetros α_i e σ^2 , a média e a variância, respectivamente, de $X_i | Y_i = y_i$, $i = 1, \dots, n$. A distribuição de Y_i pode ser escrita na forma da equação (2.1) com

$$\theta_i = \log(\mu_{iY}), \quad b(\theta_i) = \exp(\theta_i), \quad a(\phi) = 1 \quad \text{e} \quad c(y_i, \phi) = -\log(y!).$$

A função de variância é $V(\mu_{iY}) = \mu_{iY}$. Considerando o preditor como na equação (4.1), e usando a função de ligação canônica, $g_1(\cdot)$, para o modelo Poisson, tem-se que

$$\mu_{iY} = g_1^{-1}(\eta_{iY}) = \exp(\eta_{iY}).$$

A função densidade de $X_i | Y_i = y_i$ assim como η_{iX} , $g_2(\cdot)$, $g_3(\cdot)$ e $h_5(\cdot)$ são análogos aos considerados no modelo Bernoulli-normal, apresentado na Seção 4.1.

4.2.1 Estimação

O componente do vetor escore de \mathbf{B}_1 é dado por

$$\begin{aligned} \mathbf{U}(\mathbf{B}_1) &= \mathbf{S}_1^\top [(\mathbf{y} - \boldsymbol{\mu}_Y) + \sigma^{-2} \mathbf{V} \mathbf{J}_Y (\mathbf{x} - \boldsymbol{\alpha})] \\ &= \sum_{i=1}^n \left[\mathbf{Z}_{i1}^\top (y_i - \mu_{iY}) - \mathbf{Z}_{i1}^\top \mu_{iY} (x_i - \alpha_i) \gamma \sigma^{-2} \right], \end{aligned} \quad (4.7)$$

com $\mathbf{V} = \text{diag}[\mu_{1Y}, \dots, \mu_{nY}]$ e \mathbf{S}_1 , \mathbf{J}_Y e \mathbf{Z}_{i1} como na equação (4.2). Os componentes do vetor escore associados aos parâmetros $\boldsymbol{\Delta}_1$, γ e σ^2 são dados pelas equações (4.3), (4.4) e (4.6), respectivamente.

Similarmente, como na Seção 4.1, a solução do sistema $\mathbf{U} = \mathbf{0}$ é encontrada usando o algoritmo escore de Fisher para a equação (4.7), e o método de mínimos quadrados ponderados para as equações (4.3), (4.4) e (4.6). A matriz de covariâncias assintótica de $\widehat{\mathbf{B}}_1$, $\widehat{\boldsymbol{\Delta}}_1$, $\widehat{\gamma}$ e $\widehat{\sigma}^2$ é dada por

$$\text{cov}(\widehat{\mathbf{B}}_1, \widehat{\boldsymbol{\Delta}}_1, \widehat{\gamma}, \widehat{\sigma}^2) = \sum_{i=1}^n \begin{bmatrix} (\mu_{iY} + \mu_{iY}^2 \gamma^2 \sigma^{-2}) \mathbf{Z}_{i1}^\top \mathbf{Z}_{i1} & -(\sigma^{-2} \gamma \mu_{iY}) \mathbf{Z}_{i1}^\top \mathbf{Z}_{i1} & \mathbf{0} & \mathbf{0} \\ -(\sigma^{-2} \gamma \mu_{iY}) \mathbf{Z}_{i1}^\top \mathbf{Z}_{i1} & \sigma^{-2} \mathbf{Z}_{i1}^\top \mathbf{Z}_{i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu_{iY} \sigma^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2\sigma^4} \end{bmatrix}^{-1}.$$

4.3 Modelo Bernoulli-exponencial

Nesta seção, apresentamos o modelo de regressão bivariado Bernoulli-exponencial, também pertencente à classe geral proposta. Considere Y_i , as covariáveis $z_{i1}, z_{i2}, \dots, z_{ip}$,

η_{iY} e $g_1(\cdot)$ como dados na Seção 4.1, $i = 1, 2, \dots, n$. Suponha também que X_i é uma variável resposta contínua e que a distribuição condicional de X_i , dado a variável resposta discreta $Y_i = y_i$, segue uma distribuição exponencial com parâmetro $1/\alpha_i$, em que $\alpha_i = E(X_i | Y_i)$, $i = 1, 2, \dots, n$. A função densidade de $X_i | Y_i = y_i$ pode ser reescrita na forma da equação (2.3) com

$$\vartheta_i = -\frac{1}{\alpha_i}, \quad b^*(\vartheta_i) = -\log(-\vartheta_i), \quad \frac{1}{a^*(\varphi)} = \frac{1}{\varphi} = 1 \quad \text{e} \quad c^*(x_i, \varphi) = 0.$$

A função de variância é dada por $Q(\alpha_i) = \alpha_i^2$.

Assuma que um conjunto de q covariáveis $t_{i1}, t_{i2}, \dots, t_{iq}$, está disponível e considere a relação dada na (2.4) com $h_3(\cdot) \equiv h_4(\cdot) \equiv 0$ e $d = q$. Neste caso, η_{iX} é formado apenas por um termo linear das q covariáveis,

$$\eta_{iX} = \delta_0 + \delta_1 t_{i1} + \dots + \delta_q t_{iq} = \mathbf{T}_{i1} \mathbf{\Delta}_1.$$

Considere $g_2(\cdot)$ a função de ligação logarítmica. Assim, $\mu_{iX} = \exp(\eta_{iX})$. Considere também que $g_3(\cdot)$ é a função de ligação identidade e $h_5(\cdot)$ é dado por

$$h_5(y_i, \mu_{iY}, \mu_{iX}, \gamma) = (1 - y_i + \mu_{iY})\mu_{iX}.$$

Da equação (2.6), tem-se que

$$\alpha_i = (1 - y_i + \mu_{iY})\mu_{iX}.$$

As escolhas de $g_3(\cdot)$ e $h_5(\cdot)$ são adequadas já que a igualdade $E[\alpha_i] = E[(1 - y_i + \mu_{iY})\mu_{iX}] = \mu_{iX}$ é verificada.

4.3.1 Estimação

Os vetores $\boldsymbol{\beta}$ e $\mathbf{\Delta}$ têm somente os coeficientes associados à parte linear, isto é, $\boldsymbol{\beta} = \mathbf{B}_1$ e $\mathbf{\Delta} = \mathbf{\Delta}_1$. Da equação (2.12), com $\mathbf{W} = \mathbf{V}$ já que $g_1(\cdot)$ é a função de ligação canônica para o modelo Bernoulli, o componente do vetor escore referente ao vetor de parâmetros \mathbf{B}_1 , é dado por

$$\begin{aligned}
U(\mathbf{B}_1) &= \mathbf{S}_1^\top \left((\mathbf{y} - \boldsymbol{\mu}_Y) + \mathbf{R}^{\frac{1}{2}} \mathbf{V} \mathbf{Q}^{-\frac{1}{2}} \mathbf{J}_Y (\mathbf{x} - \boldsymbol{\alpha}) \right) \\
&= \sum_{i=1}^n \mathbf{Z}_{i1}^\top \left[(y_i - \mu_{iY}) + \left(\frac{x_i}{\alpha_i} - 1 \right) \frac{\mu_{iY}(1 - \mu_{iY})\mu_{iX}}{\alpha_i} \right], \quad (4.8)
\end{aligned}$$

em que $\mathbf{R} = \text{diag} \left(\frac{1}{\alpha_1^2}, \dots, \frac{1}{\alpha_n^2} \right)$, $\mathbf{V} = \text{diag} [\mu_{1Y}(1 - \mu_{1Y}), \dots, \mu_{nY}(1 - \mu_{nY})]$, $\mathbf{Q} = \text{diag} (\alpha_1^2, \dots, \alpha_n^2)$, $\mathbf{J}_Y = \text{diag} (\mu_{1X}, \dots, \mu_{nX})$ e $\mathbf{S}_1 = \mathbf{Z}_1$ com $\mathbf{Z}_1 = (\mathbf{Z}_{11}^\top, \dots, \mathbf{Z}_{n1}^\top)^\top$.

O componente do vetor escore referente ao vetor de parâmetros $\boldsymbol{\Delta}_1$, da equação (2.13), é dado por

$$U(\boldsymbol{\Delta}_1) = \mathbf{S}_2^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} \mathbf{J}_X \mathbf{O} (\mathbf{x} - \boldsymbol{\alpha}) = \sum_{i=1}^n \mathbf{T}_{i1}^\top \left(\frac{x_i}{\alpha_i} - 1 \right), \quad (4.9)$$

em que $\mathbf{J}_X = \text{diag} [(1 - y_1 + \mu_{1Y}), \dots, (1 - y_n + \mu_{nY})]$, $\mathbf{O} = \text{diag} (\mu_{1X}, \dots, \mu_{nX})$ e $\mathbf{S}_2 = \mathbf{T}_1$ com $\mathbf{T}_1 = (\mathbf{T}_{11}^\top, \dots, \mathbf{T}_{n1}^\top)^\top$.

A solução do sistema $\mathbf{U} = \mathbf{0}$, em que os componentes de \mathbf{U} são dados pelas equações (4.8) e (4.9), é encontrada usando o algoritmo de Newton-Raphson, uma vez que neste caso, a matriz de informação esperada é difícil de ser obtida analiticamente. A matriz de covariâncias assintótica de $\widehat{\mathbf{B}}_1$ e $\widehat{\boldsymbol{\Delta}}_1$, usando a equação (2.20), é dada por

$$\text{cov}(\widehat{\mathbf{B}}_1, \widehat{\boldsymbol{\Delta}}_1) = \begin{bmatrix} \mathbf{M}_\beta & \mathbf{M}_{\beta\Delta} \\ \mathbf{M}_{\beta\Delta}^\top & \mathbf{M}_\Delta \end{bmatrix}^{-1},$$

em que

$$\begin{aligned}
\mathbf{M}_\beta &= \sum_{i=1}^n \mathbf{Z}_{i1}^\top \mathbf{Z}_{i1} \left\{ -\frac{2x_i(\mu_{iY}(1 - \mu_{iY}))^2}{(1 - y_i + \mu_{iY})^3 \mu_{iX}} + \frac{1}{(1 - y_i + \mu_{iY})^2} \left[(\mu_{iY}(1 - \mu_{iY}))^2 + \frac{x_i \mu_{iY}(1 - \mu_{iY})(1 - 2\mu_{iY})}{\mu_{iX}} \right] \right. \\
&\quad \left. - \frac{\mu_{iY}(1 - \mu_{iY})(1 - 2\mu_{iY})}{(1 - y_i + \mu_{iY})} - \mu_{iY}(1 - \mu_{iY}) \right\}, \\
\mathbf{M}_\Delta &= \sum_{i=1}^n \mathbf{Z}_{i2}^\top \mathbf{Z}_{i2} \left[-\frac{x_i}{(1 - y_i + \mu_{iY})} \frac{1}{\mu_{iX}} \right] \\
\text{e } \mathbf{M}_{\beta\Delta}^\top &= \sum_{i=1}^n \mathbf{Z}_{i2}^\top \mathbf{Z}_{i1} \left[-\frac{x_i}{(1 - y_i + \mu_{iY})^2} \frac{\mu_{iY}(1 - \mu_{iY})}{\mu_{iX}} \right].
\end{aligned}$$

4.4 Modelo Poisson-normal semiparamétrico

O modelo de regressão bivariado Poisson-normal semiparamétrico é uma extensão do modelo bivariado Poisson-normal, apresentado na Seção 4.2, e também pertence à classe geral de modelos introduzida no Capítulo 2.

Suponha Y_i uma variável resposta discreta seguindo uma distribuição de Poisson com média μ_{iY} , $i = 1, \dots, n$, e X_i uma variável resposta contínua, tal que $X_i | Y_i = y_i$ segue distribuição normal com parâmetros α_i e σ^2 , a média e a variância, respectivamente, de $X_i | Y_i = y_i$, $i = 1, \dots, n$, como considerados na Seção 4.2.

Assuma que um conjunto de p covariáveis $z_{i1}, z_{i2}, \dots, z_{ip}$, esteja disponível para prever Y_i e que um conjunto de q covariáveis $t_{i1}, t_{i2}, \dots, t_{iq}$, esteja disponível para prever X_i , e considere as relações dadas pelas equações (2.2) e (2.4). Assuma também que as funções de ligação $g_1(\cdot)$, $g_2(\cdot)$, $g_3(\cdot)$ e $h_5(\cdot)$ são análogas às consideradas no modelo Poisson-normal, apresentado na Seção 4.2. Note que neste caso, os preditores são mais gerais do que os considerados na Seção 4.2 no sentido de que são uma soma de termos lineares, de termos não lineares e também de termos não paramétricos das covariáveis disponíveis. Neste contexto, o modelo apresentado é uma extensão do modelo Poisson-normal.

4.4.1 Estimação

Como os preditores são formados por termos paramétricos e não paramétricos das covariáveis disponíveis, o método de estimação apresentado na Seção 3.3 será aplicado. Da equação (3.22), segue que o componente do vetor escore penalizado em termos do vetor de coeficientes de regressão $\boldsymbol{\beta} = (\mathbf{a}_Y^\top, \mathbf{B}_2^\top)^\top$ é dado por

$$\begin{aligned}
\mathbf{U}_p(\boldsymbol{\beta}) &= \mathbf{S}_1^\top [(\mathbf{y} - \boldsymbol{\mu}_Y) + \sigma^{-2} \mathbf{V} \mathbf{J}_Y (\mathbf{x} - \boldsymbol{\alpha})] - \mathbf{P}_Y \boldsymbol{\beta} \\
&= \sum_{i=1}^n [\mathbf{S}_{i1}^\top (y_i - \mu_{iY}) - \mathbf{S}_{i1}^\top \mu_{iY} (x_i - \alpha_i) \gamma \sigma^{-2}] - \mathbf{P}_Y \boldsymbol{\beta}, \quad (4.10)
\end{aligned}$$

com \mathbf{S}_1 dada pela equação (3.24), \mathbf{P}_Y sendo a matriz de penalização dada como na equação (3.22), $\mathbf{V} = \text{diag}[\mu_{1Y}, \dots, \mu_{nY}]$ e \mathbf{J}_Y dada como na equação (4.2), em que $\mu_{iY} = \exp(\mathbf{S}_{i1}^\top \boldsymbol{\beta})$. O vetor \mathbf{S}_{i1} corresponde à i -ésima linha de \mathbf{S}_1 .

Da equação (3.23), tem-se que o componente do vetor escore penalizado em termos do vetor de coeficientes de regressão $\boldsymbol{\Delta} = (\mathbf{a}_X^\top, \boldsymbol{\Delta}_2^\top)^\top$ é dado por

$$\mathbf{U}_p(\boldsymbol{\Delta}) = \sigma^{-2} \mathbf{S}_2^\top \mathbf{J}_X \mathbf{O} (\mathbf{x} - \boldsymbol{\alpha}) - \mathbf{P}_X \boldsymbol{\Delta} = \sum_{i=1}^n \mathbf{S}_{i2}^\top (x_i - \alpha_i) \sigma^{-2} - \mathbf{P}_X \boldsymbol{\Delta}, \quad (4.11)$$

com \mathbf{S}_2 dada pela equação (3.24), \mathbf{P}_X sendo a matriz de penalização dada como na equação (3.23), e \mathbf{J}_X e \mathbf{O} dadas como na equação (4.3). Neste caso $\mu_{iX} = \mathbf{S}_{i2}^\top \boldsymbol{\Delta}$ em que \mathbf{S}_{i2} é o vetor correspondente à i -ésima linha de \mathbf{S}_2 . Os componentes $\mathbf{U}_p(\gamma)$ e $\mathbf{U}_p(\sigma^2)$ do vetor escore penalizados, associados aos parâmetros γ e σ^2 , são dados pelas equações (4.4) e (4.6), respectivamente, considerando $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, \mathbf{S}_1 e \mathbf{S}_2 como nas equações (4.10) e (4.11).

A solução do sistema $\mathbf{U}_p = \mathbf{0}$ é encontrada utilizando o algoritmo escore de Fisher para a equação (4.10), e o método de mínimos quadrados ponderados para as equações (4.11), (4.4) e (4.6). Note que os parâmetros de suavidade incluídos nas matrizes de penalização \mathbf{P}_Y e \mathbf{P}_X , são escolhidos através da validação cruzada generalizada.

A matriz de informação esperada penalizada \mathbf{M}_{Ep} , calculada a partir da equação (3.25), é dada por

$$\mathbf{M}_{Ep} = \sum_{i=1}^n \begin{bmatrix} (\mu_{iY} + \mu_{iY}^2 \gamma^2 \sigma^{-2}) \mathbf{S}_{i1}^\top \mathbf{S}_{i1} + \mathbf{P}_Y & -(\sigma^{-2} \gamma \mu_{iY}) \mathbf{S}_{i1}^\top \mathbf{S}_{i2} & \mathbf{0} & \mathbf{0} \\ -(\sigma^{-2} \gamma \mu_{iY}) \mathbf{S}_{i2}^\top \mathbf{S}_{i1} & \sigma^{-2} \mathbf{S}_{i2}^\top \mathbf{S}_{i2} + \mathbf{P}_X & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu_{iY} \sigma^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2\sigma^4} \end{bmatrix}.$$

com $\mu_{iY} = \exp(\mathbf{S}_{i1}^\top \boldsymbol{\beta})$ e $\mu_{iX} = \mathbf{S}_{i2}^\top \boldsymbol{\Delta}$.

Capítulo 5

Técnicas de diagnóstico e influência

Em muitas situações, modelos de regressão são ajustados a conjuntos de dados sem que as suposições acerca dos modelos estejam satisfeitas. Além disso, muitas vezes os conjuntos de dados utilizados contêm observações discrepantes que podem influenciar no ajuste do modelo. Nestes casos, o ajuste pode não ser o mais adequado, levando a resultados equivocados. Uma etapa essencial na escolha de um modelo é a análise de diagnóstico que consiste na verificação de "possíveis afastamentos de suposições iniciais", relacionadas à parte aleatória e à parte sistemática do modelo, assim como na investigação de possíveis observações discrepantes e observações atípicas, que podem influenciar no ajuste final do modelo (Paula, 2004).

Entre as suposições feitas na classe de modelos bivariados mistos apresentada no Capítulo 2 e que devem ser satisfeitas, estão as relacionadas à independência das variáveis resposta contínuas e discretas, entre os indivíduos, as relacionadas às distribuições iniciais para as respostas, no caso, para a resposta discreta e para a resposta contínua condicionada à discreta, as relacionadas à adequação das funções de ligação e as relacionadas à forma dos preditores.

Uma das primeiras etapas em uma análise de diagnóstico é a análise de resíduos, em

que pode ser verificada a independência das variáveis respostas, entre os indivíduos e se a distribuição inicial para as respostas, as funções de ligação e formas dos preditores adotadas estão adequadas. Além disso, por meio da análise de resíduos é possível detectar a presença de pontos aberrantes. Já a verificação de pontos influentes pode ser feita utilizando técnicas tais como influência global e local, que analisam o quão os resultados (ajuste) se alteram com a deleção e/ou perturbação dos dados e com a perturbação do modelo proposto.

5.1 Resíduos

Como visto em Fitzmaurice & Laird (1995), que apresentaram um modelo bivariado construído a partir de uma distribuição marginal e de uma distribuição condicional, e em Gomes (2007), que apresentou modelos de regressão com resposta bivariada construídos a partir de cópulas, a análise de resíduos em um modelo bivariado em geral é realizada considerando os modelos separadamente, isto é, calculando as medidas de resíduos individualmente para os modelos associados às variáveis resposta discreta e contínua. Assim, é possível verificar a independência das variáveis resposta entre os indivíduos analisando separadamente os gráficos dos resíduos de Y_i e $X_i | Y_i = y_i$ *versus* a ordem de coleta dos dados, se a ordem for importante. Além disso, também é possível verificar a forma dos preditores individualmente por meio dos gráficos dos resíduos *versus* alguma função das covariáveis disponíveis, em geral a função candidata a ser inserida nos componentes sistemáticos dos modelos. Outro gráfico também bastante utilizado é o dos resíduos contra os valores preditos. Resíduos próximos de zero e com uma amplitude constante indicam um bom ajuste do modelo em estudo (Demetrio, 2002).

Propomos neste trabalho calcular os resíduos ordinários padronizados, quantis normalizados aleatorizados e componentes da desviância separadamente em cada modelo,

o associado à variável resposta discreta e o associado à variável resposta contínua condicionada à variável discreta. O resíduo quantil normalizado aleatorizado tem distribuição conhecida para grandes amostras, independente do modelo adotado, quando as estimativas dos parâmetros do modelo são consistentes. As distribuições do resíduo padronizado e do resíduo componente da desviância não são conhecidas. Na teoria dos modelos lineares generalizados as distribuições de tais resíduos são estudadas por meio de dados simulados. Na nossa classe, embora as distribuições da variável discreta e da variável contínua condicionada à discreta pertençam a família exponencial, não é correto afirmar que cada um dos modelos individualmente pertença à classe dos modelos lineares generalizados. Portanto, o comportamento dos resíduos também deve ser analisado por meio de dados simulados.

5.1.1 Resíduos ordinários padronizados

O resíduo padronizado para o modelo associado à variável resposta discreta Y_i , é dado por

$$r_{iY_i}^p = \frac{y_i - \hat{\mu}_{iY}}{\left[a(\hat{\phi}) \hat{V}_i \right]^{1/2}}, \quad (5.1)$$

em que $\hat{\mu}_{iY} = g_1^{-1}(\hat{\eta}_{iY})$ e $\hat{V}_i = V(\hat{\mu}_{iY})$ com $\hat{\eta}_{iY}$ dado pela equação (2.2) substituindo os parâmetros relacionados aos termos paramétricos e às bases B-splines (funções suaves), por suas estimativas de máxima verossimilhança. Note que $a(\hat{\phi})$ é a estimativa consistente de $a(\phi)$. Para o modelo associado à variável contínua X_i , condicionada à discreta Y_i , o resíduo padronizado é dado por

$$r_{iX_i|Y_i}^p = \frac{x_i - \hat{\alpha}_i}{\left[a^*(\hat{\varphi}) \hat{Q}_i \right]^{1/2}}, \quad (5.2)$$

em que $\hat{\alpha}_i = g_3^{-1}(\hat{\xi}_i)$ e $\hat{Q}_i = Q(\hat{\alpha}_i)$ com $\hat{\xi}_i$ dado pela equação (2.5) substituindo μ_{iY} , μ_{iX} e γ por $\hat{\mu}_{iY}$, $\hat{\mu}_{iX}$ e $\hat{\gamma}$, respectivamente. Note que $\hat{\mu}_{iY}$ é dado como na equação (5.1), $\hat{\mu}_{iX} = g_2^{-1}(\hat{\eta}_{iX})$ em que $\hat{\eta}_{iX}$ é dado pela equação (2.4) substituindo os parâmetros relacionados aos termos paramétricos e às bases B-splines (funções suaves), por suas estimativas de máxima verossimilhança e $\hat{\gamma}$ é a estimativa de máxima verossimilhança de γ . Já $a^*(\hat{\varphi})$ é a estimativa consistente de $a^*(\varphi)$.

5.1.2 Resíduos quantis normalizados aleatorizados

O resíduo quantil normalizado aleatorizado (Dunn & Smyth, 1996), para o modelo associado à Y_i é definido por

$$r_{iY_i}^q = \Phi^{-1}(\iota_i) \quad (5.3)$$

em que $\Phi(\cdot)$ é a função distribuição acumulada da distribuição normal padrão e ι_i é um valor aleatório da distribuição uniforme no intervalo $(A_i, B_i]$, com $A_i = \lim_{y \uparrow y_i} F_{Y_i}(y_i | \hat{\mu}_{iY}, a(\hat{\phi}))$ e $B_i = F_{Y_i}(y_i | \hat{\mu}_{iY}, a(\hat{\phi}))$ sendo que $F_{Y_i}(\cdot)$ é a função de distribuição acumulada do modelo associado a Y_i com $\hat{\mu}_{iY}$ e $a(\hat{\phi})$ dados como na equação (5.1). O resíduo quantil normalizado aleatorizado associado à $X_i | Y_i = y_i$ é definido por

$$r_{iX_i|Y_i}^q = \Phi^{-1} [F_{X_i|Y_i}(x_i | \hat{\alpha}_i, a^*(\hat{\varphi}))] \quad (5.4)$$

em que $F_{X_i|Y_i}(\cdot)$ é a função de distribuição acumulada do modelo associado à $X_i | Y_i = y_i$, com $\hat{\alpha}_i$ e $a^*(\hat{\varphi})$ dados como na equação (5.2). No caso em que os preditores são formados apenas por termos paramétricos, as distribuições dos resíduos $r_{iY_i}^q$ e dos resíduos $r_{iX_i|Y_i}^q$ convergem para a distribuição normal padrão desde que as estimativas de máxima verossimilhança dos parâmetros do modelos sejam consistentes.

5.1.3 Resíduos componentes da desviância

Para o modelo associado à Y_i , o resíduo componentes da desviância é definido por

$$r_{iY_i}^d = d^*(y_i | \hat{\mu}_{iY}) = \frac{d(y_i | \hat{\mu}_{iY})}{\sqrt{a(\hat{\phi})}},$$

em que $d(y_i | \hat{\mu}_{iY}) = \{\text{sinal}(y_i - \hat{\mu}_{iY})\} \sqrt{2} \{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\}^{1/2}$, com $\hat{\theta}_i = \theta_i(\hat{\mu}_{iY})$ e $\tilde{\theta}_i = \theta_i(y_i)$ sendo as estimativas de máxima verossimilhança de θ_i sob o modelo com p covariáveis e o modelo saturado, respectivamente. Já $\hat{\mu}_{iY}$ e $a(\hat{\phi})$ são como dados na equação (5.1). Para $X_i | Y_i = y_i$, o resíduo componentes da desviância é definido por

$$r_{iX_i|Y_i}^d = d^*(x_i | \hat{\alpha}_i) = \frac{d(x_i | \hat{\alpha}_i)}{\sqrt{a^*(\hat{\varphi})}},$$

em que $d(x_i | \hat{\alpha}_i) = \{\text{sinal}(x_i - \hat{\alpha}_i)\} \sqrt{2} \{x_i(\tilde{\vartheta}_i - \hat{\vartheta}_i) - b^*(\tilde{\vartheta}_i) + b^*(\hat{\vartheta}_i)\}^{1/2}$, com $\hat{\vartheta}_i = \vartheta_i(\hat{\alpha}_i)$ e $\tilde{\vartheta}_i = \vartheta_i(x_i)$ sendo as estimativas de máxima verossimilhança de ϑ_i sob o modelo com q covariáveis e o modelo saturado, respectivamente. Já $\hat{\alpha}_i$ e $a^*(\hat{\varphi})$ são como dados na equação (5.2).

Em paralelo à verificação das suposições acerca do modelo e da presença de possíveis pontos aberrantes, por meio da análise de resíduos, uma etapa importante na análise de diagnóstico é a investigação do comportamento do modelo em estudo sob a retirada individual de alguma observação ou mesmo sob pequenas perturbações nas observações ou no modelo em geral. Esta etapa é conhecida como análise de influência em que é possível detectar pontos, através de técnicas de diagnóstico, de grande influência no ajuste do modelo, isto é, detectar observações que produzem “grandes” alterações nos resultados da análise, ao serem retiradas da amostra ou então submetidas a pequenas perturbações.

Na próxima seção, aplicamos duas técnicas de diagnóstico de influência para a classe

de modelos considerada na Seção 2.1: influência local e influência global ou total. Para a classe geral com preditores contendo termos não paramétricos, o uso destas técnicas não é habitual.

5.2 Influência local

Utilizada para analisar o comportamento do modelo proposto sob pequenas perturbações nos dados ou no próprio modelo, a técnica de influência local foi proposta por Cook (1986) e tem sido bastante utilizada na literatura em geral (Ortega *et al.*, 2003; Tang *et al.*, 2006; Lee *et al.*, 2006). A técnica tem a vantagem de investigar a sensibilidade do modelo de uma maneira mais específica, isto é, investigando como o modelo se comporta a uma perturbação, por exemplo, em uma determinada covariável ou na variável resposta. Alguns esquemas de perturbações são frequentemente utilizados na análise de influência local (Cook, 1986; Thomas & Cook, 1989). Selecionar um esquema apropriado pode levar a uma inferência mais correta acerca da influência das observações. A técnica desenvolvida em Zhu *et al.* (2007) pode ser utilizada na escolha do esquema. Entretanto, como nossa classe de modelos bivariados é geral, consideramos neste trabalho dois esquemas de perturbação mais comuns utilizados: a perturbação de casos e a perturbação na variável explicativa, deixando a apresentação de outros esquemas assim como a adequabilidade das escolhas, para um trabalho futuro.

Considere $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi) = \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \mathbf{x}, \mathbf{y}, \mathbf{Z}, \mathbf{T})$ a função log-verossimilhança dada pela equação (2.11), em que os vetores $\boldsymbol{\beta}$ e $\boldsymbol{\Delta}$ e γ , ϕ e φ são os parâmetros do modelo e seja $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ um vetor de perturbações de dimensão n pertencente a um subconjunto aberto Ω de \mathbb{R}^n . Desta forma, escolhido o esquema de perturbação, denote por $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \boldsymbol{\omega})$ a função log-verossimilhança correspondente ao modelo perturbado para um dado $\boldsymbol{\omega}$. Note que, definindo $\boldsymbol{\omega}_0$ como sendo o vetor de não perturbação, segue que $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \boldsymbol{\omega}_0) = \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)$.

Sejam $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Delta}}, \hat{\gamma}, \hat{\phi}$ e $\hat{\varphi}$ as estimativas de máxima verossimilhança sob $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)$ e $\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \hat{\boldsymbol{\Delta}}_{\boldsymbol{\omega}}, \hat{\gamma}_{\boldsymbol{\omega}}, \hat{\phi}_{\boldsymbol{\omega}}$ e $\hat{\varphi}_{\boldsymbol{\omega}}$ as estimativas de máxima verossimilhança sob $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \boldsymbol{\omega})$, em que $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \boldsymbol{\omega})$ é assumida continuamente diferenciável de classe C_2 . É possível comparar as estimativas $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Delta}}, \hat{\gamma}, \hat{\phi}, \hat{\varphi}$ com as estimativas $\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \hat{\boldsymbol{\Delta}}_{\boldsymbol{\omega}}, \hat{\gamma}_{\boldsymbol{\omega}}, \hat{\phi}_{\boldsymbol{\omega}}, \hat{\varphi}_{\boldsymbol{\omega}}$, para $\boldsymbol{\omega}$ variando em Ω , utilizando o “afastamento das verossimilhanças” (*likelihood displacement*), dada por

$$LD(\boldsymbol{\omega}) = 2 \left[\ell(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Delta}}, \hat{\gamma}, \hat{\phi}, \hat{\varphi}) - \ell(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \hat{\boldsymbol{\Delta}}_{\boldsymbol{\omega}}, \hat{\gamma}_{\boldsymbol{\omega}}, \hat{\phi}_{\boldsymbol{\omega}}, \hat{\varphi}_{\boldsymbol{\omega}}) \right]. \quad (5.5)$$

O objetivo da técnica de influência local é analisar o comportamento da função $LD(\boldsymbol{\omega})$ em uma vizinhança de $\boldsymbol{\omega}_0$. Para isto, seja \boldsymbol{v} um vetor de direção com norma igual a 1 e considere o gráfico de $LD(\boldsymbol{\omega}_0 + \mathcal{K}\boldsymbol{v})$ contra \mathcal{K} , em que $\mathcal{K} \in \mathbb{R}$. Para cada vetor de direção \boldsymbol{v} , é possível encontrar a curvatura normal em torno de $\mathcal{K} = 0$, denotada por C_v , já que $LD(\boldsymbol{\omega}_0 + \mathcal{K}\boldsymbol{v})$ possui ponto de mínimo local em $\mathcal{K} = 0$. Assim, o interesse está na direção \boldsymbol{v}_{\max} que produz a maior curvatura normal $C_{v_{\max}}$. O gráfico de \boldsymbol{v}_{\max} contra a ordem das observações indica quais observações, sob pequenas perturbações, exercem notável influência sobre $LD(\boldsymbol{\omega})$.

Como observado em Cook (1986), a curvatura normal na direção \boldsymbol{v} é dada por

$$C_v = 2|\boldsymbol{v}^\top \boldsymbol{F}\boldsymbol{v}|, \quad (5.6)$$

em que $\boldsymbol{F} = \boldsymbol{\Lambda}^\top \boldsymbol{M}^{-1} \boldsymbol{\Lambda}$, com \boldsymbol{M} sendo a matriz de informação observada sob o modelo

postulado, dada pela equação (2.20) e $\mathbf{\Lambda}$ a matriz dada por

$$\mathbf{\Lambda} = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi | \boldsymbol{\omega})}{\partial \boldsymbol{\Delta} \partial \boldsymbol{\omega}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi | \boldsymbol{\omega})}{\partial \gamma \partial \boldsymbol{\omega}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi | \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top} \\ \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi | \boldsymbol{\omega})}{\partial \varphi \partial \boldsymbol{\omega}^\top} \end{bmatrix}, \quad (5.7)$$

avaliadas em $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$, $\gamma = \widehat{\gamma}$, $\phi = \widehat{\phi}$, $\varphi = \widehat{\varphi}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. O vetor \mathbf{v}_{\max} corresponde ao autovetor associado ao maior autovalor de \mathbf{F} .

É possível também avaliar a influência da i -ésima observação por meio da curvatura na direção v_i , que corresponde ao vetor de zeros com o valor um na i -ésima posição (Lesaffre & Verbeke, 1998). A curvatura nesta direção, denotada de influência local total do i -ésimo indivíduo, é dada por $C_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi) = 2|\boldsymbol{\Lambda}_i^\top \mathbf{M}^{-1} \boldsymbol{\Lambda}_i|$, com $\boldsymbol{\Lambda}_i^\top$ sendo a i -ésima linha da matriz $\boldsymbol{\Lambda}^\top$, cuja matriz transposta $\boldsymbol{\Lambda}$ é dada pela equação (5.7). A i -ésima observação é considerada influente se $C_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi) \geq 2 \sum_{i=1}^n C_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)/n$ (Verbeke & Molenberghs, 2000).

Em relação aos esquemas de perturbações, adotamos dois: a perturbação de casos e a perturbação na variável explicativa. Na perturbação de casos, a função log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi | \boldsymbol{\omega}) = \sum_i^n \omega_i l_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi), \quad (5.8)$$

com $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$, em que $0 < \omega_i \leq 1$ para $i = 1, \dots, n$, $l_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)$ dada como na equação (2.11) e $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$ o vetor de não perturbação. Neste caso, a

matriz Λ dada em (5.7) é escrita como

$$\Lambda = \begin{bmatrix} \mathbf{S}_1^\top \left(\frac{1}{a(\hat{\phi})} \widehat{\mathbf{W}}^{\frac{1}{2}} \widehat{\mathbf{V}}^{-\frac{1}{2}} \mathbf{D}_{(\mathbf{y}-\hat{\boldsymbol{\mu}})} + \frac{1}{a^*(\hat{\varphi})} \widehat{\mathbf{R}}^{\frac{1}{2}} \widehat{\mathbf{W}}^{\frac{1}{2}} \widehat{\mathbf{V}}^{\frac{1}{2}} \widehat{\mathbf{Q}}^{-\frac{1}{2}} \widehat{\mathbf{J}}_Y \mathbf{D}_{(\mathbf{x}-\hat{\boldsymbol{\alpha}})} \right) \\ \frac{1}{a^*(\hat{\varphi})} \mathbf{S}_2^\top \widehat{\mathbf{R}}^{\frac{1}{2}} \widehat{\mathbf{Q}}^{-\frac{1}{2}} \widehat{\mathbf{J}}_X \widehat{\mathbf{O}} \mathbf{D}_{(\mathbf{x}-\hat{\boldsymbol{\alpha}})} \\ \frac{1}{a^*(\hat{\varphi})} \widehat{\mathbf{J}}_\gamma^\top \widehat{\mathbf{R}}^{\frac{1}{2}} \widehat{\mathbf{Q}}^{-\frac{1}{2}} \mathbf{D}_{(\mathbf{x}-\hat{\boldsymbol{\alpha}})} \\ - \frac{a'(\hat{\phi})}{a(\hat{\phi})^2} (\mathbf{y}^\top \mathbf{D}_{(\hat{\boldsymbol{\theta}})} - \hat{\mathbf{b}} \mathbf{I}) + \hat{\mathbf{c}} \mathbf{I} \\ - \frac{a^{*'}(\hat{\varphi})}{a^*(\hat{\varphi})^2} (\mathbf{x}^\top \mathbf{D}_{(\hat{\boldsymbol{\vartheta}})} - \hat{\mathbf{b}}^* \mathbf{I}) + \hat{\mathbf{c}}^* \mathbf{I} \end{bmatrix}, \quad (5.9)$$

em que $\widehat{\mathbf{W}}$, $\widehat{\mathbf{V}}$, $\widehat{\mathbf{R}}$, $\widehat{\mathbf{Q}}$, $\widehat{\mathbf{J}}_Y$, $\widehat{\mathbf{J}}_X$, $\widehat{\mathbf{O}}$, $\widehat{\mathbf{J}}_\gamma$, $\hat{\boldsymbol{\theta}}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{c}}$, $\hat{\boldsymbol{\vartheta}}$, $\hat{\mathbf{b}}^*$, $\hat{\mathbf{c}}^*$, $\hat{\boldsymbol{\mu}}$ e $\hat{\boldsymbol{\alpha}}$ são, respectivamente, as matrizes e vetores \mathbf{W} , \mathbf{V} , \mathbf{R} , \mathbf{Q} , \mathbf{J}_Y , \mathbf{J}_X , \mathbf{O} , \mathbf{J}_γ , $\boldsymbol{\theta}$, \mathbf{b} , \mathbf{c} , $\boldsymbol{\vartheta}$, \mathbf{b}^* , \mathbf{c}^* , $\boldsymbol{\mu}$ e $\boldsymbol{\alpha}$, dados na Subseção 2.1.1, avaliados em $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\boldsymbol{\Delta} = \hat{\boldsymbol{\Delta}}$, $\gamma = \hat{\gamma}$, $\phi = \hat{\phi}$ e $\varphi = \hat{\varphi}$. Já $\mathbf{D}_{(\mathbf{y}-\hat{\boldsymbol{\mu}})}$, $\mathbf{D}_{(\mathbf{x}-\hat{\boldsymbol{\alpha}})}$, $\mathbf{D}_{(\hat{\boldsymbol{\theta}})}$ e $\mathbf{D}_{(\hat{\boldsymbol{\vartheta}})}$ são matrizes diagonais cujos elementos são formados respectivamente, pelos elementos dos vetores $(\mathbf{y}-\boldsymbol{\mu})$, $(\mathbf{x}-\boldsymbol{\alpha})$, $\boldsymbol{\theta}$ e $\boldsymbol{\vartheta}$, também avaliados em $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\boldsymbol{\Delta} = \hat{\boldsymbol{\Delta}}$, $\gamma = \hat{\gamma}$, $\phi = \hat{\phi}$ e $\varphi = \hat{\varphi}$.

No caso da perturbação na variável explicativa, altera-se uma ou mais colunas da matriz de dados \mathbf{Z} e/ou \mathbf{T} , definidas na Subseção 2.1.1, adicionando a cada variável explicativa de interesse o vetor de perturbação $\boldsymbol{\omega}$ multiplicado por um fator de escala ϖ (Thomas & Cook, 1990). Este fator de escala age como um peso para o vetor de perturbação e geralmente é considerado como sendo a estimativa do desvio padrão da variável explicativa perturbada. Sem perda de generalidade, suponha que a variável explicativa $\mathbf{z}_j = (z_{1j}, z_{2j}, \dots, z_{nj})^\top$ seja perturbada em que \mathbf{z}_j é a j -ésima coluna da matriz de covariáveis \mathbf{Z}_1 , dada na Subseção 2.1.1. Seja ϖ o desvio padrão amostral dos valores observados da covariável \mathbf{z}_j . Segue que os novos valores da covariável são dados por

$$z_{ij}^* = z_{ij} + \varpi \omega_i \quad \text{para } i = 1, \dots, n. \quad (5.10)$$

A função log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid \boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid x_i, y_i, \mathbf{Z}_{i1}^*, \mathbf{Z}_{i2}, \mathbf{T}_{i1}, \mathbf{T}_{i2}), \quad (5.11)$$

em que $\ell_i(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi \mid x_i, y_i, \mathbf{Z}_{i1}^*, \mathbf{Z}_{i2}, \mathbf{T}_{i1}, \mathbf{T}_{i2})$ é dada como na equação (2.11), considerando $\mathbf{Z}_{i1}^* = (1, z_{i1}, \dots, z_{ij-1}, z_{ij}^*, z_{ij+1}, \dots, z_{ik})$, ou seja, a variável explicativa perturbada ao invés da variável explicativa sem perturbação. Note que este tipo de perturbação influencia a função log-verossimilhança perturbada apenas através de θ_i , que está relacionado à média da variável resposta discreta Y_i . Além disso, o vetor de não perturbação é dado por $\boldsymbol{\omega}_0 = \mathbf{0}_{(n \times 1)}$.

A matriz \mathbf{A} é calculada como indicada na equação (5.7), considerando a função log-verossimilhança perturbada dada na equação (5.11).

5.3 Influência global ou total

Também utilizada no diagnóstico de influência, a técnica de influência global ou total tem como finalidade analisar a sensibilidade do modelo em relação a cada uma das observações, de uma maneira mais geral. Esta análise é feita utilizando a técnica de deleção de casos (Cook, 1977), em que cada uma das observação é deletada individualmente da amostra e então, ajustando o modelo em estudo com e sem a observação deletada, avalia-se o impacto da retirada da observação nas estimativas dos parâmetros do modelo. Este impacto é quantificado em medidas. Neste trabalho, considera-se duas: a distância de Cook generalizada e o afastamento das verossimilhanças.

Considerando a classe de modelos bivariados com respostas mistas apresentada na Seção 2.1, seja $\ell_{[i]}(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)$ a função log-verossimilhança dada na equação (2.11), desconsiderando a i -ésima observação e $\hat{\boldsymbol{\beta}}_{[i]}$, $\hat{\boldsymbol{\Delta}}_{[i]}$, $\hat{\gamma}_{[i]}$, $\hat{\phi}_{[i]}$ e $\hat{\varphi}_{[i]}$ os respectivos estimadores de máxima verossimilhança de $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, γ , ϕ e φ , sob $\ell_{[i]}(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)$. A distância

de Cook generalizada, para a i -ésima observação, é definida por

$$DCG_i = \left[(\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\boldsymbol{\Delta}}_{[i]}^\top, \widehat{\gamma}_{[i]}, \widehat{\phi}_{[i]}, \widehat{\varphi}_{[i]}) - (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\Delta}}^\top, \widehat{\gamma}, \widehat{\phi}, \widehat{\varphi}) \right] \widehat{\boldsymbol{M}} \left[(\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\boldsymbol{\Delta}}_{[i]}^\top, \widehat{\gamma}_{[i]}, \widehat{\phi}_{[i]}, \widehat{\varphi}_{[i]}) - (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\Delta}}^\top, \widehat{\gamma}, \widehat{\phi}, \widehat{\varphi}) \right]^\top,$$

em que $\widehat{\boldsymbol{M}}$ é a matriz de informação observada, dada pela equação (2.20), avaliada em $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$, $\gamma = \widehat{\gamma}$, $\phi = \widehat{\phi}$ e $\varphi = \widehat{\varphi}$, os estimadores de máxima verossimilhança sob $\ell(\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \phi, \varphi)$.

Outra medida, já utilizada na Seção 5.2, é o afastamento das verossimilhanças, definida para a i -ésima observação como

$$DV_i = 2\{\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Delta}}, \widehat{\gamma}, \widehat{\phi}, \widehat{\varphi}) - \ell(\widehat{\boldsymbol{\beta}}_{[i]}, \widehat{\boldsymbol{\Delta}}_{[i]}, \widehat{\gamma}_{[i]}, \widehat{\phi}_{[i]}, \widehat{\varphi}_{[i]})\}.$$

Assim, quanto menores as medidas DCG_i e DV_i obtidas, menos influente é a i -ésima observação no ajuste do modelo.

Capítulo 6

Estudo de simulação

Neste capítulo, realizamos um estudo com dados simulados considerando o modelo Bernoulli-exponencial, apresentado na Seção 4.3, e o modelo Poisson-normal semiparamétrico, dado na Seção 4.4. No caso do modelo Bernoulli-exponencial o objetivo deste estudo é analisar o comportamento dos estimadores de máxima verossimilhança dos parâmetros do modelo em relação ao viés, à raiz quadrada do erro quadrático médio ($\sqrt{\text{EQM}}$), ao desvio padrão das estimativas (DP), à média dos erros padrão assintóticos (EPa) e à probabilidade de cobertura dos intervalos de confiança assintóticos (PCa) em alguns cenários predeterminados, e também analisar a performance dos resíduos apresentados na Seção 5.1. No caso do modelo Poisson-normal semiparamétrico, o estudo é dividido em duas partes. Na primeira, é considerado um cenário em que os preditores são compostos apenas por termos lineares e não lineares paramétricos e assim, o objetivo também é analisar o comportamento dos estimadores de máxima verossimilhança dos parâmetros do modelo em relação à viés, $\sqrt{\text{EQM}}$, DP, EPa e PCa e a performance dos resíduos, como realizado para o modelo Bernoulli-exponencial. Na segunda parte, três cenários são propostos em que os preditores são compostos por termos paramétricos e não paramétricos. Nestes casos, o objetivo é analisar o ajuste dos termos não paramétricos (funções suaves dos dados) e o comportamento dos estimadores de

máxima verossimilhança referente à parte paramétrica do modelo, em relação à viés, $\sqrt{\text{EQM}}$ e DP, além de analisar a performance dos resíduos apresentados na Seção 5.1.

O viés simulado é calculado fazendo a diferença entre a média das estimativas e o verdadeiro valor do parâmetro. A $\sqrt{\text{EQM}}$ simulada é dada pela expressão

$$\sqrt{\text{EQM}} = \left[\sum_{i=1}^N (\hat{b}_i - b)^2 / N \right]^{1/2},$$

em que N é o número de amostras simuladas, b é o valor verdadeiro do parâmetro e \hat{b}_i é a estimativa de máxima verossimilhança de b , na i -ésima amostra simulada. O desvio padrão das estimativas dos parâmetros do modelo é calculado pela expressão

$$\text{DP} = \left[\sum_{i=1}^N \frac{(\hat{b}_i - \bar{\hat{b}})^2}{N - 1} \right]^{1/2},$$

em que $\bar{\hat{b}}$ é a média da amostra $\hat{b}_1, \dots, \hat{b}_N$. Os erros padrão assintóticos das estimativas dos parâmetros, em cada amostra simulada, são obtidas utilizando a estimativa da matriz de covariâncias assintótica, dada pela equação (2.20). Por fim, a estimativa da probabilidade de cobertura dos intervalos de confiança assintóticos de um parâmetro é encontrada calculando os intervalos de confiança assintóticos das estimativas em cada amostra gerada e então verificando a proporção de intervalos que contem o verdadeiro valor do parâmetro. A probabilidade de cobertura nominal considerada é de 95%.

Diferentes cenários e tamanhos amostrais são considerados, em que procuramos representar um grande número de situações. Para todas as situações tratadas neste capítulo, na estimação pelo método de máxima verossimilhança é adotado que os algoritmos Newton-Raphson e score de Fisher, utilizados para encontrar as soluções numéricas das equações de estimação, se encerram quando a diferença entre as estimativas dos parâmetros do passo atual e do passo anterior é menor do que 10^{-2} . Gráficos dos resíduos apresentados nas Subseções 5.1.1, 5.1.2 e 5.1.3 *versus* as obser-

vações, para algumas das amostras geradas, são mostradas. Toda parte computacional é desenvolvida em linguagem R (R Core Team, 2015).

6.1 Modelo Bernoulli-exponencial

Para o modelo bivariado Bernoulli-exponencial, quatro cenários são apresentados. Nos três primeiros cenários, são consideradas amostras de tamanho 50, 100, 200 e 500. Para cada configuração e tamanho amostral, 10000 conjuntos de dados são gerados. Os valores do viés simulado, da $\sqrt{\text{EQM}}$ simulado, do DP, do EPa médio e da PCa estimada, de cada parâmetro, são apresentados. Por fim, no quarto cenário é ilustrado técnicas de resíduos na detecção da inadequabilidade do componente sistemático proposto.

Cenário 1: São consideradas três covariáveis dependentes para prever cada uma das respostas Y_i e X_i . Os valores das covariáveis são gerados de uma distribuição uniforme discreta com parâmetro 10, considerando uma correlação próxima de 0,5 entre os valores gerados. Para isto, utiliza-se a função *ordsample* do pacote *GenOrd* no software R. Os valores dos coeficientes de regressão adotados são $\beta = (0,50; 0,01; 0,05; 0,10)$. Para Δ , dois vetores de valores são considerados $\Delta^1 = (0,10; -0,20; 0,03; 0,05)$ e $\Delta^2 = (1,0; 0,5; 0,2; -0,1)$. Nos três vetores apresentados, o primeiro valor corresponde ao termo do intercepto. Para estas escolhas, o valor médio de Y_i é próximo de 0,7 enquanto que o valor médio de X_i é 0,6028 com uma pequena variabilidade e 329,80 com uma grande variabilidade, para Δ^1 e Δ^2 , respectivamente. Os resultados são apresentados nas Tabelas 6.1 e 6.2. Os comportamentos dos resíduos para algumas das amostras geradas são ilustrados nas Figuras 6.1 e 6.2.

Tabela 6.1: Viés, $\sqrt{\text{EQM}}$, DP, EPa médio e PCa - Cenário 1 com $\beta = (0, 50; 0, 01; 0, 05; 0, 10)$ e $\Delta^1 = (0, 10; -0, 20; 0, 03; 0, 05)$.

Viés simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	-0,0407	-0,0067	0,0100	0,0014	-0,0615	0,0290	-0,0044	-0,0063
100	0,0203	0,0003	0,0040	0,0056	-0,0285	0,0011	-0,0005	0,0002
200	0,0055	-0,0010	0,0027	0,0032	-0,0057	-0,0006	0,0002	-0,0005
500	0,0023	0,0005	0,0004	0,0009	-0,0030	-0,0001	0,0003	-0,0003
$\sqrt{\text{EQM}}$ simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,8739	0,2353	0,3555	0,2049	0,3807	0,1011	0,0616	0,0578
100	0,6043	0,0964	0,1086	0,0986	0,2633	0,0463	0,0441	0,0432
200	0,4215	0,0653	0,0684	0,0645	0,1854	0,0301	0,0299	0,0284
500	0,2524	0,0403	0,0408	0,0405	0,1173	0,0181	0,0189	0,0180
DP								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,8730	0,2352	0,3554	0,2049	0,3757	0,0969	0,0614	0,0575
100	0,6040	0,0964	0,1086	0,0985	0,2618	0,0463	0,0441	0,0433
200	0,4215	0,0653	0,0684	0,0644	0,1854	0,0301	0,0299	0,0284
500	0,2524	0,0403	0,0408	0,0405	0,1173	0,0181	0,0189	0,0180
EPa médio								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,6962	0,1313	0,1423	0,1185	0,3436	0,0628	0,0550	0,0507
100	0,5784	0,0917	0,1041	0,0930	0,2612	0,0462	0,0438	0,0426
200	0,4138	0,0636	0,0667	0,0631	0,1877	0,0302	0,0299	0,0286
500	0,2510	0,0399	0,0406	0,0400	0,1177	0,0182	0,0189	0,0182
PCa estimada (%)								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	81,89	81,33	81,61	81,82	81,26	81,72	81,55	81,88
100	94,87	94,77	94,79	94,82	94,74	94,94	94,78	94,63
200	94,96	94,55	95,10	94,98	95,38	95,04	95,07	95,19
500	95,00	94,78	94,99	94,90	95,14	95,09	94,90	95,14

Tabela 6.2: Viés, $\sqrt{\text{EQM}}$, DP, EPa médio e PCa - Cenário 1 com $\beta = (0, 50; 0, 01; 0, 05; 0, 10)$ e $\Delta^2 = (1, 0; 0, 5; 0, 2; -0, 1)$.

Viés simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,0118	-0,0025	0,0151	0,0212	-0,0463	0,0005	-0,0009	0,0007
100	0,0048	0,0017	0,0030	0,0068	-0,0259	0,0001	0,0007	-0,0002
200	0,0098	-0,0013	0,0023	0,0033	-0,0114	-0,0003	-0,0001	0,0003
500	0,0023	0,0001	0,0005	0,0014	-0,0032	0,0002	-0,0002	-0,0002
$\sqrt{\text{EQM}}$ simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,8799	0,1718	0,2034	0,1672	0,4105	0,0734	0,0646	0,0601
100	0,6006	0,0955	0,1085	0,0977	0,2641	0,0467	0,0441	0,0425
200	0,4239	0,0653	0,0681	0,0646	0,1891	0,0304	0,0299	0,0289
500	0,2509	0,0399	0,0411	0,0401	0,1180	0,0180	0,0190	0,0185
DP								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,8798	0,1718	0,2028	0,1658	0,4079	0,0734	0,0646	0,0601
100	0,6006	0,0955	0,1085	0,0975	0,2628	0,0467	0,0441	0,0425
200	0,4238	0,0653	0,0681	0,0646	0,1888	0,0304	0,0299	0,0289
500	0,2509	0,0399	0,0411	0,0401	0,1180	0,0180	0,0190	0,0185
EPa médio								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	0,8074	0,1527	0,1654	0,1387	0,3993	0,0730	0,0640	0,0589
100	0,5770	0,0915	0,1039	0,0928	0,2613	0,0462	0,0438	0,0426
200	0,4139	0,0636	0,0667	0,0631	0,1877	0,0302	0,0298	0,0286
500	0,2511	0,0399	0,0406	0,0400	0,1178	0,0182	0,0189	0,0182
PCa estimada (%)								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	95,08	94,68	94,65	95,55	94,31	94,92	95,06	94,74
100	94,91	94,82	94,98	95,24	94,48	94,53	94,82	95,30
200	94,89	95,19	94,67	94,97	94,89	94,97	95,04	94,80
500	95,12	95,23	95,03	95,06	94,92	94,96	94,93	94,43

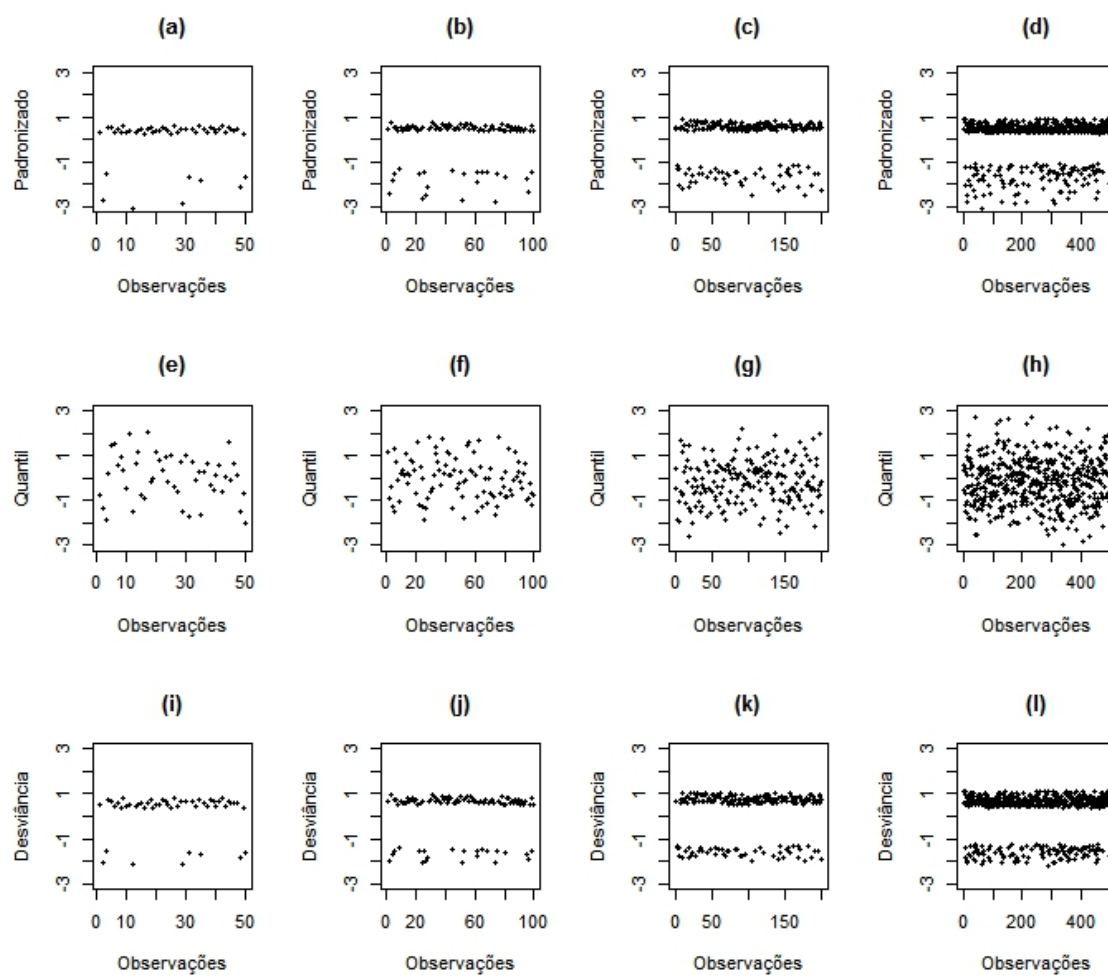


Figura 6.1: Resíduos *versus* observações - Modelo associado à Y_i - Cenário 1 com (a), (e) e (i): $n = 50$ e Δ^1 ; (b), (f) e (j): $n = 100$ e Δ^2 ; (c), (g) e (k): $n = 200$ e Δ^1 ; (d), (h) e (l): $n = 500$ e Δ^2 .

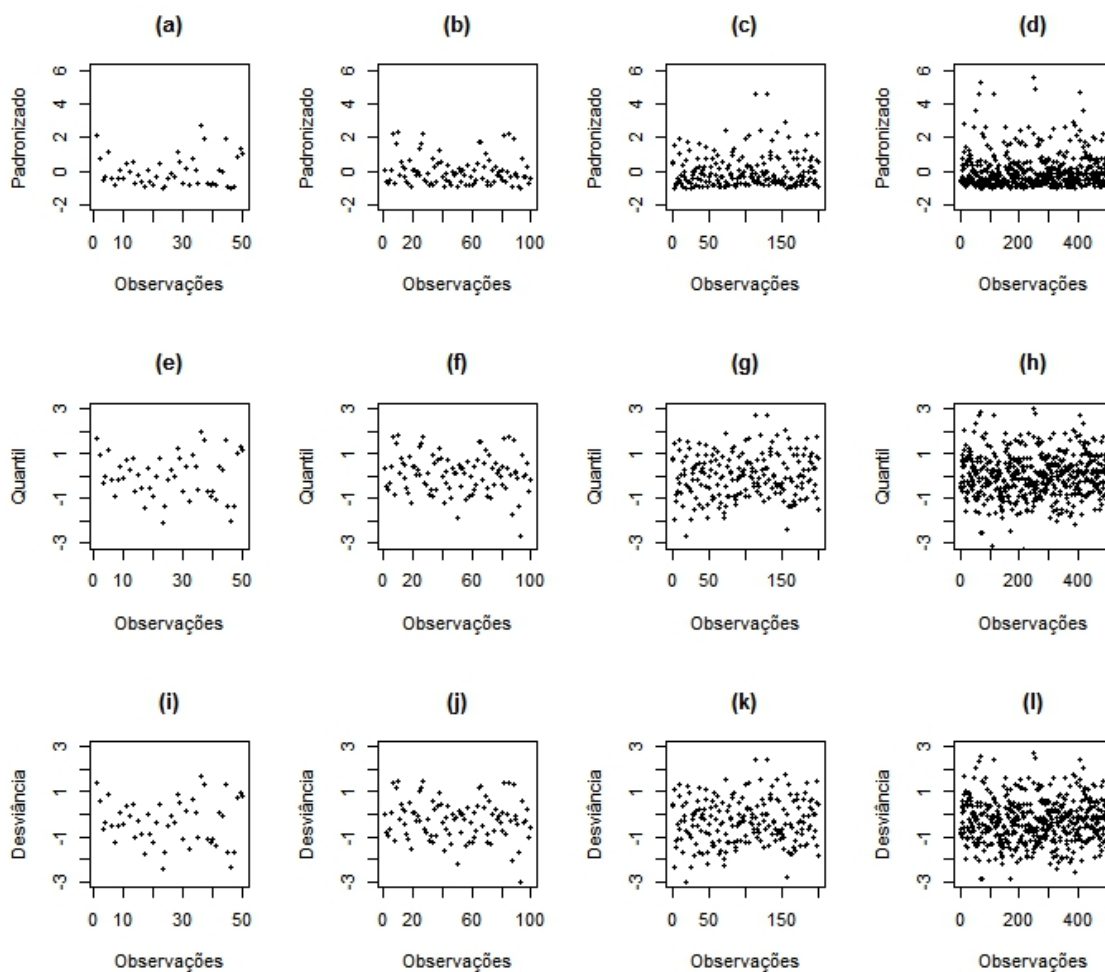


Figura 6.2: Resíduos *versus* observações - Modelo associado à $X_i | Y_i = y_i$ - Cenário 1 com (a), (e) e (i): $n = 50$ e Δ^1 ; (b), (f) e (j): $n = 100$ e Δ^2 ; (c), (g) e (k): $n = 200$ e Δ^1 ; (d), (h) e (l): $n = 500$ e Δ^2 .

Cenário 2: São consideradas novamente três covariáveis dependentes para prever cada uma das respostas Y_i e X_i . Porém, os valores das covariáveis são gerados de uma distribuição contínua normal multivariada com vetor de médias iguais a 4, variâncias iguais a 5 e covariâncias iguais a 2,5 resultando em uma correlação próxima de 0,5 entre os valores gerados. Os valores dos coeficientes de regressão adotados são $\beta = (0, 10; 0, 01; 0, 05; 0, 20)$. Para Δ , são considerados $\Delta^1 = (0, 30; -0, 20; 0, 03; 0, 05)$ e $\Delta^2 = (0, 9; 0, 5; -0, 1; 0, 5)$. O primeiro valor em cada vetor corresponde ao termo do

intercepto. Para estas escolhas, o valor médio de Y_i é próximo de 0,7 enquanto que o valor médio de X_i é 0,8435 com uma pequena variabilidade e 492,375 com uma grande variabilidade entre os valores gerados, para Δ^1 e Δ^2 , respectivamente. Os resultados são apresentados nas Tabelas 6.3 e 6.4. Os comportamentos dos resíduos neste cenário são análogos aos obtidos no cenário 1 e portanto, os gráficos não serão apresentados.

Tabela 6.3: Viés, $\sqrt{\text{EQM}}$, DP, EPa médio e PCa - Cenário 2 com $\beta = (0, 10; 0, 01; 0, 05; 0, 20)$ e $\Delta^1 = (0, 30; -0, 20; 0, 03; 0, 05)$.

Viés simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	-0,0753	0,0135	0,0051	0,0343	-0,0325	-0,0002	-0,0010	-0,0018
100	-0,0296	0,0014	0,0018	0,0167	-0,0172	-0,0015	0,0003	-0,0002
200	-0,0123	0,0012	0,0007	0,0074	-0,0091	-0,0010	0,0001	0,0004
500	-0,0078	0,0001	0,0009	0,0032	-0,0038	-0,0001	0,0003	-0,0002
$\sqrt{\text{EQM}}$ simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	1,1366	0,2439	0,2115	0,2330	0,4339	0,0889	0,0960	0,0782
100	0,6078	0,1434	0,1399	0,1352	0,2786	0,0583	0,0585	0,0592
200	0,4172	0,0909	0,0913	0,0923	0,1855	0,0412	0,0409	0,0438
500	0,2442	0,0561	0,0543	0,0573	0,1171	0,0251	0,0262	0,0260
DP								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	1,1341	0,2436	0,2114	0,2305	0,4327	0,0889	0,0960	0,0782
100	0,6071	0,1434	0,1399	0,1341	0,2781	0,0583	0,0585	0,0592
200	0,4170	0,0909	0,0913	0,0920	0,1853	0,0412	0,0409	0,0438
500	0,2441	0,0561	0,0543	0,0572	0,1171	0,0251	0,0262	0,0260
EPa médio								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	1,0078	0,2102	0,1866	0,1926	0,4316	0,0884	0,0948	0,0782
100	0,5824	0,1347	0,1317	0,1274	0,2754	0,0582	0,0580	0,0590
200	0,4072	0,0883	0,0883	0,0887	0,1858	0,0412	0,0406	0,0438
500	0,2405	0,0559	0,0538	0,0559	0,1163	0,0252	0,0262	0,0261
PCa estimada (%)								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	94,02	93,98	94,26	94,76	95,03	94,95	94,44	95,29
100	94,51	94,44	94,58	94,68	94,82	95,01	95,08	94,88
200	94,62	94,73	94,73	94,53	94,98	95,02	94,62	95,21
500	94,82	95,29	94,84	94,26	94,85	95,14	95,12	95,21

Tabela 6.4: Viés, $\sqrt{\text{EQM}}$, DP, EPa médio e PCa - Cenário 2 com $\beta = (0, 10; 0, 01; 0, 05; 0, 20)$ e $\Delta^2 = (0, 9; 0, 5; -0, 1; 0, 5)$.

Viés simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	-0,0674	0,0134	0,0050	0,0298	-0,0233	-0,0006	-0,0010	-0,0035
100	-0,0094	0,0013	0,0031	0,0124	-0,0183	-0,0013	0,0001	0,0001
200	-0,0085	-0,0004	0,0038	0,0053	-0,0091	-0,0004	0,0001	-0,0004
500	0,0001	0,0005	0,0003	0,0017	-0,0034	-0,0003	0,0001	-0,0001
$\sqrt{\text{EQM}}$ simulado								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	1,1165	0,2357	0,2084	0,2202	0,4333	0,0882	0,0959	0,0782
100	0,6064	0,1423	0,1387	0,1346	0,2760	0,0577	0,0580	0,0593
200	0,4127	0,0910	0,0910	0,0910	0,1857	0,0415	0,0408	0,0437
500	0,2423	0,0566	0,0544	0,0564	0,1168	0,0251	0,0265	0,0264
DP								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	1,1145	0,2353	0,2084	0,2182	0,4327	0,0882	0,0959	0,0781
100	0,6063	0,1423	0,1387	0,1340	0,2754	0,0577	0,0580	0,0593
200	0,4126	0,0910	0,0909	0,0909	0,1855	0,0415	0,0408	0,0437
500	0,2423	0,0566	0,0544	0,0563	0,1168	0,0251	0,0265	0,0264
EPa médio								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	1,0041	0,2090	0,1859	0,1912	0,4317	0,0884	0,0948	0,0781
100	0,5841	0,1351	0,1321	0,1277	0,2753	0,0582	0,0580	0,0590
200	0,4070	0,0882	0,0883	0,0887	0,1859	0,0412	0,0405	0,0438
500	0,2406	0,0559	0,0538	0,0559	0,1163	0,0252	0,0262	0,0261
PCa estimada (%)								
n	β_0	β_1	β_2	β_3	δ_0	δ_1	δ_2	δ_3
50	94,29	94,55	94,70	94,96	95,18	94,84	94,76	94,80
100	94,61	94,56	94,75	94,76	94,96	95,19	94,93	94,91
200	94,95	94,70	94,58	94,95	94,97	95,05	94,94	95,10
500	95,08	94,80	94,77	94,96	94,89	95,16	94,65	94,78

Cenário 3: Procurando reproduzir as configurações do conjunto de dados apresentado na Seção 7, considera-se que quatro covariáveis binárias estão disponíveis para predizer Y_i e X_i . Neste caso, gera-se inicialmente valores de uma distribuição uniforme discreta com parâmetro 5, representando 5 categorias distintas e então, atribui-se os valores 0 ou 1 às variáveis binárias, de acordo com as categorias geradas. Os valores adotados para os coeficientes de regressão são $\beta = (0, 05; 0, 07; 0, 01; 0, 05; 0, 20)$. Para Δ , adota-se

dois vetores: $\Delta^1 = (0, 01; 0, 05; -0, 20; 0, 03; 0, 05)$ e $\Delta^2 = (4, 0; 1, 5; -1, 0; 2, 9; 5, 0)$ O primeiro valor em cada vetor corresponde ao termo do intercepto. Para estas escolhas, o valor médio de Y_i é próximo de 0,53 enquanto que o valor médio de X_i é 0,9764 com uma pequena variabilidade e 1478,74 com uma grande variabilidade entre os valores gerados, para Δ^1 e Δ^2 , respectivamente. Os resultados são apresentados nas Tabelas 6.5 e 6.6 enquanto que os comportamentos dos resíduos para algumas das amostras geradas são apresentados nas Figuras 6.3 e 6.4.

Tabela 6.5: Viés, \sqrt{EQM} , DP, EPa médio e PCa - Cenário 3 com $\beta = (0, 05; 0, 07; 0, 01; 0, 05; 0, 20)$ e $\Delta^1 = (0, 01; 0, 05; -0, 20; 0, 03; 0, 05)$.

Viés simulado										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	-0,0190	0,0100	-0,0133	-0,0124	0,0201	-0,0496	-0,0174	0,0036	-0,0435	-0,0230
100	-0,0150	0,0057	0,0085	0,0066	0,0203	-0,0276	-0,0087	-0,0027	-0,0069	-0,0037
200	-0,0123	0,0050	0,0121	0,0100	0,0150	-0,0186	0,0038	0,0077	0,0051	0,0020
500	-0,0004	0,0007	-0,0047	-0,0008	0,0008	-0,0067	0,0014	0,0037	0,0020	-0,0008
\sqrt{EQM} simulado										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	0,6105	0,7647	0,9037	0,9039	0,8457	0,3164	0,4635	0,4039	0,5362	0,4663
100	0,4528	0,5738	0,5938	0,6269	0,5570	0,2186	0,3320	0,3016	0,3158	0,3149
200	0,2861	0,3889	0,3928	0,3855	0,3693	0,1773	0,2383	0,2255	0,2354	0,2328
500	0,1749	0,2403	0,2325	0,2369	0,2332	0,1111	0,1474	0,1433	0,1443	0,1492
DP										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	0,6103	0,7647	0,9036	0,9038	0,8455	0,3125	0,4632	0,4039	0,5344	0,4657
100	0,4526	0,5738	0,5938	0,6269	0,5567	0,2168	0,3319	0,3016	0,3158	0,3149
200	0,2859	0,3889	0,3926	0,3854	0,3690	0,1764	0,2383	0,2254	0,2354	0,2328
500	0,1750	0,2403	0,2325	0,2369	0,2332	0,1109	0,1474	0,1433	0,1443	0,1492
EPa médio										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	0,5762	0,7267	0,8556	0,8544	0,8143	0,3066	0,4545	0,3966	0,5134	0,4670
100	0,4360	0,5576	0,5766	0,6019	0,5384	0,2146	0,3263	0,2988	0,3104	0,3100
200	0,2798	0,3817	0,3807	0,3787	0,3643	0,1753	0,2369	0,2227	0,2340	0,2326
500	0,1737	0,2381	0,2304	0,2351	0,2328	0,1116	0,1479	0,1440	0,1446	0,1492
PCa estimada (%)										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	96,41	95,36	95,71	95,49	95,87	94,20	94,45	94,68	93,81	94,98
100	95,58	95,20	94,93	94,92	95,00	94,55	94,55	94,64	94,45	94,73
200	94,99	95,07	94,62	94,79	95,02	94,58	94,63	94,89	94,88	94,87
500	95,05	95,05	95,05	94,99	95,04	94,81	95,19	95,14	95,02	94,90

Tabela 6.6: Viés, $\sqrt{\text{EQM}}$, DP, EPa médio e PCa - Cenário 3 com $\beta = (0, 05; 0, 07; 0, 01; 0, 05; 0, 20)$ e $\Delta^2 = (4, 0; 1, 5; -1, 0; 2, 9; 5, 0)$.

Viés simulado										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	-0,0169	0,0087	-0,0137	-0,0013	0,0137	-0,0531	-0,0069	0,0070	-0,0402	-0,0204
100	-0,0099	0,0123	0,0036	0,0005	0,0175	-0,0289	-0,0081	-0,0022	-0,0076	-0,0057
200	-0,0057	0,0059	0,0033	-0,0008	0,0064	-0,0197	0,0023	0,0060	0,0067	0,0032
500	-0,0029	0,0038	0,0029	0,0012	0,0069	-0,0071	0,0015	0,0014	0,0005	0,0005
$\sqrt{\text{EQM}}$ simulado										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	0,6194	0,7670	0,8919	0,8964	0,8568	0,3154	0,4648	0,4041	0,5345	0,4808
100	0,4588	0,5795	0,6062	0,6268	0,5627	0,2187	0,3291	0,3028	0,3169	0,3144
200	0,2827	0,3869	0,3875	0,3822	0,3684	0,1758	0,2385	0,2218	0,2361	0,2325
500	0,1731	0,2348	0,2291	0,2354	0,2347	0,1128	0,1487	0,1449	0,1460	0,1507
DP										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	0,6192	0,7670	0,8919	0,8965	0,8568	0,3109	0,4647	0,4040	0,5330	0,4804
100	0,4587	0,5794	0,6062	0,6268	0,5625	0,2168	0,3290	0,3028	0,3168	0,3143
200	0,2827	0,3869	0,3875	0,3822	0,3684	0,1747	0,2385	0,2217	0,2360	0,2325
500	0,1731	0,2348	0,2291	0,2354	0,2346	0,1126	0,1487	0,1449	0,1460	0,1507
EPa médio										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	0,5773	0,7275	0,8524	0,8554	0,8145	0,3065	0,4544	0,3966	0,5133	0,4667
100	0,4371	0,5590	0,5781	0,6028	0,5396	0,2145	0,3263	0,2988	0,3103	0,3100
200	0,2798	0,3817	0,3808	0,3787	0,3644	0,1752	0,2369	0,2227	0,2340	0,2325
500	0,1735	0,2380	0,2303	0,2350	0,2328	0,1116	0,1479	0,1440	0,1446	0,1492
PCa estimada (%)										
n	β_0	β_1	β_2	β_3	β_4	δ_0	δ_1	δ_2	δ_3	δ_4
50	96,28	95,31	95,98	95,84	95,74	94,18	94,50	94,41	93,89	94,27
100	95,30	94,96	94,95	95,02	94,85	94,57	94,69	94,69	94,26	94,34
200	95,08	94,86	94,78	94,96	95,03	94,82	95,02	95,05	94,53	94,77
500	95,42	95,42	95,09	94,77	94,79	94,78	94,80	94,87	94,83	94,93

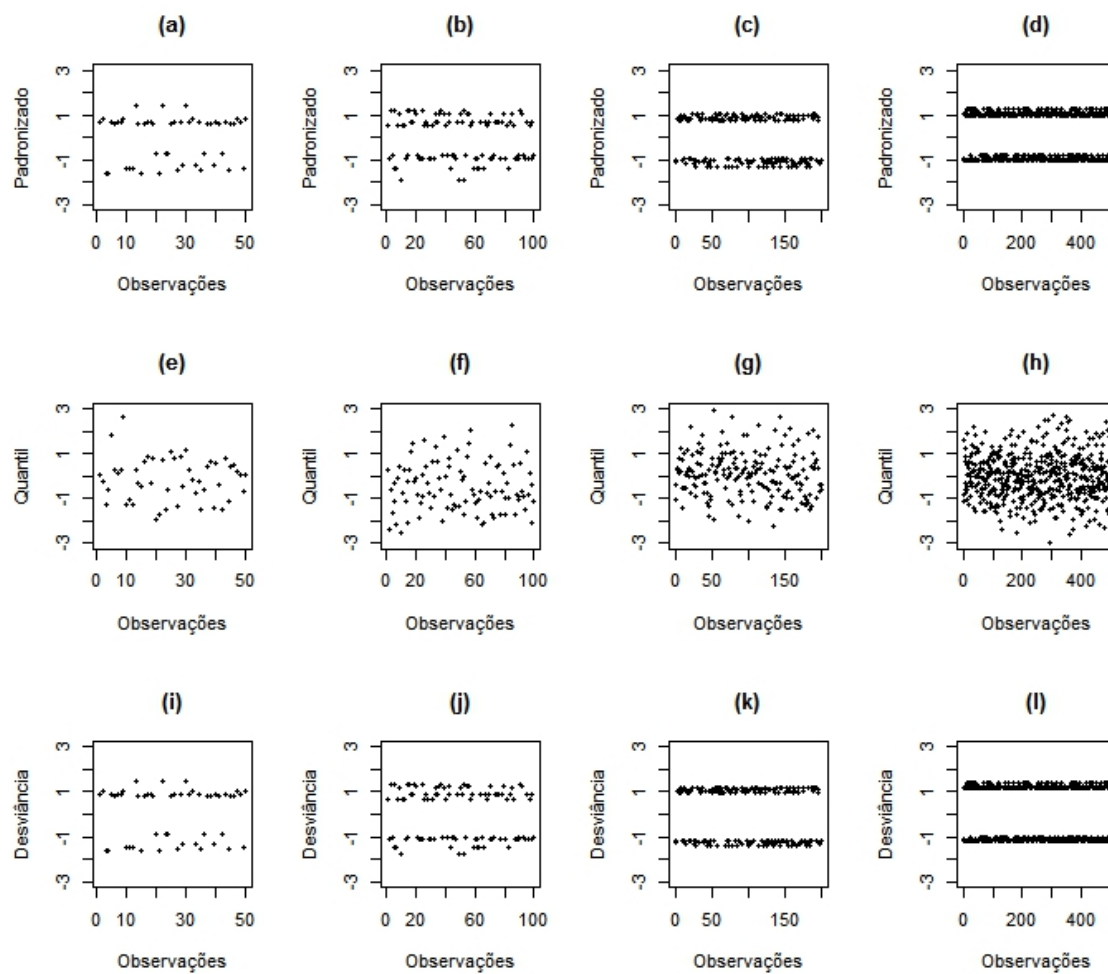


Figura 6.3: Resíduos *versus* observações - Modelo associado à Y_i - Cenário 3 com (a), (e) e (i): $n = 50$ e Δ^1 ; (b), (f) e (j): $n = 100$ e Δ^1 ; (c), (g) e (k): $n = 200$ e Δ^2 ; (d), (h) e (l): $n = 500$ e Δ^2 .

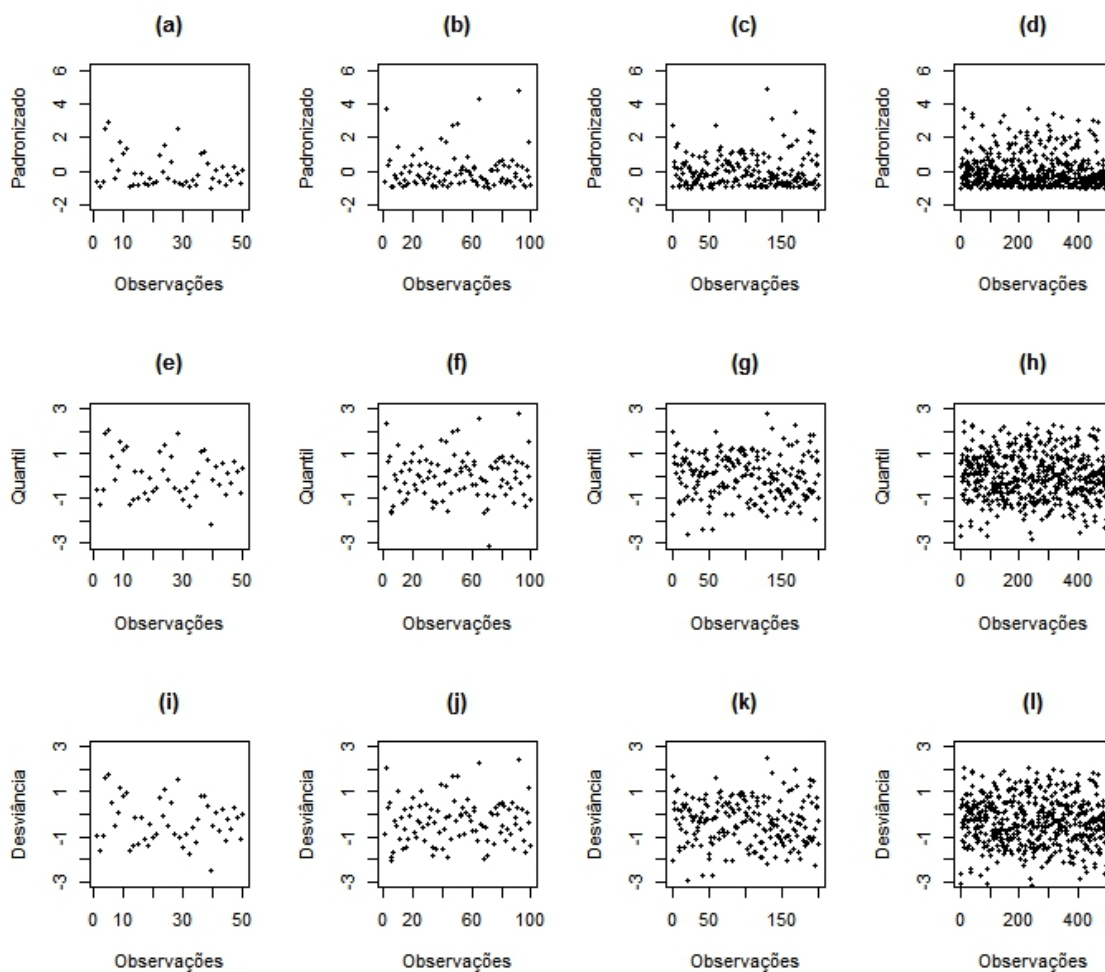


Figura 6.4: Resíduos *versus* observações - Modelo associado à $X_i | Y_i = y_i$ - Cenário 3 com (a), (e) e (i): $n = 50$ e Δ^1 ; (b), (f) e (j): $n = 100$ e Δ^1 ; (c), (g) e (k): $n = 200$ e Δ^2 ; (d), (h) e (l): $n = 500$ e Δ^2 .

Analisando as Tabelas 6.1, 6.2, 6.3, 6.4, 6.5 e 6.6, observa-se que os valores do viés simulado são relativamente pequenos mesmo para amostras de tamanho 50. Além disso, estes valores tendem a diminuir com o aumento do tamanho amostral, nos três cenários analisados, como esperado. Em relação à $\sqrt{\text{EQM}}$ e ao DP, nota-se uma semelhança nos valores obtidos em uma mesma configuração (cenário e tamanho amostral). Embora próximos, os valores do EPA médio nas amostras de tamanhos menores são diferentes em geral, mas tendem a se aproximar dos valores da $\sqrt{\text{EQM}}$ e do DP a medida que

o tamanho da amostra aumenta. Em todos os cenários analisados é possível ver que os valores da $\sqrt{\text{EQM}}$, do DP e do EPa médio diminuem com o aumento amostral, como esperado. Em relação as estimativas da PCa, nota-se que os valores obtidos são próximos em geral, do valor da cobertura nominal considerado (95%), exceto no Cenário 1 com $\Delta = \Delta^1$ e $n = 50$, em que os valores obtidos estão em torno de 82%. Desta forma, o comportamento dos estimadores de máxima verossimilhança dos parâmetros do modelo em estudo se mostra satisfatório, nas condições analisadas.

Em relação ao modelo associado à Y_i , analisando as Figuras 6.1 e 6.3, nota-se que o comportamento dos resíduos padronizado e componentes da desviância é similar em todos os cenários analisados, com valores obtidos concentrados em duas faixas, coerente em dados com resposta binária (Paula, 2004). Entretanto, nota-se uma assimetria nas distribuições destes resíduos, mesmo para grandes amostras, nas configurações dos cenários 1 e 2. Nestes casos, os valores médios obtidos em cada cenário e tamanho amostral são próximos a 0, em que os valores estão concentrados nas faixas entre -3 e -1 e entre 0 e 2, em geral. Por outro lado, no cenário 3, em que as covariáveis são binárias, as distribuições dos resíduos padrão e componentes da desviância tendem a ser simétricas a medida que o tamanho amostral aumenta, em que os valores obtidos estão nas faixas entre -2 e 0 e entre 0 e 2. Já o comportamento do resíduo quantil é coerente com a distribuição teórica em todos os cenários analisados, com valores próximos de 0 e concentrados entre -3 e 3. O teste de normalidade Kolmogorov-Smirnov é aplicado aos resíduos quantil, comprovando a normalidade destes, para a maioria das amostras geradas. Em relação ao modelo associado à $X_i | Y_i = i$, nota-se analisando as Figuras 6.2 e 6.4 uma assimetria na distribuição do resíduo padronizado, em todos os cenários e tamanho amostral analisados. Neste caso, os valores obtidos variam entre -2 e 6 com valores concentrados em torno de 0. O resíduo quantil tem comportamento coerente com sua distribuição teórica, com valores em torno de -3 e 3 e concentrados em 0. Assim como para os resíduos de Y_i , o teste de normalidade Kolmogorov-Smirnov

é aplicada, comprovando a normalidade dos resíduos na maioria das amostras geradas. Comportamento similar é obtido nos resíduos componentes da desviância em que os valores estão concentrados em 0 e distribuídos em uma faixa de -3 e 3.

Cenário 4: Com o objetivo de analisar a adequabilidade da forma do preditor associado à variável resposta contínua e/ou discreta, em relação ao conjunto de variáveis explicativas disponíveis, é gerado um conjunto de dados em que um termo quadrático é inserido no preditor associado a variável resposta contínua X_i . Desta forma, ajusta-se o modelo considerando o preditor sem termos quadráticos e então, analisa-se o comportamento dos resíduos. Os dados são gerados como no cenário 2, porém com $\beta = (0,001; 0,01; 0,005; 0,06)$ e $\Delta = (0,03; -0,02; 0,003; 0,7; 0,1)$, em que 0,1 corresponde ao coeficiente δ_4 associado ao termo t_{i3}^2 , adicionado ao preditor. O comportamento dos resíduos é mostrado na Figura 6.5, em que curvas suavizadas utilizando a técnica Lowess (Cleveland, 1979), é inserida para detectar a tendência dos valores obtidos.

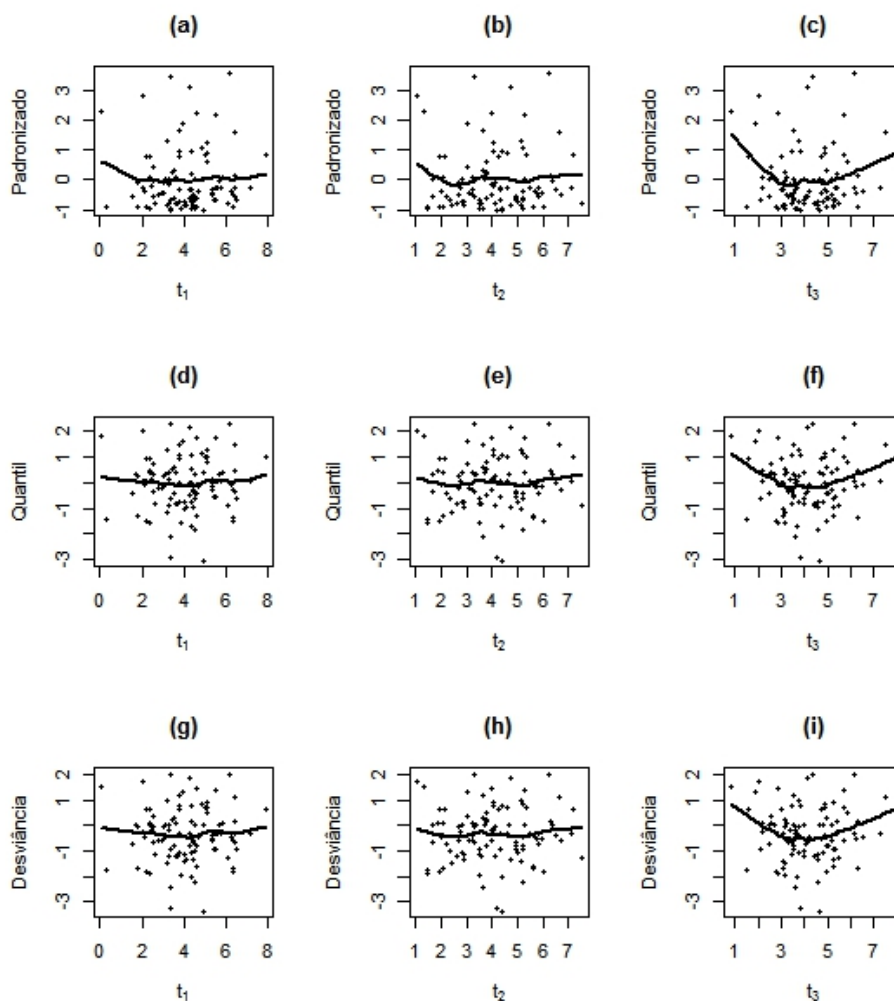


Figura 6.5: (a), (b) e (c): Resíduos padronizado *versus* variáveis explicativas t_1 , t_2 e t_3 , respectivamente; (d), (e) e (f): Resíduos quantil *versus* variáveis explicativas t_1 , t_2 e t_3 , respectivamente e (g), (h) e (i): Resíduos componentes da desviância *versus* variáveis explicativas t_1 , t_2 e t_3 , respectivamente.

Analisando a Figura 6.5 é possível notar que nos gráficos (c), (f) e (i), o comportamento dos resíduos tem a forma quadrática, indicativo que um termo quadrático associado a variável explicativa t_3 , deve ser inserido no preditor correspondente a X_i , como esperado. Nos demais gráficos, o comportamento é linear indicando a adequabilidade dos termos utilizados.

6.2 Modelo Poisson-normal semiparamétrico

Para o modelo Poisson-normal semiparamétrico, propomos quatro distintos cenários. No primeiro cenário estudado, é adotado que os preditores são formados apenas por termos paramétricos. Neste caso, amostras de tamanho $n = 50, 100, 200$ e 500 são consideradas sendo que para cada configuração, 1000 conjuntos de dados são gerados. Já nos cenários 2, 3 e 4, são adotados preditores compostos por termos paramétricos e não paramétricos, sendo consideradas amostras de tamanho $n = 100, 200$ e 500 . Em cada cenário, 500 amostras são geradas para cada tamanho amostral adotado.

Para o cenário 1, os valores do viés simulado, da $\sqrt{\text{EQM}}$ simulado, do DP, do EPa médio e da PCa estimada, associados a cada parâmetro do modelo, são apresentados. Já para os demais cenários analisados, apresentam-se os valores do viés simulado, da $\sqrt{\text{EQM}}$ simulado e do DP, referentes à parte paramétrica do modelo. No caso dos cenários 2, 3 e 4, em que os preditores contêm termos não paramétricos, a performance dos ajustes do modelo é analisada calculando o logaritmo do erro quadrático médio empírico, definido em cada réplica $l = 1, 2, \dots, 500$, por

$$\text{EQM}_{\text{emp}}^{(l)} = \frac{1}{n} \sum_{i=1}^n (\eta_i - \hat{\eta}_i^{(l)})^2, \quad (6.1)$$

em que η_i é o valor do preditor real, utilizado para gerar os dados, associado à i -ésima observação e $\hat{\eta}_i^{(l)}$ é o valor do preditor estimado na l -ésima amostra, associado à i -ésima observação. Valores pequenos de $\text{EQM}_{\text{emp}}^{(l)}$ indicam uma proximidade entre as curvas reais e ajustadas, e portanto, um ajuste adequado do modelo.

Cenário 1: São consideradas duas covariáveis para predizer cada uma das respostas Y_i e X_i , em que os preditores são formados por um intercepto, por um termo linear e por uma função exponencial das covariáveis disponíveis, ou seja, $\eta_{iY} = \beta_0 + \beta_1 z_{i1} + \exp(\beta_2 z_{i2})$ e $\eta_{iX} = \delta_0 + \delta_1 t_{i1} + \exp(\delta_2 t_{i2})$, $i = 1, 2, \dots, n$. Os valores das covariáveis z_{i1} e

t_{i1} são gerados de uma distribuição uniforme contínua (0;1) enquanto que os valores das covariáveis z_{i2} e t_{i2} são gerados de uma distribuição uniforme discreta (2;50). Os valores dos coeficientes de regressão adotados são $\beta = (0,5; -0,3; 0,03)$ e $\Delta = (1; 0,5; 0,1)$. Para os demais parâmetros são adotados que $\gamma = 0,5$ e $\sigma^2 = 0,5$. Na Figura 6.6 tem-se os gráficos dos termos lineares e não lineares dos preditores, em função das covariáveis. Os resultados são apresentados na Tabela 6.7. Os gráficos dos resíduos para algumas das amostras geradas são apresentadas nas Figuras 6.7 e 6.8.

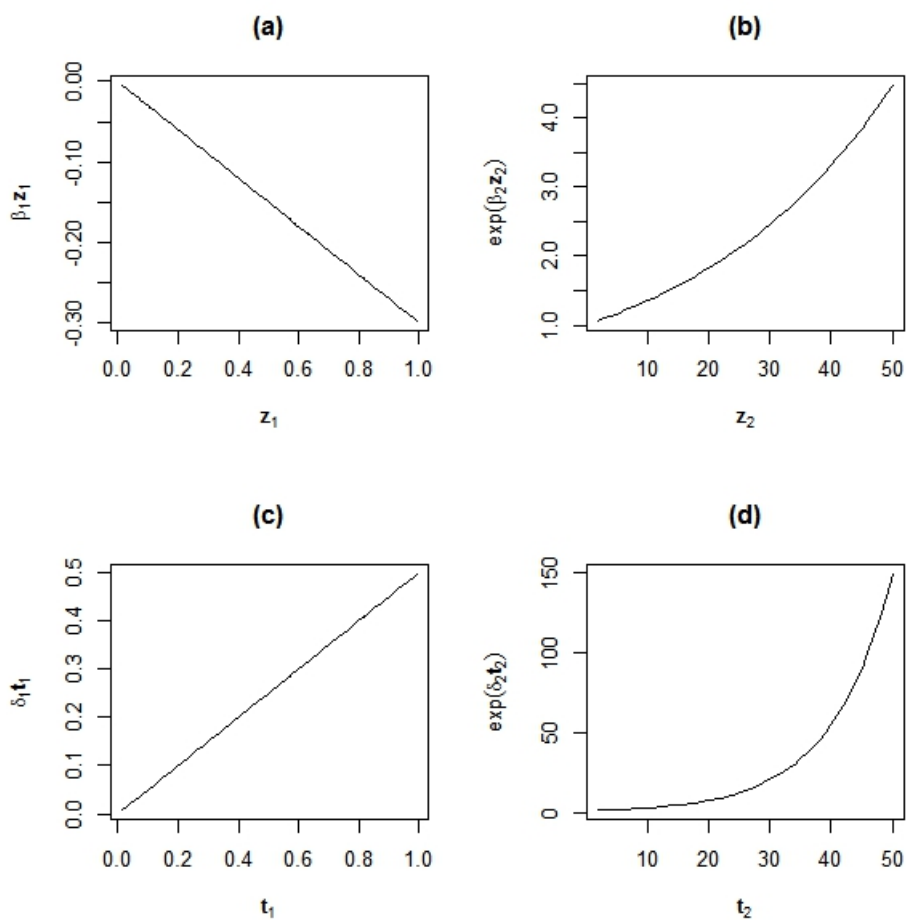


Figura 6.6: (a): termo linear de η_Y , (b): termo não linear de η_Y , (c): termo linear de η_X e (d): termo não linear de η_X - Cenário 1.

Tabela 6.7: Cenário 1 com $\beta = (0, 5; -0, 3; 0, 03)$ e $\Delta = (1; 0, 5; 0, 1)$.

Viés simulado								
n	β_0	β_1	β_2	δ_0	δ_1	δ_2	γ	σ^2
50	0,0004	0,0048	-0,0001	0,0132	-0,0872	0,0000	0,0021	-0,2540
100	-0,0040	0,0119	0,0000	-0,0162	0,0579	0,0000	0,0016	-0,1247
200	0,0006	-0,0008	0,0000	0,0282	-0,0526	0,0000	0,0007	-0,0587
500	0,0000	0,0040	0,0000	-0,0017	0,0345	0,0000	0,0008	-0,0280
$\sqrt{\text{EQM}}$ simulado								
n	β_0	β_1	β_2	δ_0	δ_1	δ_2	γ	σ^2
50	0,0679	0,2595	0,0013	0,6526	3,1908	0,0007	0,0450	0,5254
100	0,0799	0,1922	0,0006	0,5524	1,5757	0,0004	0,0338	0,3681
200	0,0575	0,1382	0,0004	0,3732	1,0309	0,0002	0,0240	0,2512
500	0,0378	0,1012	0,0003	0,2430	0,6911	0,0002	0,0146	0,1602
DP								
n	β_0	β_1	β_2	δ_0	δ_1	δ_2	γ	σ^2
50	0,0679	0,2596	0,0013	0,6527	3,1912	0,0007	0,0449	0,4601
100	0,0799	0,1919	0,0006	0,5524	1,5754	0,0004	0,0337	0,3465
200	0,0576	0,1382	0,0004	0,3723	1,0301	0,0002	0,0240	0,2443
500	0,0378	0,1011	0,0003	0,2431	0,6906	0,0002	0,0145	0,1579
EPa médio								
n	β_0	β_1	β_2	δ_0	δ_1	δ_2	γ	σ^2
50	0,0651	0,2407	0,0012	0,6197	2,9398	0,0007	0,0394	0,4494
100	0,0795	0,1894	0,0006	0,5307	1,4922	0,0004	0,0313	0,3361
200	0,0588	0,1399	0,0004	0,3797	1,0159	0,0002	0,0225	0,2442
500	0,0390	0,1034	0,0003	0,2457	0,6984	0,0002	0,0139	0,1565
PCa estimada (%)								
n	β_0	β_1	β_2	δ_0	δ_1	δ_2	γ	σ^2
50	94,3	92,5	92,4	93,2	92,1	92,4	90,5	82,4
100	94,7	94,6	94,6	93,8	94,3	94,7	93,2	89,2
200	95,6	95,7	95,1	95,4	95,4	94,4	93,2	93,4
500	95,5	95,2	94,9	95,3	94,5	94,0	93,3	93,8

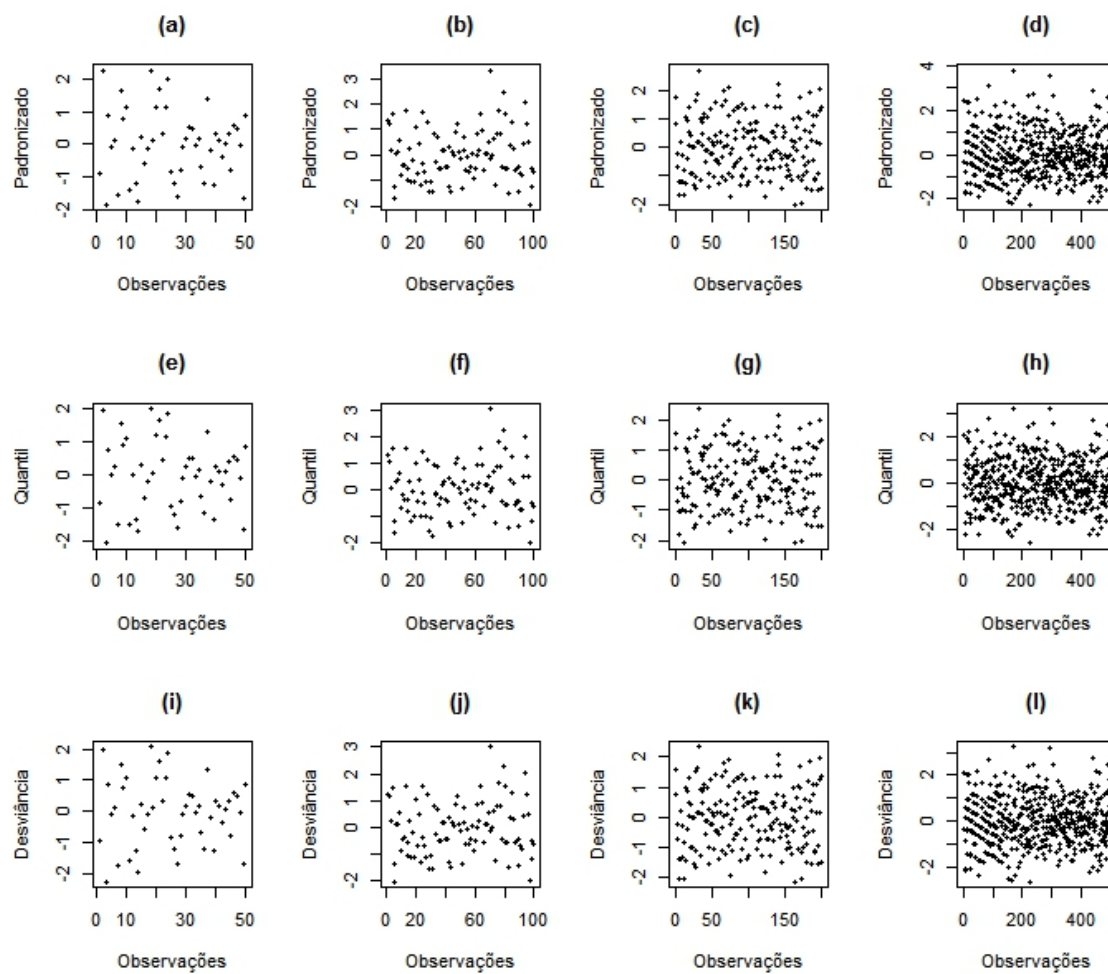


Figura 6.7: Resíduos *versus* observações - Modelo associado à Y_i - Cenário 1 com (a), (e) e (i): $n = 50$; (b), (f) e (j): $n = 100$; (c), (g) e (k): $n = 200$; (d), (h) e (l): $n = 500$.

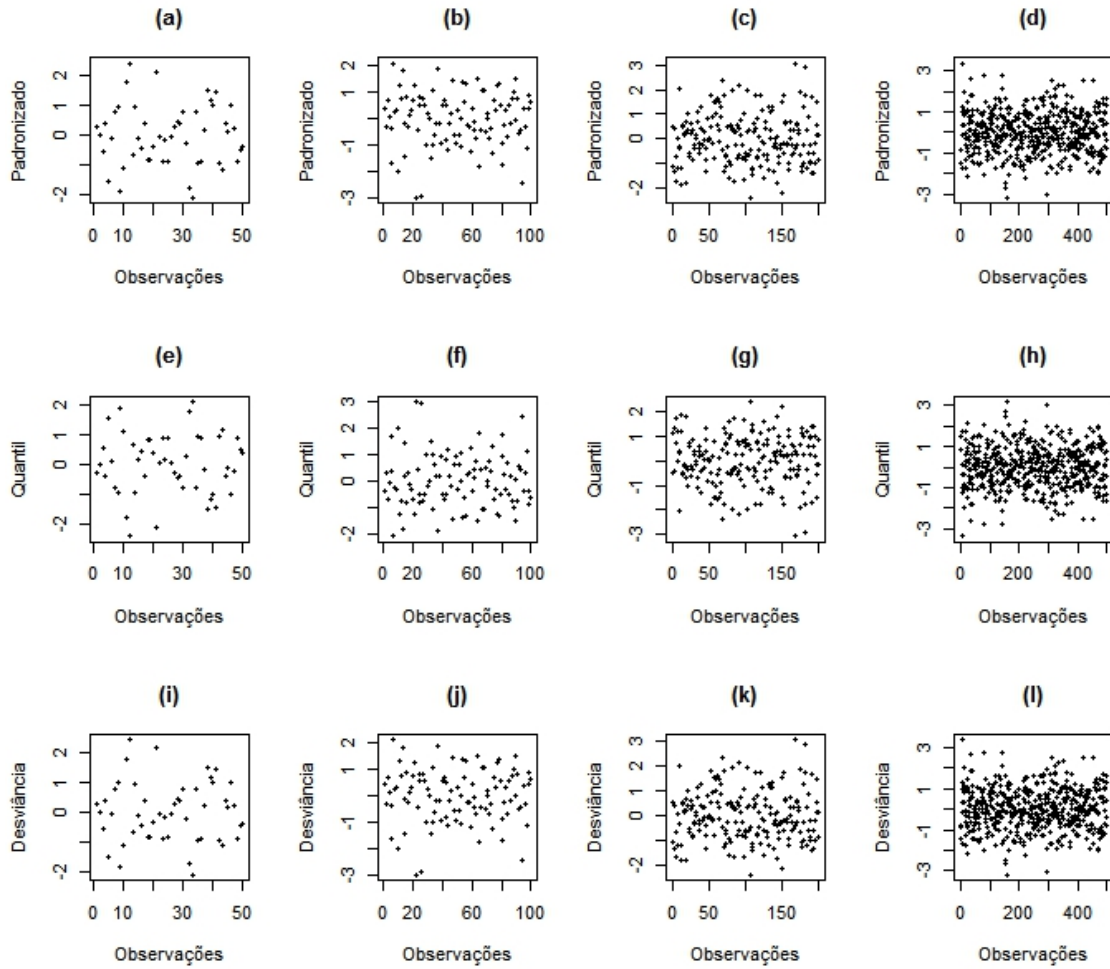


Figura 6.8: Resíduos *versus* observações - Modelo associado à $X_i | Y_i = y_i$ - Cenário 1 com (a), (e) e (i): $n = 50$; (b), (f) e (j): $n = 100$; (c), (g) e (k): $n = 200$; (d), (h) e (l): $n = 500$.

Cenário 2: Neste cenário, é considerado uma única covariável para prever a resposta discreta Y_i e uma única covariável para prever a resposta contínua X_i , em que os preditores são formados apenas por um termo não paramétrico das covariáveis disponíveis. Desta forma, tem-se que $\eta_{iY} = s_1(z_{i1})$ e $\eta_{iX} = u_1(t_{i1})$, $i = 1, 2, \dots, n$. Ambas as covariáveis são geradas de uma distribuição uniforme contínua, no intervalo $(0; 10)$. Além disso, é adotado que $s_1(\mathbf{z}_1) = \text{seno}(0,5\pi\mathbf{z}_1)$ e $u_1(\mathbf{t}_1) = \text{seno}(0,5\pi\mathbf{t}_1)$, em que $\text{seno}(\cdot)$ é a função seno. Para os demais parâmetros do modelo, é adotado os valores

$\gamma = 0,5$ e $\sigma^2 = 0,05$. Na Figura 6.9 apresentamos os gráficos de ambas as funções.

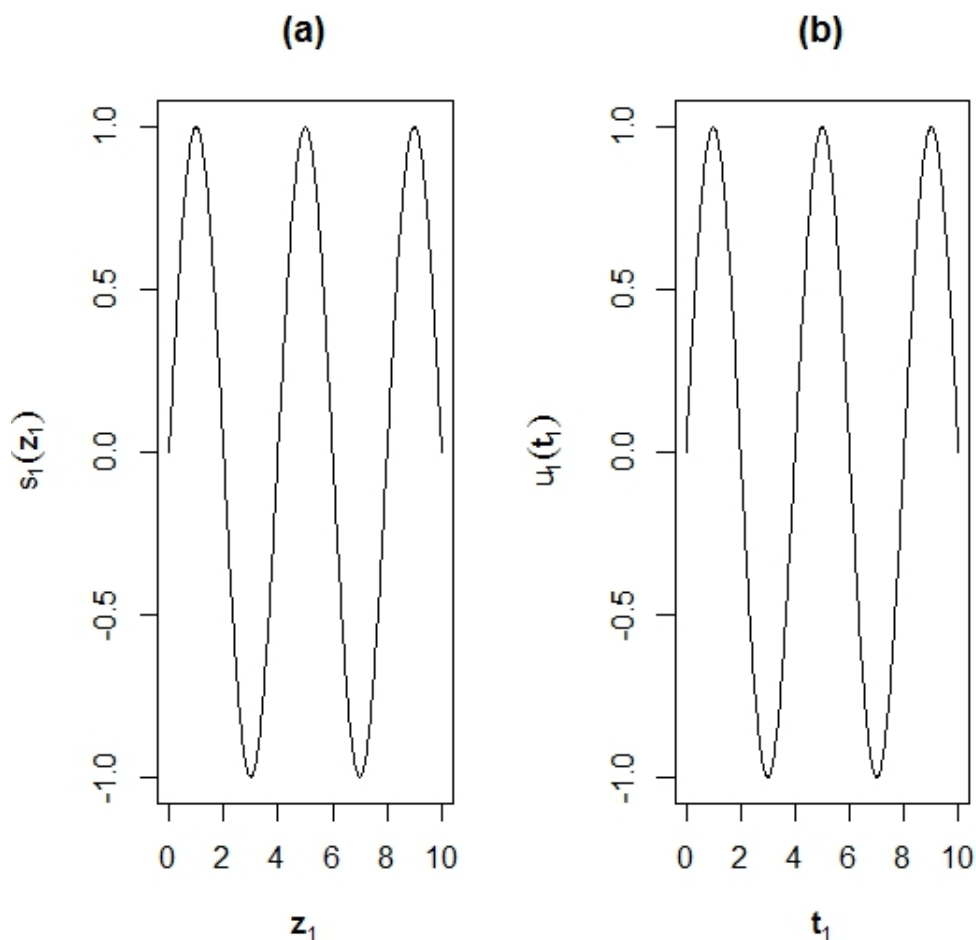


Figura 6.9: (a): curva real $s_1(\cdot)$ utilizada em η_Y e (b): curva real $u_1(\cdot)$ utilizada em η_X - Cenário 2.

Na aplicação da técnica P-splines para suavizar ambos os preditores, é considerado uma Base B-spline de ordem 3, gerada a partir de 20 nós. É adotado uma ordem de penalização igual a 2 para ambas as matrizes de penalização. Esta configuração é utilizada para os três tamanhos de amostra considerados.

Com o objetivo de analisar a performance do ajuste do modelo nos 500 conjuntos de dados gerados, apresentamos na Figura 6.10 os gráficos de caixa dos valores do

logaritmo de $EQM_{emp}^{(l)}$ para os três tamanhos de amostras considerados, calculados em relação a cada um dos preditores η_Y e η_X . O gráfico do ajuste em uma das amostras geradas é mostrado na Figura 6.11. Os ajustes obtidos nas demais amostras, para os diferentes tamanhos amostrais considerados, são similares aos apresentados na Figura 6.11 e portanto, não serão apresentados. Nas Figuras 6.12 e 6.13, são ilustrados os gráficos dos resíduos *versus* as observações, para algumas das amostras geradas. Na Tabela 6.8 são apresentados os valores do Viés, \sqrt{EQM} e DP, associados aos parâmetros γ e σ^2 .

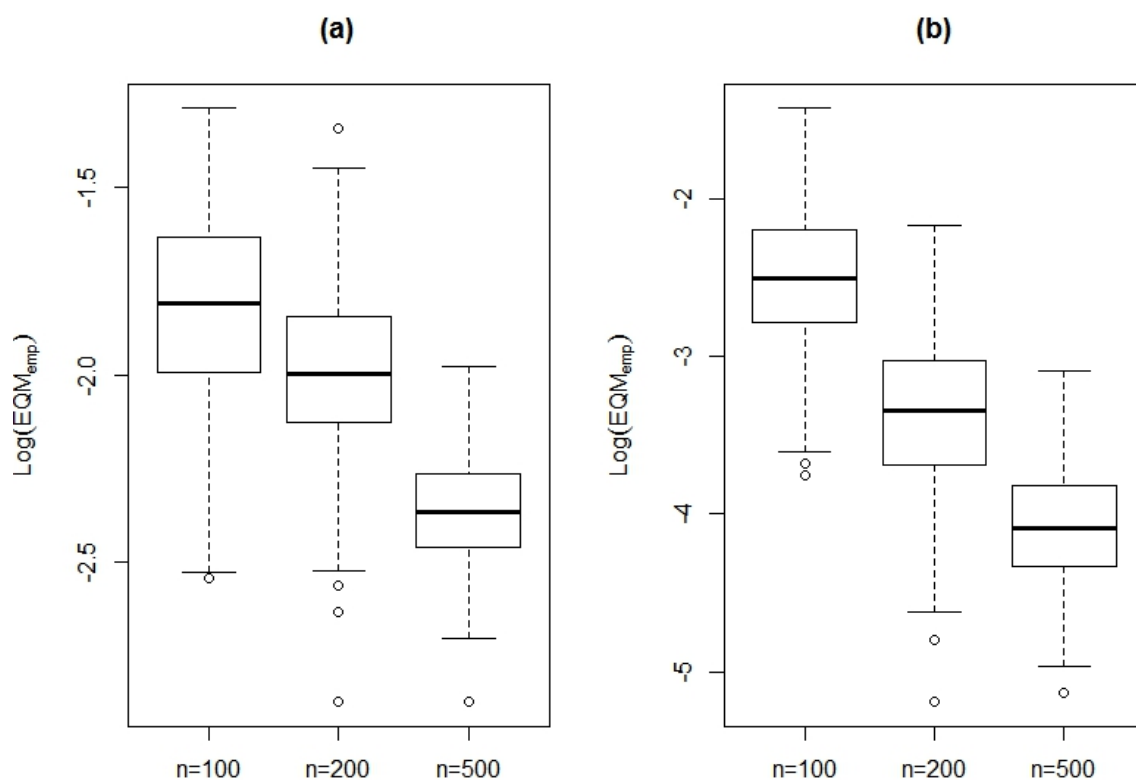


Figura 6.10: (a): gráfico de caixa do $\log(EQM_{emp})$ associado a η_Y e (b): gráfico de caixa do $\log(EQM_{emp})$ associado a η_X - 500 conjuntos de dados com $n = 100, 200$ e 500 - Cenário 2.

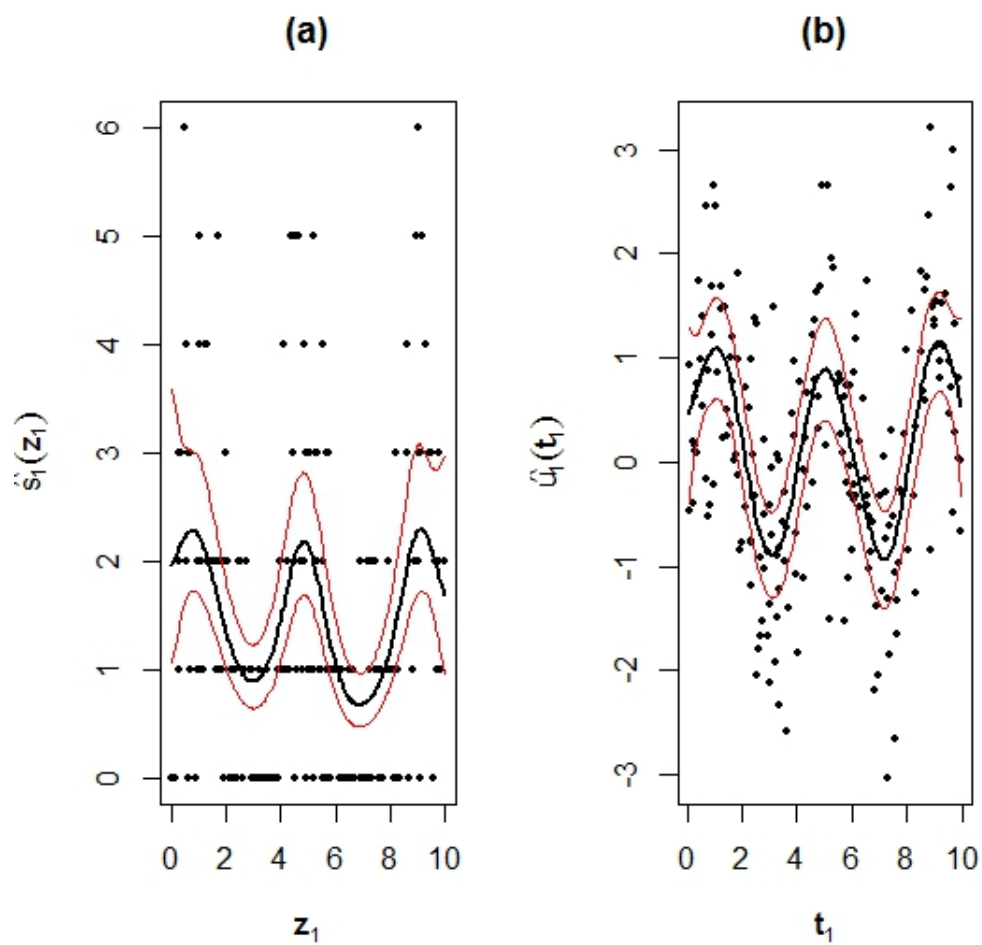


Figura 6.11: (a): curva $s_1(\cdot)$ ajustada, banda de confiança e observações em uma das amostras geradas e (b): curva $u_1(\cdot)$ ajustada, banda de confiança e observações em uma das amostras geradas - Cenário 2 com $n = 200$.

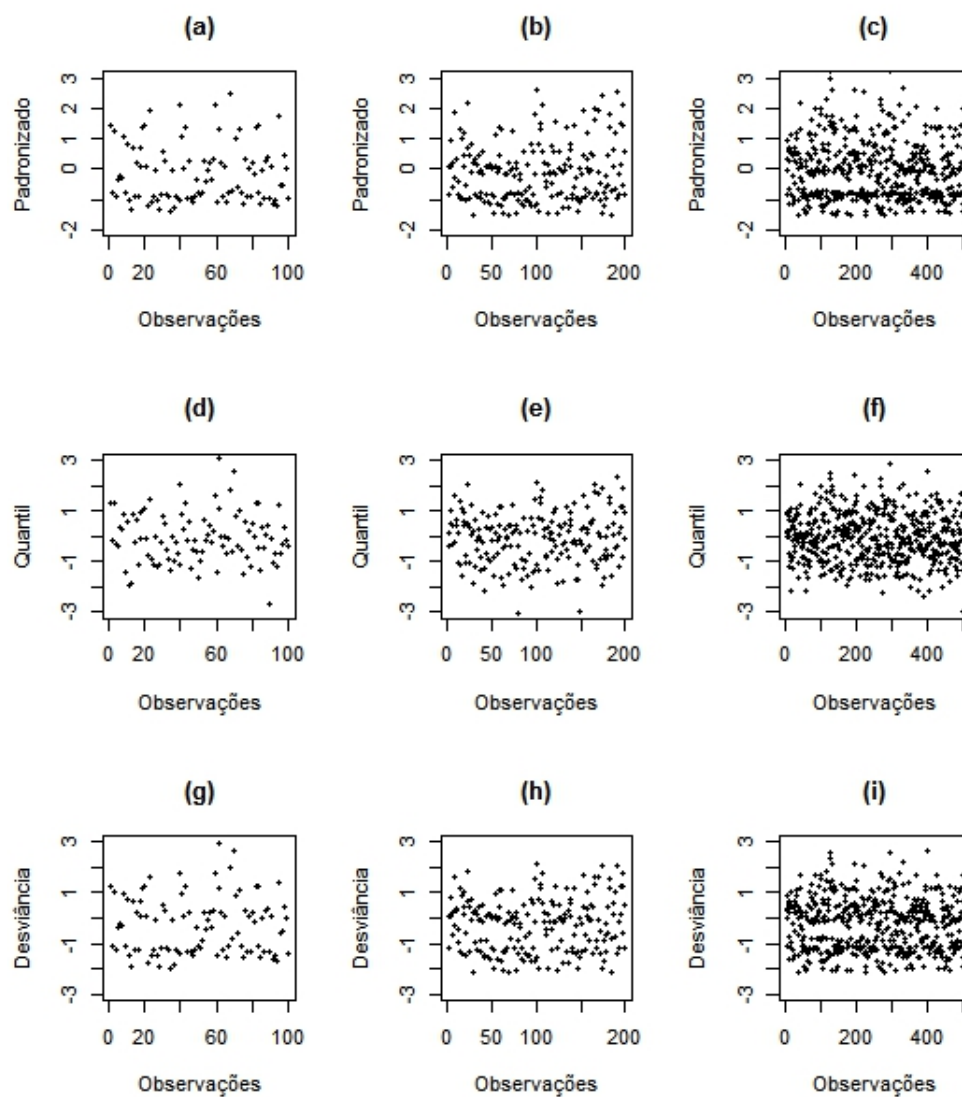


Figura 6.12: Resíduos *versus* observações - Modelo associado à Y_i - Cenário 2 com (a), (d) e (g): $n = 100$; (b), (e) e (h): $n = 200$; (c), (f) e (i): $n = 500$.

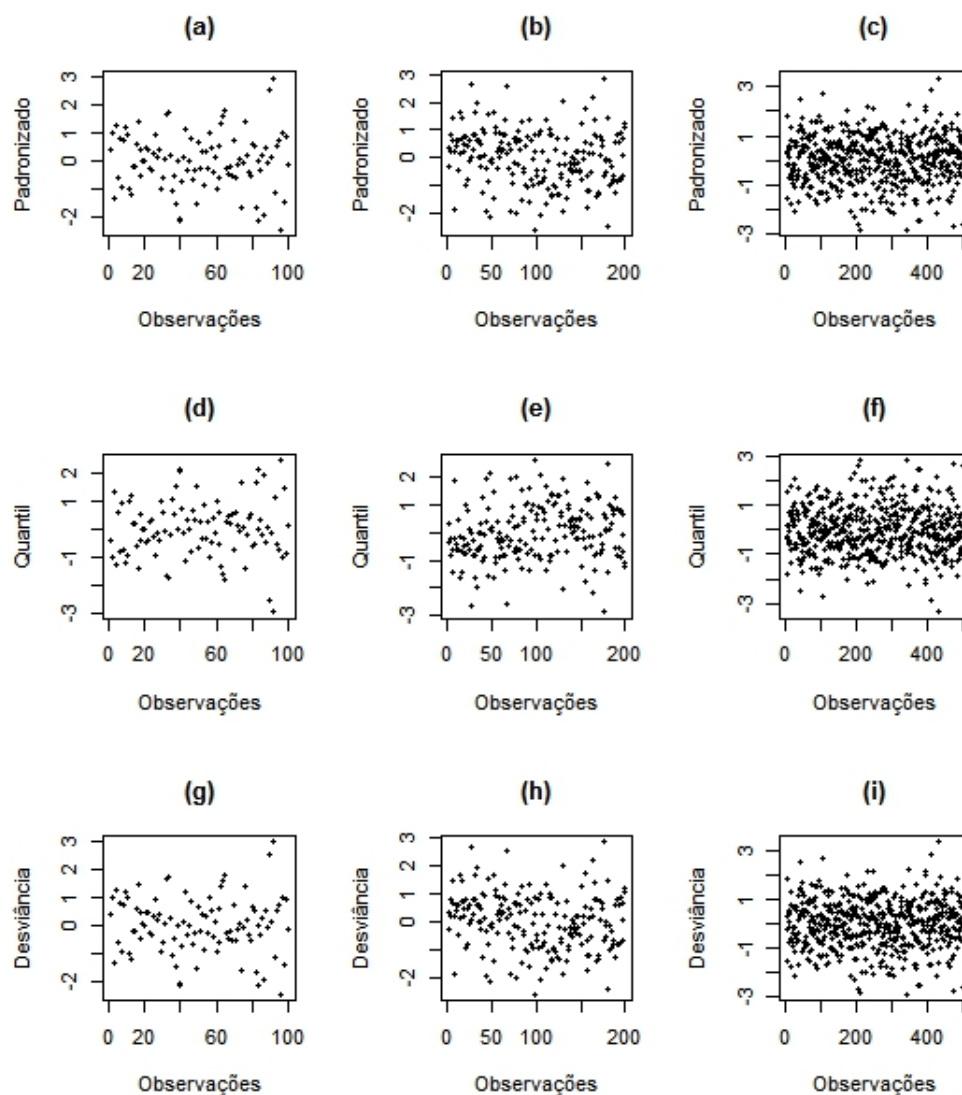


Figura 6.13: Resíduos *versus* observações - Modelo associado à $X_i | Y_i = y_i$ - Cenário 2 com (a), (d) e (g): $n = 100$; (b), (e) e (h): $n = 200$; (c), (f) e (i): $n = 500$.

Tabela 6.8: Viés, \sqrt{EQM} e DP - Cenário 2.

n	Viés simulado		\sqrt{EQM} simulado		DP	
	γ	σ^2	γ	σ^2	γ	σ^2
100	-0,0784	0,0695	0,0963	0,0915	0,0560	0,0597
200	-0,0385	0,0050	0,0569	0,0391	0,0420	0,0389
500	-0,0181	-0,0044	0,0321	0,0324	0,0267	0,0322

Cenário 3: Neste cenário, considera-se que duas covariáveis estão disponíveis para prever as respostas Y_i e X_i , em que os preditores são formados por um intercepto e

por duas funções suaves das covariáveis disponíveis, ou seja, $\eta_{iY} = \beta_0 + s_1(z_{i1}) + s_2(z_{i2})$ e $\eta_{iX} = \delta_0 + u_1(t_{i1}) + u_2(t_{i2})$, $i = 1, 2, \dots, n$. Os valores das covariáveis z_{i1} e t_{i1} são gerados de uma distribuição uniforme contínua (0;1) enquanto que os valores das covariáveis z_{i2} e t_{i2} são gerados de uma distribuição uniforme contínua (0;4). Para a geração de $\boldsymbol{\eta}_Y$, considera-se as curvas reais $s_1(\mathbf{z}_1) = 2\text{seno}(2\pi\mathbf{z}_1)$ e $s_2(\mathbf{z}_2) = 0,5\text{cosseno}(5\mathbf{z}_2)$, em que $\text{seno}(\cdot)$ e $\text{cosseno}(\cdot)$ são as funções seno e cosseno, respectivamente. Além disso, considera-se um valor de intercepto $\beta_0 = 1$. De maneira análoga, considera-se $u_1(\mathbf{t}_1) = 2\text{seno}(2\pi\mathbf{t}_1)$ e $u_2(\mathbf{t}_2) = 0,3\text{cosseno}(2\mathbf{t}_2)$, as curvas reais relacionadas à $\boldsymbol{\eta}_X$, e $\delta_0 = 2$, o respectivo valor do intercepto. É possível observar a forma das quatro curvas, em função das covariáveis geradas, na Figura 6.14. Para os demais parâmetros são adotados que $\gamma = 0,5$ e $\sigma^2 = 0,05$.

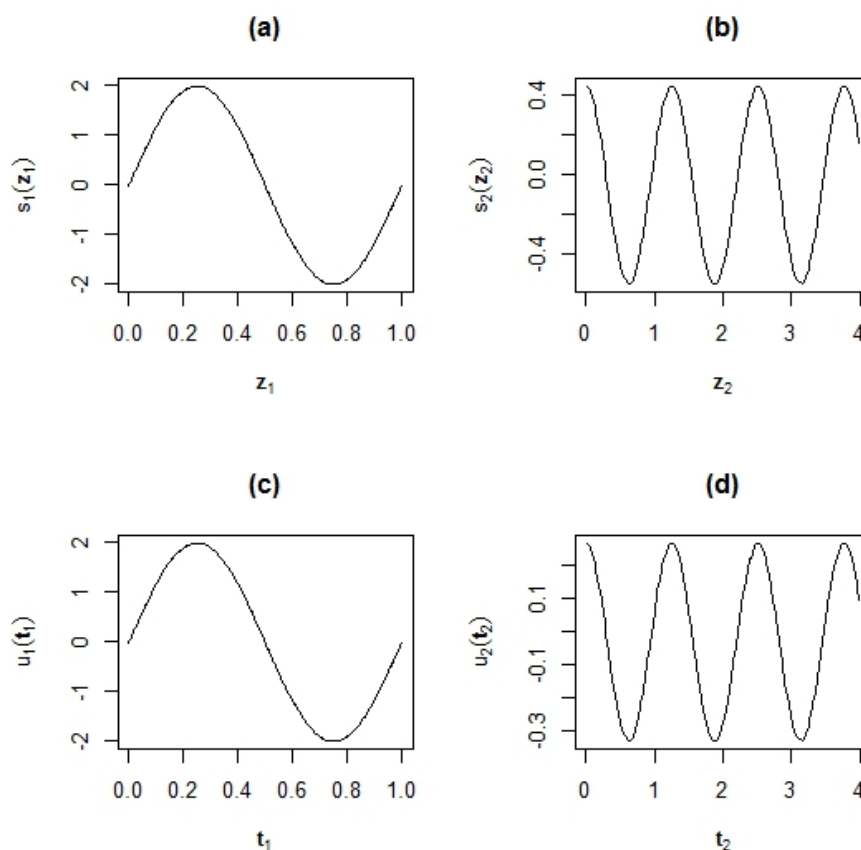


Figura 6.14: (a) e (b): curvas reais $s_1(\cdot)$ e $s_2(\cdot)$, respectivamente, utilizadas em $\boldsymbol{\eta}_Y$, (c) e (d): curvas reais $u_1(\cdot)$ e $u_2(\cdot)$, respectivamente, utilizadas em $\boldsymbol{\eta}_X$ - Cenário 3.

No ajuste do modelo, é considerado que as bases B-splines utilizadas para suavizar as curvas $s_1(\cdot)$, $s_2(\cdot)$, $u_1(\cdot)$ e $u_2(\cdot)$, são de grau 2 e são construídas considerando 12 nós, sendo que as amostras são de tamanhos $n = 100, 200$ e 500 . A ordem de penalização adotada é 2 em todos os casos. Esta configuração é mantida no ajuste dos 500 conjuntos de dados, nos três tamanhos de amostra considerados. Novamente, a performance do ajuste do modelo nas 500 réplicas é analisada através dos gráficos de caixa dos valores do logaritmo de $\text{EQM}_{\text{emp}}^{(l)}$, em relação aos preditores $\boldsymbol{\eta}_Y$ e $\boldsymbol{\eta}_X$, para cada tamanho amostral. Estes gráficos são apresentados na Figura 6.15.

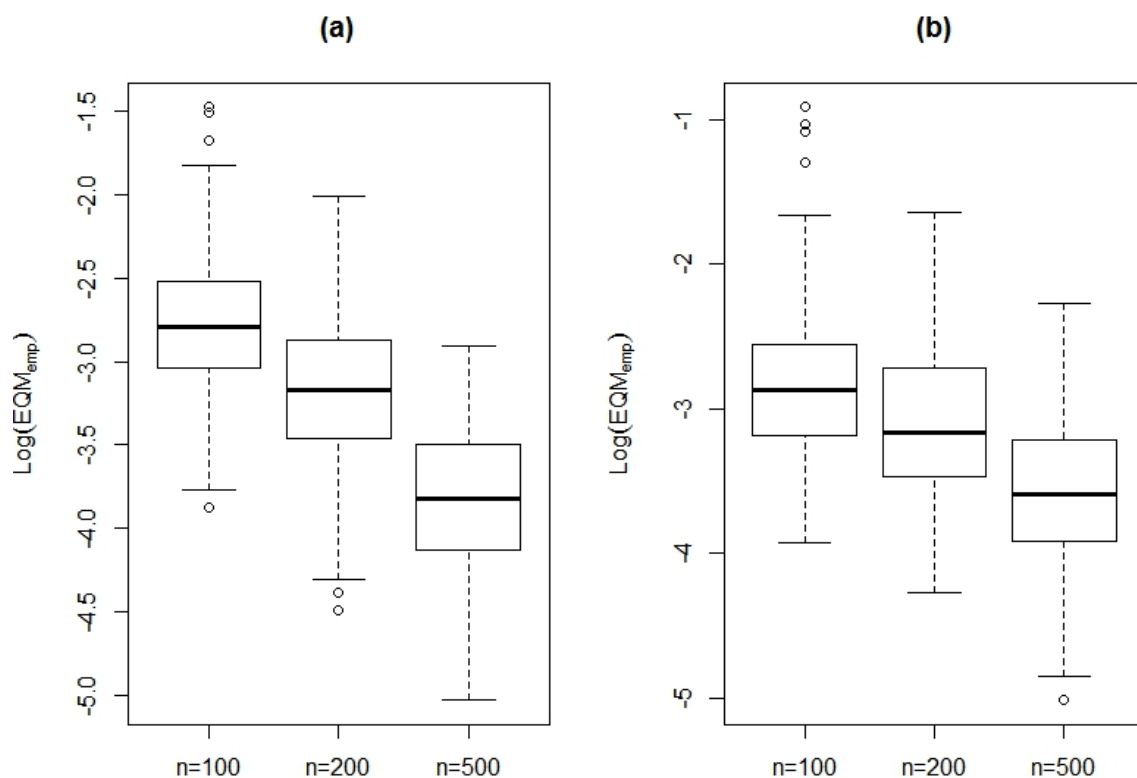


Figura 6.15: (a): gráfico de caixa do $\log(\text{EQM}_{\text{emp}})$ associado a $\boldsymbol{\eta}_Y$ e (b): gráfico de caixa do $\log(\text{EQM}_{\text{emp}})$ associado a $\boldsymbol{\eta}_X$ - 500 conjuntos de dados com $n = 100, 200$ e 500 - Cenário 3.

No caso em que o preditor é formado por uma soma de funções suaves das covariáveis, pode-se analisar o comportamento do ajuste de cada uma das funções suaves

individualmente, analisado os gráficos dos resíduos parciais. Estes gráficos, para uma das amostras geradas, são mostrados na Figura 6.16. Para as demais amostras geradas, o comportamento das curvas ajustadas é semelhante ao obtido na Figura 6.16, com as curvas ajustadas conseguindo representar bem as tendências. Os comportamentos dos resíduos neste cenário são análogos aos obtidos no cenário 2, e portanto, os gráficos não serão apresentados. Os valores do Viés, $\sqrt{\text{EQM}}$ e DP, referentes aos parâmetros γ e σ^2 do modelo, são apresentados na Tabela 6.9.

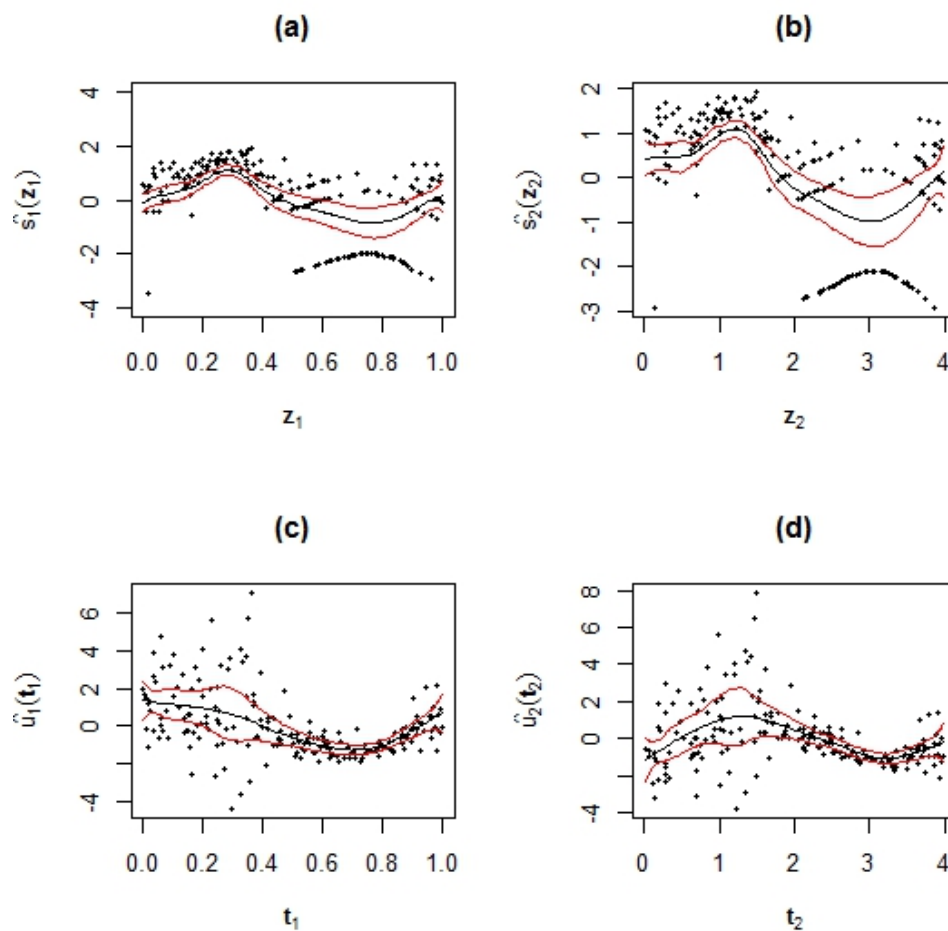


Figura 6.16: (a) e (b): curvas $s_1(\cdot)$ e $s_2(\cdot)$ ajustadas, respectivamente, bandas de confiança e observações em uma das amostras geradas e (c) e (d): curvas $u_1(\cdot)$ e $u_2(\cdot)$ ajustadas, respectivamente, bandas de confiança e observações em uma das amostras geradas - Cenário 3 com $n = 200$.

Tabela 6.9: Viés, $\sqrt{\text{EQM}}$ e DP - Cenário 3.

n	Viés simulado		$\sqrt{\text{EQM}}$ simulado		DP	
	γ	σ^2	γ	σ^2	γ	σ^2
100	0,0006	0,0007	0,0101	0,0043	0,0101	0,0043
200	0,0004	-0,0029	0,0067	0,0048	0,0067	0,0038
500	0,0001	-0,0027	0,0042	0,0041	0,0042	0,0031

Cenário 4: Considera-se que duas covariáveis estão disponíveis para prever as respostas Y_i e X_i , em que os preditores são formados por um intercepto, por uma função suave e por uma função exponencial das covariáveis disponíveis. Assim, tem-se que $\eta_{iY} = \beta_0 + s_1(z_{i1}) + \exp(\beta_2 z_{i2})$ e $\eta_{iX} = \delta_0 + u_1(t_{i1}) + \exp(\delta_2 t_{i2})$, $i = 1, 2, \dots, n$. Os valores das covariáveis z_{i1} e t_{i1} são gerados de uma distribuição uniforme contínua (0;1) enquanto que os valores das covariáveis z_{i2} e t_{i2} são gerados de uma distribuição uniforme discreta (1;20). São adotados que $s_1(\mathbf{z}_1) = \text{seno}(2\pi\mathbf{z}_1)$ e $u_1(\mathbf{t}_1) = 4\text{cosseno}(2\pi\mathbf{t}_1)$. Além disso, considera-se que $\beta_0 = 1$; $\beta_2 = 0,07$; $\delta_0 = 3$; $\delta_2 = 0,06$; $\gamma = 0,7$ e $\sigma^2 = 0,5$. É possível observar a forma das quatro curvas, em função das covariáveis geradas, na Figura 6.17.

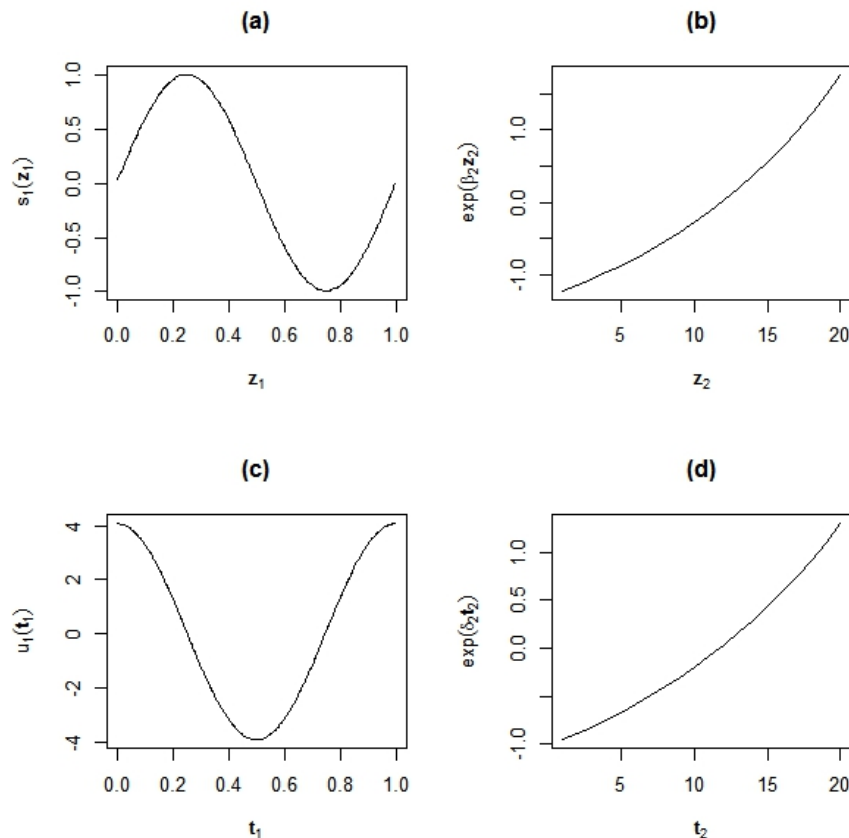


Figura 6.17: (a) e (b): curva real $s_1(\cdot)$ e função não linear, respectivamente, utilizadas em $\boldsymbol{\eta}_Y$, (c) e (d): curva real $u_1(\cdot)$ e função não linear, respectivamente, utilizadas em $\boldsymbol{\eta}_X$ - Cenário 4.

São considerados bases B-splines de grau 2 e também construídas considerando 12 nós. Novamente, a ordem de penalização adotada é 2. Diferentemente dos cenários 2 e 3, neste cenário são considerados também os ajustes dos modelos individuais, associados à resposta discreta Y_i e à resposta contínua X_i , dado à discreta Y_i . Nestes casos, a metodologia apresentada na Seção 3.2 é utilizada no proceso de estimação. As performances dos ajustes do modelo conjuntamente e individualmente nas 500 réplicas são analisadas através dos gráficos de caixa dos valores do logaritmo de $\text{EQM}_{\text{emp}}^{(l)}$, em relação aos preditores $\boldsymbol{\eta}_Y$ e $\boldsymbol{\eta}_X$, para cada tamanho amostral. Estes gráficos são apresentados na Figura 6.18.

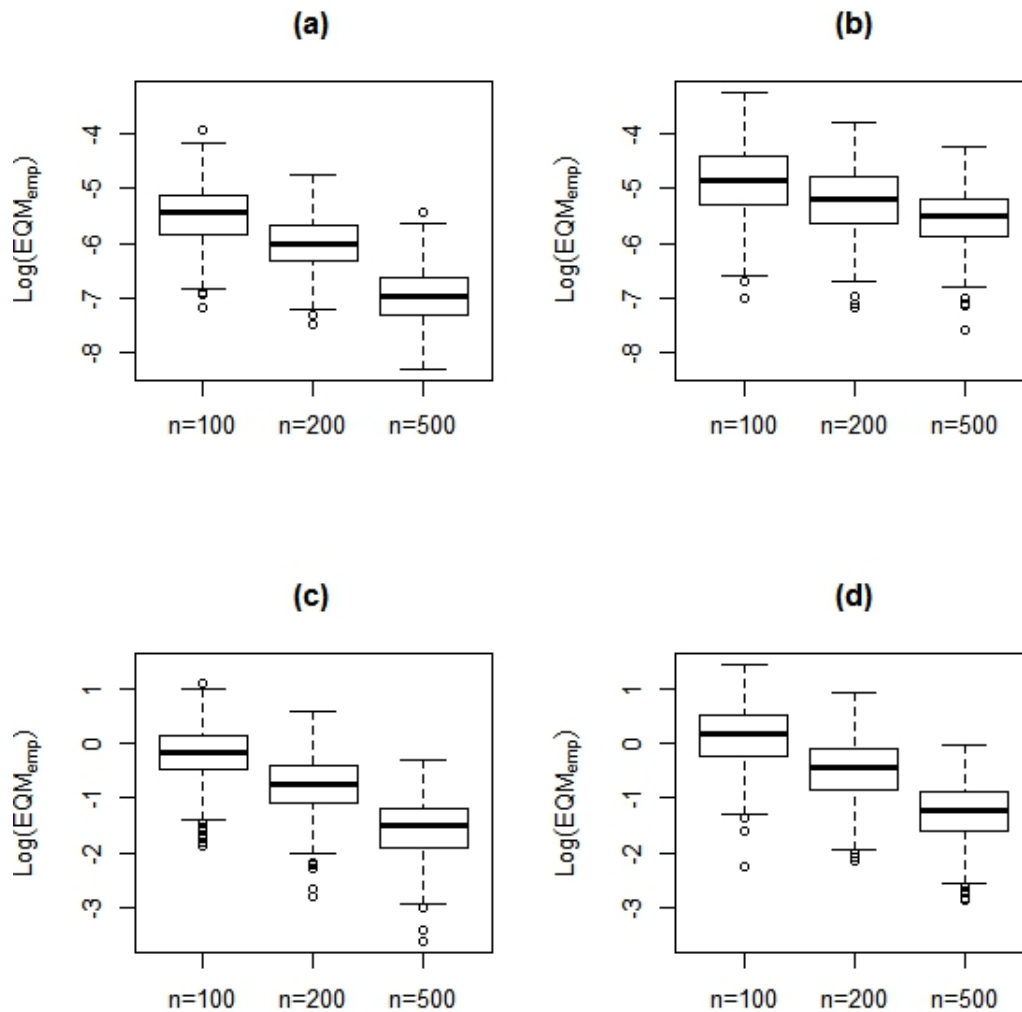


Figura 6.18: (a) e (b): gráficos de caixa do $\log(\text{EQM}_{\text{emp}})$ associado a $\boldsymbol{\eta}_Y$, considerando o ajuste conjunto e individual, respectivamente, (c) e (d): gráficos de caixa do $\log(\text{EQM}_{\text{emp}})$ associado a $\boldsymbol{\eta}_X$, considerando o ajuste conjunto e individual, respectivamente - 500 conjuntos de dados com $n = 100, 200$ e 500 - Cenário 4.

O comportamento do ajuste de cada uma das funções suaves e dos termos não lineares são analisados individualmente, através dos gráficos dos resíduos parciais *versus* as covariáveis. Estes gráficos, para uma das amostras geradas de tamanho $n = 200$ e 500 , são mostrados nas Figuras 6.19 e 6.20, respectivamente, em que a curva preta indica o ajuste obtido conjuntamente e a curva azul indica o ajuste obtido considerando

os modelos individualmente.

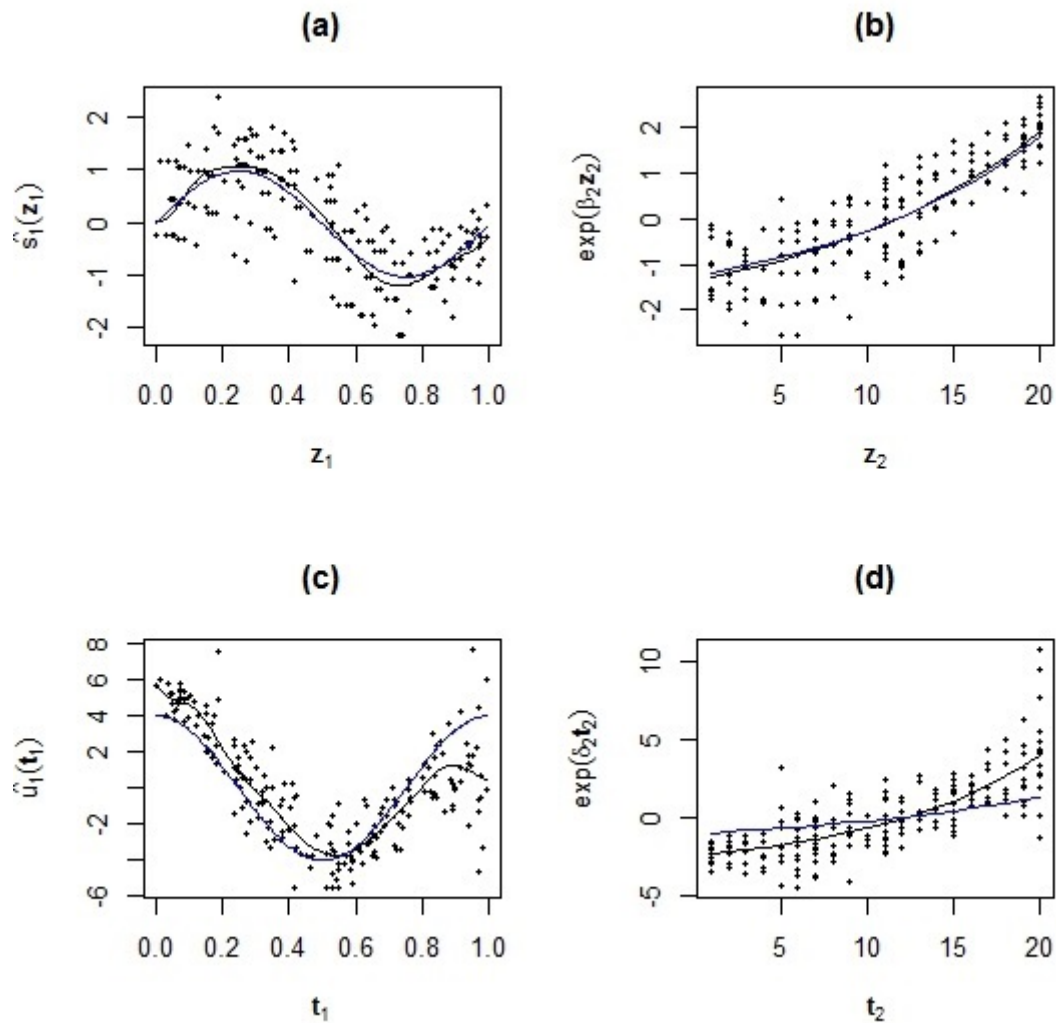


Figura 6.19: (a) e (b): curva $s_1(\cdot)$ e função não linear ajustadas, respectivamente, associadas a $\boldsymbol{\eta}_Y$ e (c) e (d): curva $u_1(\cdot)$ e função não linear ajustadas, respectivamente, associadas a $\boldsymbol{\eta}_X$ - curva preta indica o ajuste obtido considerando o modelo conjunto e curva azul indica o ajuste obtido considerando os modelos individualmente - Cenário 4 com $n = 200$.

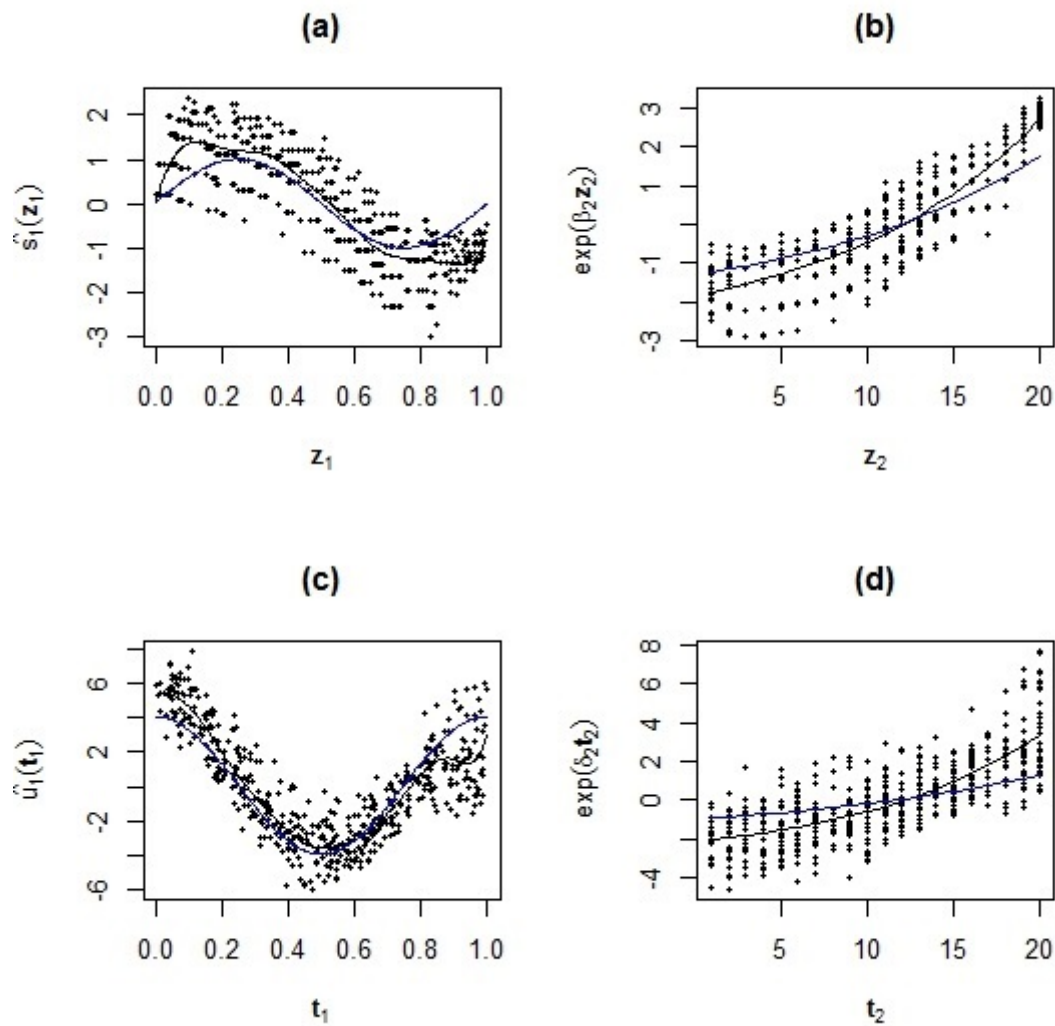


Figura 6.20: (a) e (b): curva $s_1(\cdot)$ e função não linear ajustadas, respectivamente, associadas a η_Y e (c) e (d): curva $u_1(\cdot)$ e função não linear ajustadas, respectivamente, associadas a η_X - curva preta indica o ajuste obtido considerando o modelo conjunto e curva azul indica o ajuste obtido considerando os modelos individualmente - Cenário 4 com $n = 500$.

Os valores do Viés, $\sqrt{\text{EQM}}$ e DP, referentes aos parâmetros β_2 , δ_2 , γ e σ^2 do modelo, são apresentados na Tabela 6.10. Já os comportamentos dos resíduos para algumas das amostras geradas são apresentados nas Figuras 6.21 e 6.22.

Tabela 6.10: Viés, $\sqrt{\text{EQM}}$ e DP - Cenário 4,

Viés simulado				
n	β_2	δ_2	γ	σ^2
100	0,0093	0,0362	0,0048	-0,0771
200	0,0061	0,0157	0,0014	-0,0399
500	0,0031	0,0079	0,0015	-0,0155
$\sqrt{\text{EQM}}$ simulado				
n	β_2	δ_2	γ	σ^2
100	0,0151	0,1058	0,0438	0,1006
200	0,0107	0,0680	0,0281	0,0632
500	0,0076	0,0531	0,0185	0,0344
DP				
n	β_2	δ_2	γ	σ^2
100	0,0119	0,1047	0,0435	0,0648
200	0,0088	0,0576	0,0281	0,0490
500	0,0070	0,0526	0,0185	0,0308

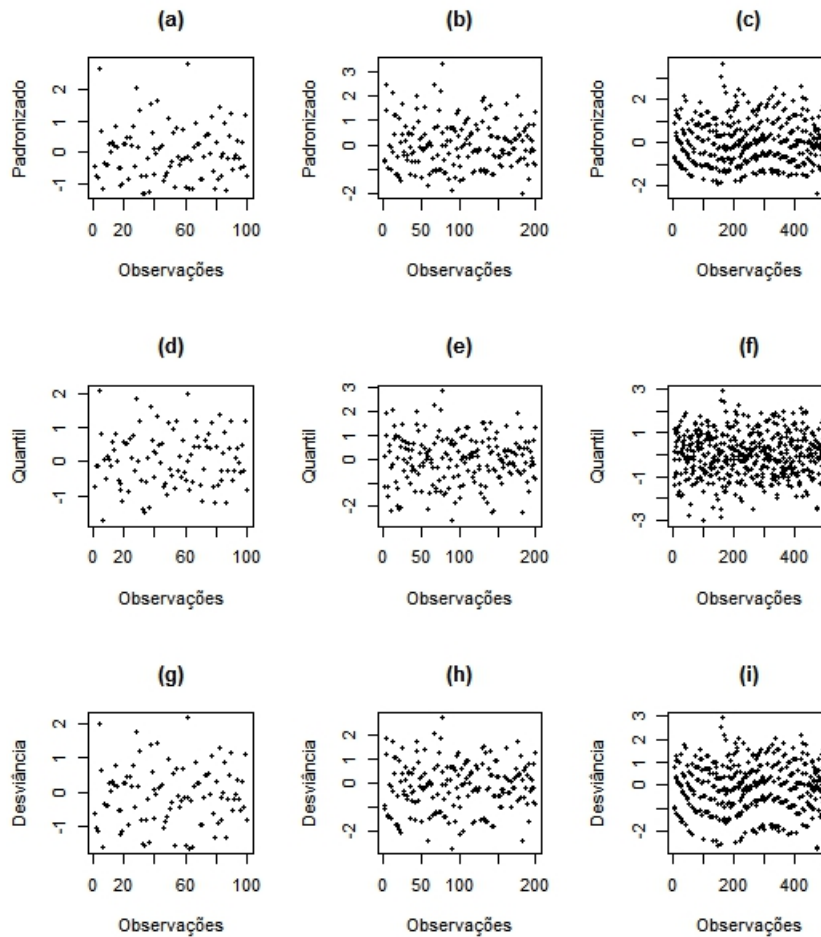


Figura 6.21: Resíduos *versus* observações - Modelo associado à Y_i - Cenário 4 com (a), (d) e (g): $n = 100$; (b), (e) e (h): $n = 200$; (c), (f) e (i): $n = 500$.

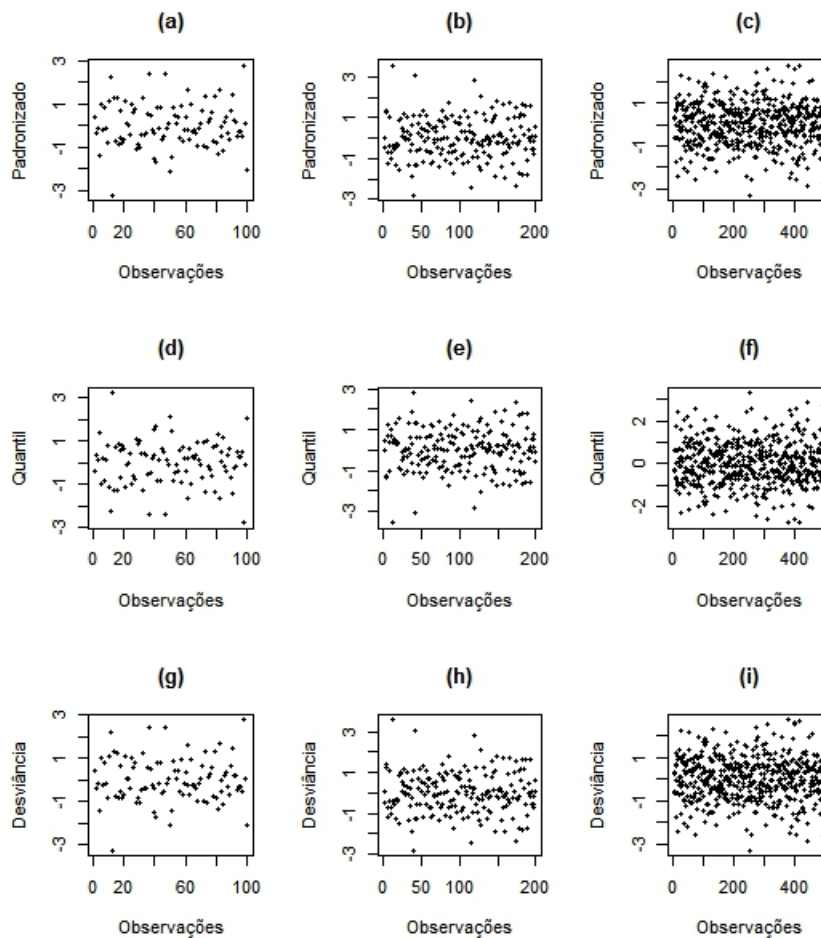


Figura 6.22: Resíduos *versus* observações - Modelo associado à $X_i | Y_i = y_i$ - Cenário 2 com (a), (d) e (g): $n = 100$; (b), (e) e (h): $n = 200$; (c), (f) e (i): $n = 500$.

No cenário 1 em que os preditores são formados apenas por termos lineares e não lineares paramétricos das covariáveis disponíveis, é possível notar analisando a Tabela 6.7 que os valores do viés simulado são relativamente pequenos e em geral, diminuem com o aumento do tamanho amostral. Além disso, notam-se que os valores de \sqrt{EQM} , DP e EPa médio são próximos e tendem a diminuir e se tornarem mais próximos a medida que o tamanho da amostra aumenta, como esperado. Em relação as estimativas da PCA, tem-se que os valores obtidos tendem ao valor da cobertura nominal considerado (95%) a medida que o tamanho de n aumenta, indicando o bom comportamento

dos estimadores de máxima verossimilhança dos parâmetros do modelo no cenário em estudo. Os valores dos resíduos associados a ambos os modelos marginal e condicional estão na faixa entre -3 e 3, e concentrados em torno de 0 em geral, como pode ser observado nas Figuras 6.7 e 6.8. A normalidade dos três resíduos considerados, em ambos os modelos marginais, é verificada em mais de 95% das amostras geradas, utilizando o teste de normalidade Kolmogorov-Smirnov.

Já nos cenários 2, 3 e 4, em que os preditores são formados por termos paramétricos e não paramétricos, é possível observar analisando as Figuras 6.10, 6.15 e 6.18 que os valores do EQM_{emp} são relativamente pequenos e em geral, diminuem a medida que o tamanho da amostra aumenta, como desejado. Além disso, nota-se que em geral os valores do EQM_{emp} obtidos dos ajustes do modelo bivariado são menores do que os valores obtidos dos ajustes dos modelos marginais, indicando que o ajuste conjunto tende a ser melhor se comparado ao ajuste realizado separadamente. Isto também pode ser notado observado as Figuras 6.11, 6.16, 6.19 e 6.20 em que é possível verificar que as curvas ajustadas representam bem as tendências dos dados.

Em relação aos parâmetros γ e σ^2 no caso dos cenários 2 e 3, e dos parâmetros β_2 , δ_2 , γ e σ^2 no cenário 4, pode-se notar pelas tabelas 6.8, 6.9 e 6.10 que os valores do viés simulado são pequenos e tendem a diminuir com o aumento do tamanho amostral assim como os valores da \sqrt{EQM} e do DP que são semelhantes em geral e diminuem com o aumento amostral. Desta forma, o comportamento dos estimadores de máxima verossimilhança associados à parte paramétrica do modelo e os ajustes das curvas suaves associadas à parte não paramétrica, se mostram satisfatórios para dados nas condições dos cenários estudados.

Analisando as Figuras 6.12, 6.13, 6.21 e 6.22 nota-se que o comportamento dos resíduos padronizado, quantil e componentes da desviância em geral são similares nos três cenários estudados, com valores concentrados em torno de 0 e variando em uma faixa

entre -3 e 3. Aplicando-se o teste de normalidade Kolmogorov-Smirnov, é verificada a normalidade dos resíduos em mais de 95% das amostras geradas, para os três tamanhos de amostras considerados.

Capítulo 7

Aplicação

Neste capítulo, ajustamos o modelo Bernoulli-exponencial apresentado na Seção 4.3, a um conjunto de dados reais fornecido por uma operadora de planos de saúde. O conjunto de dados é composto por informações sobre internações de 300 pacientes, em que estão disponíveis o valor do gasto total, em reais, com o paciente no período de internação, a utilização ou não do centro cirúrgico pelo paciente e a idade do paciente, em anos, no momento da internação. Para a operadora, é de grande importância conhecer a relação entre o gasto total obtido e a utilização ou não do centro cirúrgico, bem como a contribuição da idade do paciente nestas duas variáveis. Na Tabela 7.1 é apresentada uma análise descritiva dos dados.

Tabela 7.1: Análise descritiva dos dados.

	Mínimo	Mediana	Média	Desvio Padrão	Máximo
Gasto total (reais)	0	1758	3979,00	10596,65	147000
Idade (anos)	6	58	54,71	21,91	103

A idade do paciente é categorizada em menor do que 25 anos no grupo 1, entre 26-44 anos no grupo 2, entre 45-64 anos no grupo 3, entre 65-83 anos no grupo 4, e acima de 84 anos no grupo 5. Os histogramas dos gastos totais, dos pacientes que utilizaram o centro cirúrgico e dos pacientes que não utilizaram o centro cirúrgico, são dados na Figura 7.1. Já na Figura 7.2, são apresentados os gráficos de caixa dos gastos

totais, dos pacientes que utilizaram ou não o centro cirúrgico, em cada grupo de idade considerado.

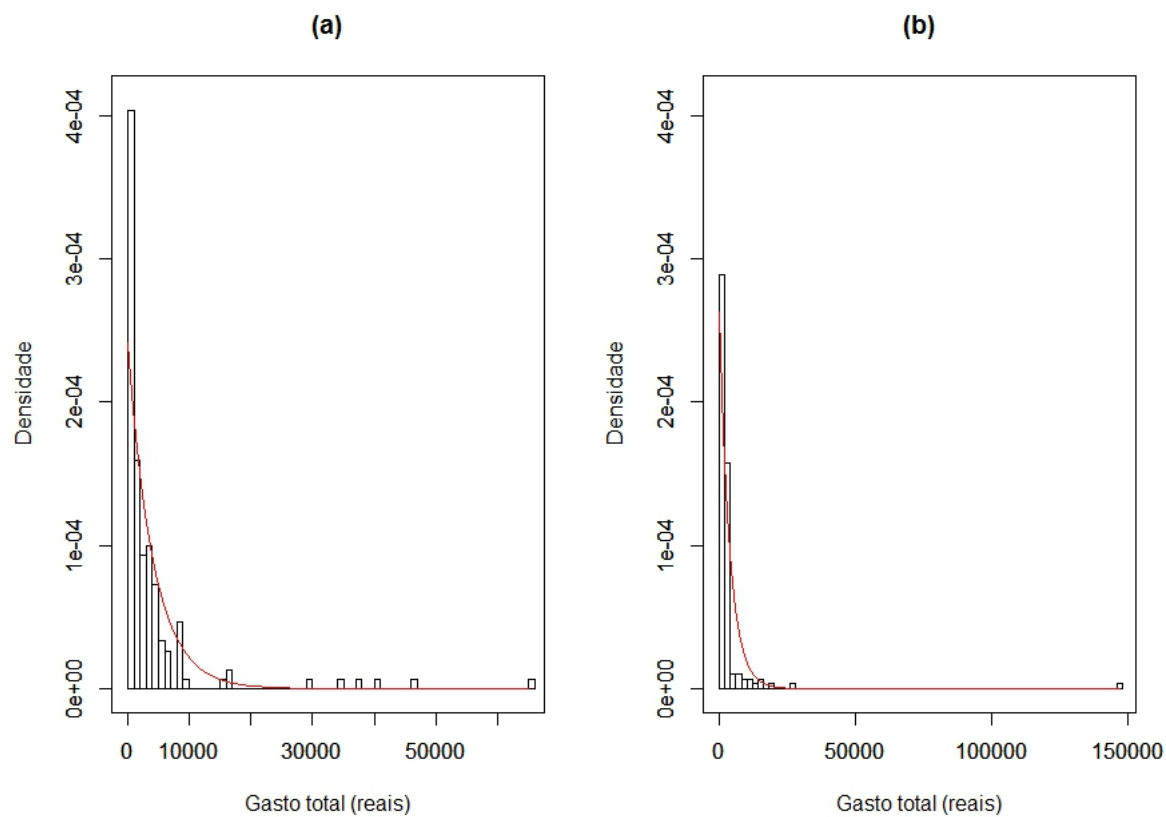


Figura 7.1: (a): histograma dos gastos totais dos pacientes que não utilizaram o centro cirúrgico e (b): histograma dos gastos totais dos pacientes que utilizaram o centro cirúrgico.

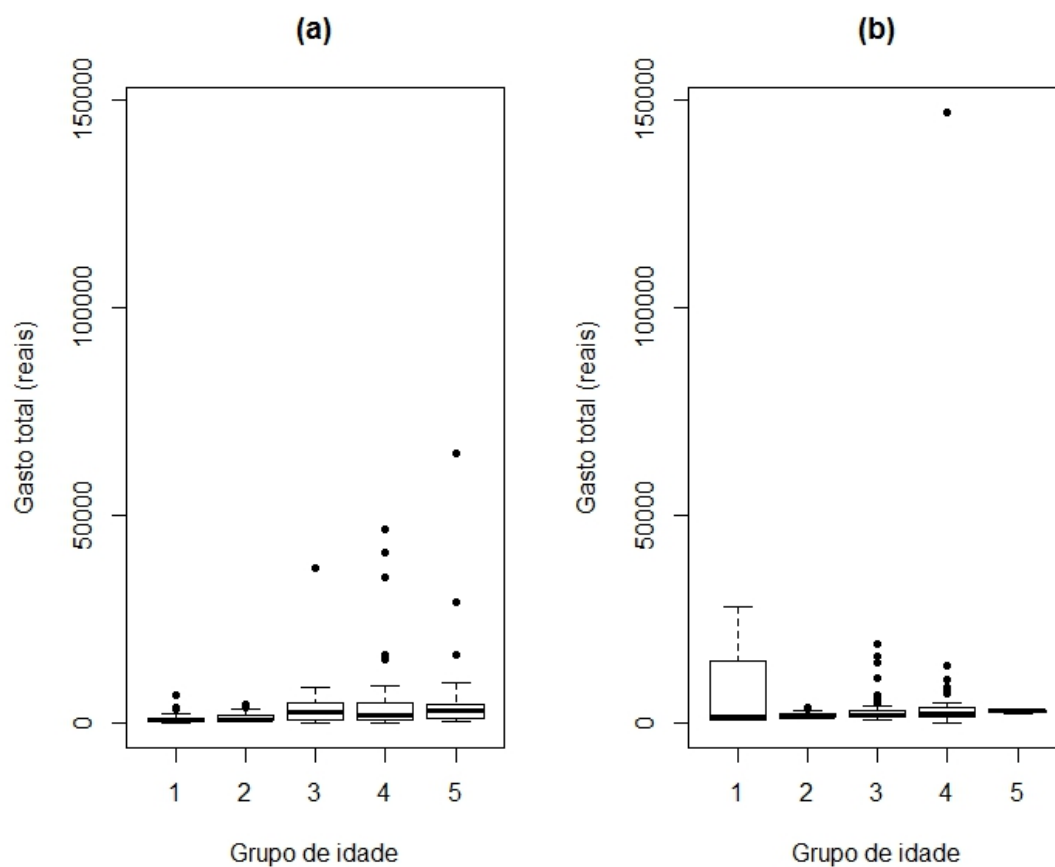


Figura 7.2: (a): gráfico de caixa dos gastos totais dos pacientes que não utilizaram o centro cirúrgico, com idades nos grupos 1, 2, 3, 4 e 5, respectivamente; (b): gráfico de caixa dos gastos totais dos pacientes que utilizaram o centro cirúrgico, com idades nos grupos 1, 2, 3, 4 e 5, respectivamente.

É possível notar pela Figura 7.1 que a soma dos gastos totais dos pacientes que não utilizaram o centro cirúrgico é maior se comparada com a soma dos gastos totais dos pacientes que utilizaram, embora o maior valor gasto pertença a um paciente que utilizou o centro cirúrgico. Além disso, nota-se pela Figura 7.2 que os gastos totais médios se alteram dependendo do grupo de idade considerado. Desta forma, ajustamos o modelo Bernoulli-exponencial ao conjunto de dados reais, em que gasto total é considerada a variável resposta contínua, a utilização ou não do centro cirúrgico é considerada a variável resposta discreta e grupo de idades, a variável explicativa.

Como existem 5 grupos de idades, 4 variáveis *dummy* estão disponíveis (z_{i1} , z_{i2} , z_{i3} e z_{i4}) para explicar as variáveis resposta gasto total e utilização ou não do centro cirúrgico, sendo $z_{ij} = 0$ ou 1, para $j = 1, \dots, 4$ e $i = 1, \dots, 300$. Na Tabela 7.2 é mostrado as estimativas de máxima verossimilhança dos parâmetros do modelo e os respectivos erros padrão estimados e intervalos de confiança assintóticos.

Tabela 7.2: Estimativas dos coeficientes de regressão, erros padrão assintóticos (EPa) e intervalos de confiança assintóticos (IC) - Dados operadora de planos de saúde.

Coefficientes	β_0	β_1	β_2	β_3	β_4
Estimativas	-0,3835	1,5516	0,7742	0,6542	-1,0315
(EPa)	(0,2736)	(0,3940)	(0,3335)	(0,3341)	(0,4846)
(IC 95%)	(-0,9198; 0,1527)	(0,7793; 2,3238)	(0,1205; 1,4278)	(-0,0006; 1,3090)	(-1,9813; -0,0817)
Coefficientes	δ_0	δ_1	δ_2	δ_3	δ_4
Estimativas	8,0683	-0,4922	0,2400	0,8393	0,8266
(EPa)	(0,2262)	(0,2670)	(0,2562)	(0,2599)	(0,3254)
(IC 95%)	(7,6250; 8,5117)	(-1,0156; 0,0311)	(-0,2621; 0,7422)	(0,3299; 1,3487)	(0,1888; 1,4643)

Uma análise de resíduos considerando o ajuste inicial é realizada. Os gráficos dos resíduos dados nas Subseções 5.1.1, 5.1.2 e 5.1.3, *versus* as observações, para o modelo associado à variável discreta (uso ou não do centro cirúrgico) e para o modelo associado à variável contínua (gastos totais), dada à discreta, são apresentadas na Figura 7.3.

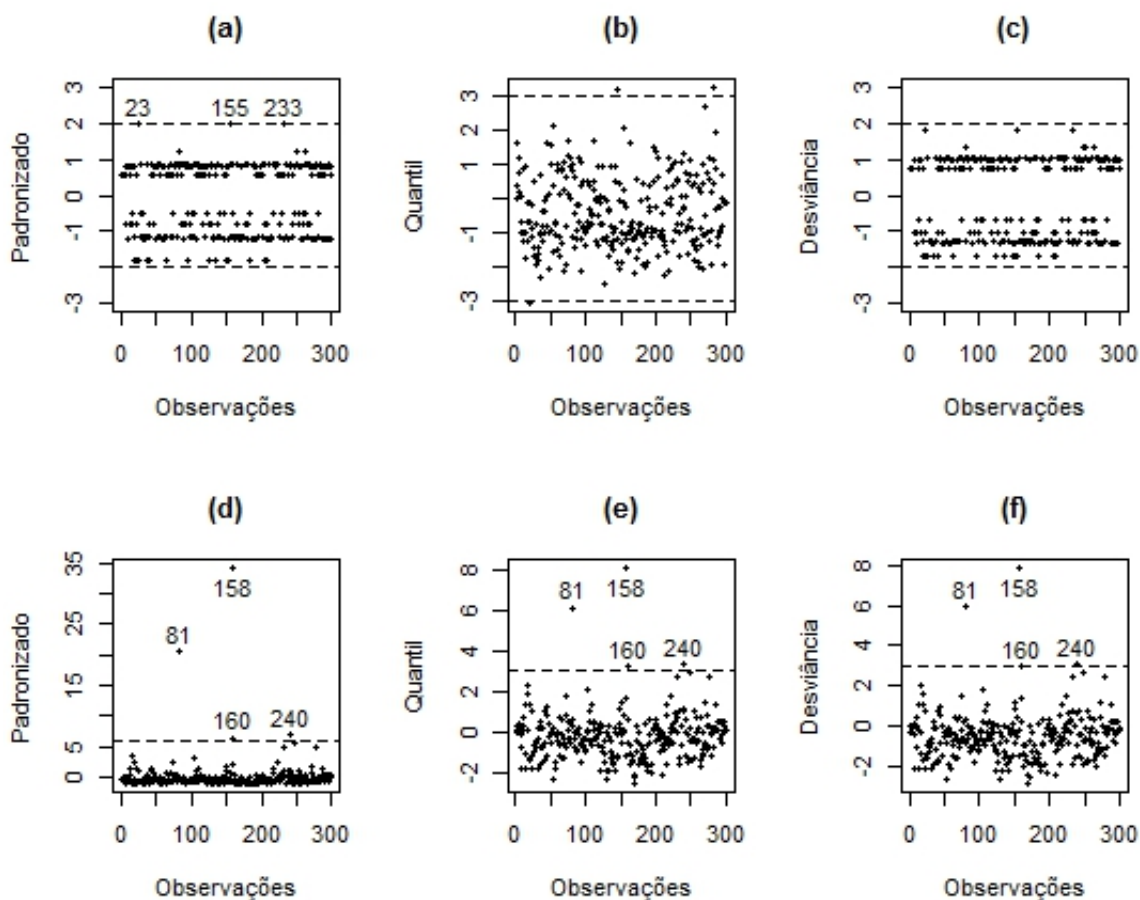


Figura 7.3: (a), (b) e (c): resíduos padronizado, quantil normalizado e componentes da desviância, respectivamente, *versus* observações - Modelo associado à Y ; (d), (e) e (f): resíduos padronizado, quantil normalizado e componentes da desviância, respectivamente, *versus* observações - Modelo associado à $X | Y = y$.

É possível notar que em relação ao modelo associado à Y_i , os valores dos resíduos padronizado referentes às observações 23, 155 e 233 estão no extremo da faixa de valores esperados, como visto no estudo de simulação na Seção 6.1. Em relação aos resíduos associado ao modelo de $X_i | Y_i = y_i$, os valores referentes às observações 81, 158, 160 e 240 estão fora da faixa de valores esperados, indicando que tais pontos devem ser analisados quanto a sua influência no ajuste. Desta forma, as técnicas de influência local e global, dadas nas Seções 5.2 e 5.3, são aplicadas. Os gráficos da direção

\mathbf{v}_{\max} , da curvatura na direção v_i , $i = 1, \dots, n$, da distância de Cook generalizada e do afastamento das verossimilhanças, são apresentados na Figura 7.4.

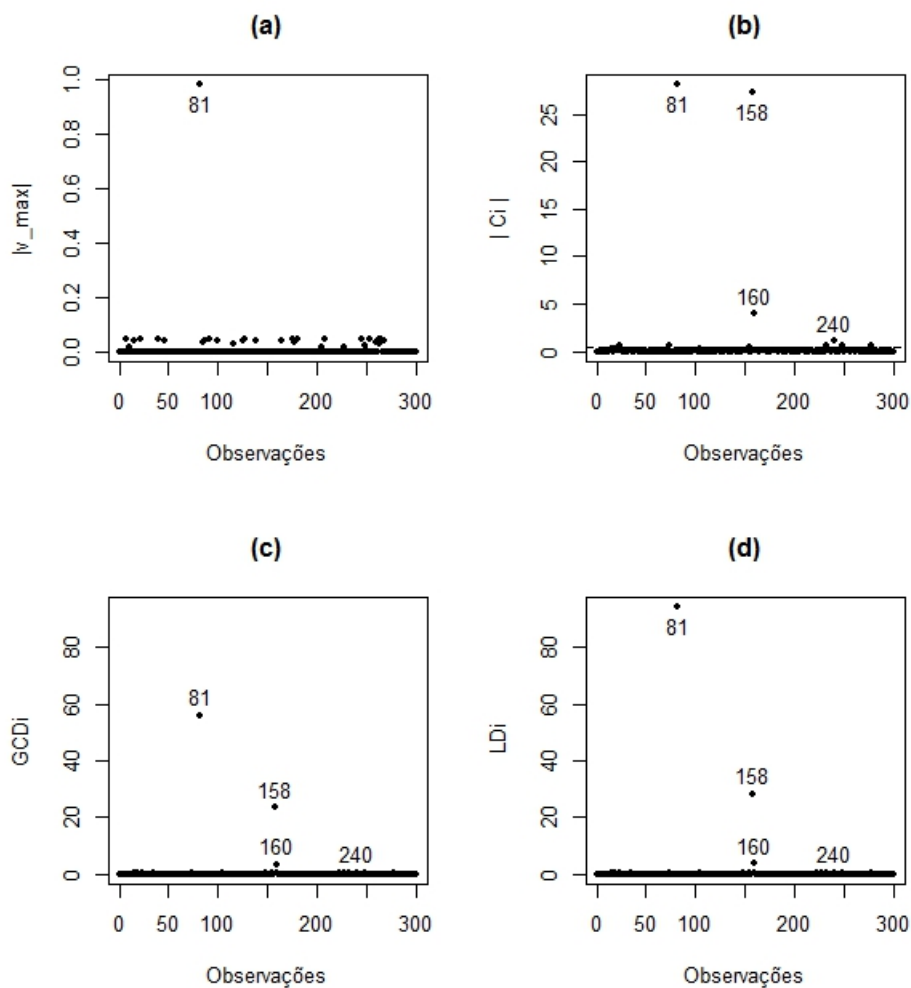


Figura 7.4: (a): direção \mathbf{v}_{\max} versus as observações; (b): curvatura na direção v_i versus as observações; (c): distância de Cook generalizada versus as observações e (d): afastamento das verossimilhanças versus as observações.

Analisando a Figura 7.4, é possível notar que as observações 81 e 158 são fortemente influentes. As observações 160 e 240 também são consideradas influentes, porém, menos do que as observações 81 e 158. Para complementar a análise, é realizado um estudo de exclusão de casos em que as observações em análise são retiradas da amostra e então, calcula-se a mudança percentual das estimativas com a retirada das observações. Na

Tabela 7.3 são apresentadas as mudanças relativas nas estimativas dos parâmetros do modelo, retirando individualmente as observações.

Tabela 7.3: Mudança percentual nos valores das estimativas de máxima-verossimilhança.

Observação excluída	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
81	178,01%	44,00%	88,20%	104,35%	-66,20%
158	0,00%	0,00%	0,00%	-25,81%	0,00%
160	0,00%	0,00%	0,00%	0,00%	-22,97%
240	0,00%	0,00%	-4,39%	0,00%	0,00%
Observação excluída	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_3$	$\hat{\delta}_4$
81	-11,47%	-187,98%	385,53%	110,25%	111,95%
158	0,00%	0,00%	0,00%	-53,34%	0,00%
160	0,00%	0,00%	0,00%	0,00%	-49,69%
240	0,00%	0,00%	-26,61%	0,00%	0,00%

Como visto nas Figuras 7.3 e 7.4, e na Tabela 7.3, as observações 81, 158, 160 e 240 são influentes no ajuste do modelo, com mudanças relativas altas principalmente na observação 81 em que há grande mudança em todos os coeficientes. Note que as variáveis explicativas neste ajuste são variáveis *dummy*, o que explica a grande quantidade de valores zeros, na mudança relativa de alguns parâmetros. A observação 160 se refere ao paciente com maior gasto total, dentre os pacientes que não utilizaram o centro cirúrgico, enquanto que as observações 81, 158 e 240 se referem aos pacientes com os três maiores gastos totais, se considerados apenas os indivíduos que utilizaram o centro cirúrgico. Desta forma, baseado na análise das Figuras 7.3 e 7.4 e na Tabela 7.3, decidimos ajustar o modelo sem os pacientes 81, 158, 160 e 240. Na Tabela 7.4 são apresentados os valores das estimativas de máxima verossimilhança, os erros padrão estimados e os intervalos de confiança assintóticos referentes ao novo ajuste, enquanto que na Figura 7.3 são apresentados os gráficos dos resíduos *versus* as observações.

Tabela 7.4: Estimativas dos coeficientes de regressão, erros padrão assintóticos (EPa), e intervalos de confiança assintóticos (IC) - Dados operadora de planos de saúde sem os pacientes 81, 158, 160, e 240,

Coefficientes	β_0	β_1	β_2	β_3	β_4
Estimativas	-1,0662	2,2342	1,4229	1,1681	-0,1118
(EPa)	(0,3153)	(0,4240)	(0,3690)	(0,3720)	(0,4941)
(IC 95%)	(-1,6842; -0,4482)	(1,4032; 3,0653)	(0,6997; 2,1461)	(0,4389; 1,8972)	(-1,0803; 0,8566)
Coefficientes	δ_0	δ_1	δ_2	δ_3	δ_4
Estimativas	7,1430	0,4331	1,1014	1,3169	1,3412
(EPa)	(0,2162)	(0,2586)	(0,2477)	(0,2514)	(0,3257)
(IC 95%)	(6,7193; 7,5668)	(-0,0738; 0,9399)	(0,6160; 1,5869)	(0,8242; 1,8096)	(0,7028; 1,9795)

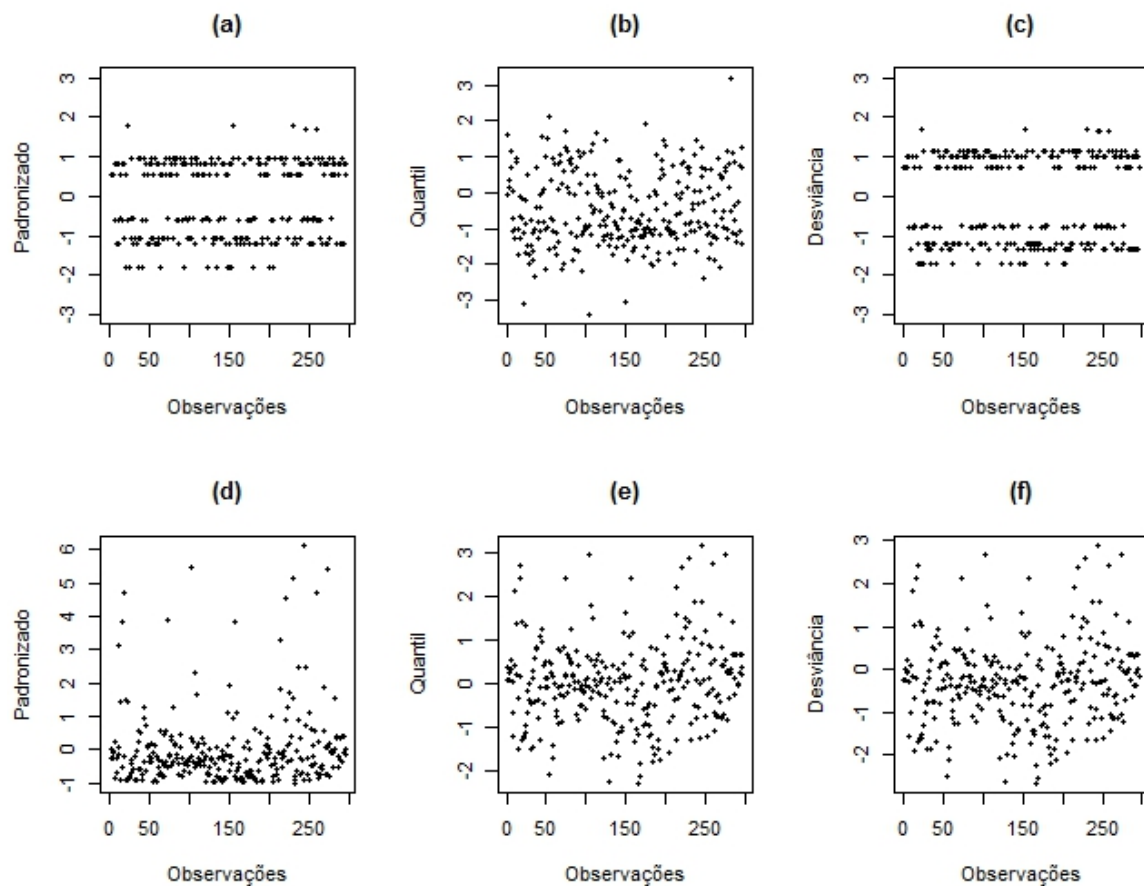


Figura 7.5: (a), (b) e (c): resíduos padronizado, quantil normalizado e componentes da desviância, respectivamente, *versus* observações - Modelo associado à Y . (d), (e) e (f): resíduos padronizado, quantil normalizado e componentes da desviância, respectivamente, *versus* observações - Modelo associado à $X | Y = y$ - ajuste final.

O comportamento dos resíduos apresentados na Figura 7.5 está coerente com os resultados encontrados considerando dados simulados. De fato, em relação à Y_i , os

valores dos resíduos padronizado e componentes da desviância estão concentrados em duas faixas de -2 a 0 e de 0 a 2 e os valores dos resíduos quantil estão próximos de 0 e concentrados entre -3 e 3, enquanto que em relação à $X_i | Y_i = y_i$, os valores do resíduo padronizado variam entre de -2 a 6 e estão concentrados em torno de 0 e os valores dos resíduos quantil e componentes da desviância estão concentrados em 0 e distribuídos em uma faixa de -3 e 3. Desta forma, o ajuste é satisfatório.

Capítulo 8

Conclusão

Neste trabalho propomos uma classe ampla e geral de modelos de regressão bivariados com respostas discreta e contínua, cuja distribuição conjunta é construída a partir do produto de uma distribuição marginal e de uma distribuição condicional. Consideramos que uma das variáveis resposta é discreta e sua distribuição pertencente à família exponencial de distribuições uniparamétrica ou biparamétrica, enquanto que a outra variável resposta é contínua e sua distribuição, condicionada à variável aleatória discreta, pertencente à família exponencial uniparamétrica ou biparamétrica.

Adotamos que covariáveis estão relacionadas às médias das variáveis resposta discreta e contínua por meio de funções de ligação, em que os componentes sistemáticos de ambas as respostas (relacionados às médias marginais) são compostos por um termo linear, como nos modelos lineares generalizados, por um termo não linear paramétrico e por um termo não paramétrico. Além disso, introduzimos uma estrutura de dependência entre as respostas discreta e contínua por meio da média da distribuição condicional.

Para o caso em que os preditores são formados apenas por termos paramétricos das covariáveis disponíveis, o método de máxima verossimilhança é considerado para

a estimação dos parâmetros do modelo em que o vetor escore e a matriz de informação observada ou esperada são apresentados. Por outro lado, quando os preditores são formados por termos lineares e por termos não paramétricos (funções suaves) das covariáveis disponíveis, é proposto o uso da técnica de suavização P-splines para suavizar conjuntamente as curvas do modelo bivariado. Utilizando resultados já conhecidos na literatura para o caso univariado, é possível reduzir o problema não paramétrico em um problema linear, em que a função log-verossimilhança conjunta é penalizada. Desta forma, pode-se utilizar a metodologia desenvolvida para o caso paramétrico para estimar os modelos pertencentes à classe.

No caso mais geral, em que os preditores são compostos por termos lineares e não lineares paramétricos e por termos não paramétricos, não é possível representar os preditores somente por termos lineares. Entretanto, a metodologia apresentada na Subseção 2.1.1 é utilizada já que o problema semiparamétrico se reduz a um problema paramétrico, com preditores formados por termos lineares e não lineares. Note porém, que a inserção de termos paramétricos não lineares nos preditores pode gerar dificuldades na obtenção da inversa da matriz de informação observada, necessária no processo de estimação. Este tópico será estudado com mais detalhes em um trabalho futuro.

Posteriormente, ilustramos dois modelos existentes na literatura que são casos particulares da classe de modelos bivariados proposta e apresentamos dois novos modelos, um paramétrico e outro semiparamétrico, que também pertencem à classe. Então, propomos os resíduos padronizados, quantis aleatorizados e componentes da desviância, além das técnicas de influência local e influência global ou total, para os modelos pertencentes à classe bivariada proposta. Note que as distribuições dos resíduos não são conhecidas neste caso e assim, seu comportamento é estudado através de estudo de simulação.

Um estudo com dados simulados é então realizado para analisar o comportamento

dos estimadores de máxima verossimilhança dos parâmetros dos modelos Bernoulli-exponencial e Poisson-normal semiparamétrico, para analisar o ajuste dos termos não paramétricos (funções suaves dos dados), no caso do modelo Poisson-normal semiparamétrico, e também para analisar a performance dos resíduos apresentados na Seção 5.1, em alguns cenários predeterminados. A eficiência da técnica P-splines nos cenários estudados é verificada assim como as propriedades dos estimadores de máxima verossimilhança, no caso paramétrico. Por fim, ajustamos o modelo Bernoulli-exponencial a um conjunto de dados reais em que o ajuste se mostrou satisfatório.

Portanto, em geral a classe se mostrou flexível em problemas envolvendo dados que não seguem distribuição normal, mostrando ser uma ferramenta útil em problemas mais gerais envolvendo dados bivariados.

Capítulo 9

Propostas de trabalhos futuros

Como propostas de trabalhos futuros, listamos os seguintes tópicos:

- Aperfeiçoar o uso da técnica P-splines em modelos bivariados semiparamétricos;
 - Propriedades assintóticas;
 - Matriz de covariâncias para as curvas ajustadas;
 - Propor maneiras mais eficientes de se determinar os parâmetros de suavidade;
- Estender a classe bivariada para permitir dados agrupados;
- Estender a classe bivariada a uma classe multivariada;
- Apresentar métodos de estimação bayesiana para a classe proposta;
- Propor resíduos e técnicas gráficas bivariadas para a classe bivariada geral;
- Propor técnicas de diagnóstico para a classe bivariada geral, com termos não paramétricos inclusos;
- Introduzir outros tipos de perturbações no estudo de influência local;

- Adaptar critérios de seleção de modelos para a classe proposta;
- Organizar os códigos e criar um pacote com todas as ferramentas estudadas implementadas.

Apêndice A

Derivadas da função log-verossimilhança

Os componentes do vetor escore $\mathbf{U}(\boldsymbol{\beta})$, $\mathbf{U}(\boldsymbol{\Delta})$, $\mathbf{U}(\boldsymbol{\gamma})$, $\mathbf{U}(\boldsymbol{\phi})$ e $\mathbf{U}(\boldsymbol{\varphi})$ em forma matricial são dados pelas equações (2.12) – (2.16) em que

$$\begin{aligned}\boldsymbol{\mu}_Y &= (\mu_{1Y}, \mu_{2Y}, \dots, \mu_{nY})^\top, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)^\top, \quad \mathbf{J}_\gamma = \left(\frac{d\xi_1}{d\gamma}, \frac{d\xi_2}{d\gamma}, \dots, \frac{d\xi_n}{d\gamma} \right)^\top, \\ \mathbf{1} &= (1, 1, \dots, 1)^\top, \quad \boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)^\top, \\ \mathbf{b} &= (b(\theta_1), b(\theta_2), \dots, b(\theta_n))^\top, \quad \boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_n)^\top, \quad \mathbf{b}^* = (b^*(\vartheta_1), b^*(\vartheta_2), \dots, b^*(\vartheta_n))^\top, \\ \mathbf{c} &= (c'(y_1, \phi), c'(y_2, \phi), \dots, c'(y_n, \phi))^\top \text{ e } \mathbf{c}^* = (c^{*'}(x_1, \varphi), c^{*'}(x_2, \varphi), \dots, c^{*'}(x_n, \varphi))^\top,\end{aligned}$$

são vetores n -dimensional,

$$\mathbf{S}_1 = \begin{pmatrix} \frac{\partial \eta_{1Y}}{\partial \beta_0} & \frac{\partial \eta_{2Y}}{\partial \beta_0} & \cdots & \frac{\partial \eta_{nY}}{\partial \beta_0} \\ \frac{\partial \eta_{1Y}}{\partial \beta_1} & \frac{\partial \eta_{2Y}}{\partial \beta_1} & \cdots & \frac{\partial \eta_{nY}}{\partial \beta_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_{1Y}}{\partial \beta_p} & \frac{\partial \eta_{2Y}}{\partial \beta_p} & \cdots & \frac{\partial \eta_{nY}}{\partial \beta_p} \end{pmatrix}^\top \text{ e } \mathbf{S}_2 = \begin{pmatrix} \frac{\partial \eta_{1X}}{\partial \delta_0} & \frac{\partial \eta_{2X}}{\partial \delta_0} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_0} \\ \frac{\partial \eta_{1X}}{\partial \delta_1} & \frac{\partial \eta_{2X}}{\partial \delta_1} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \eta_{1X}}{\partial \delta_q} & \frac{\partial \eta_{2X}}{\partial \delta_q} & \cdots & \frac{\partial \eta_{nX}}{\partial \delta_q} \end{pmatrix}^\top,$$

são matrizes $n \times (p + 1)$ e $n \times (q + 1)$, respectivamente, e

$$\begin{aligned} \mathbf{W} &= \text{diag}(w_1, w_2, \dots, w_n), \quad \mathbf{V} = \text{diag}(V_1, V_2, \dots, V_n), \quad \mathbf{R} = \text{diag}(r_1, r_2, \dots, r_n), \\ \mathbf{Q} &= \text{diag}(Q_1, Q_2, \dots, Q_n), \quad \mathbf{O} = \text{diag}\left(\frac{d\mu_{1X}}{d\eta_{1X}}, \frac{d\mu_{2X}}{d\eta_{2X}}, \dots, \frac{d\mu_{nX}}{d\eta_{nX}}\right), \\ \mathbf{J}_Y &= \text{diag}\left(\frac{d\xi_1}{d\mu_{1Y}}, \frac{d\xi_2}{d\mu_{2Y}}, \dots, \frac{d\xi_n}{d\mu_{nY}}\right) \quad \text{e} \quad \mathbf{J}_X = \text{diag}\left(\frac{d\xi_1}{d\mu_{1X}}, \frac{d\xi_2}{d\mu_{2X}}, \dots, \frac{d\xi_n}{d\mu_{nX}}\right), \end{aligned}$$

são matrizes diagonal, em que $V_i = V(\mu_{iY}) = b''(\theta_i) = \frac{d\mu_{iY}}{d\theta_i}$, é a função de variância de Y_i e $Q_i = Q(\alpha_i) = b^{*\prime\prime}(\vartheta_i) = \frac{d\alpha_i}{d\vartheta_i}$ é a função de variância de $X_i | Y_i$, w_i e r_i são pesos definidos por

$$w_i = \frac{1}{V_i} \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \quad \text{e} \quad r_i = \frac{1}{Q_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2, \quad i = 1, \dots, n.$$

Os elementos dos vetores $\mathbf{U}(\boldsymbol{\beta})$ e $\mathbf{U}(\boldsymbol{\Delta})$ são dados por

$$\begin{aligned} \frac{\partial \ell(\cdot)}{\partial \beta_{j_1}} &= \frac{1}{a(\phi)} \sum_{i=1}^n (y_i - \mu_{iY}) \frac{d\theta_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + \frac{1}{a^*(\varphi)} \sum_{i=1}^n (x_i - \alpha_i) \frac{d\vartheta_i}{d\alpha_i} \frac{d\alpha_i}{d\xi_i} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \quad \text{e} \\ \frac{\partial \ell(\cdot)}{\partial \delta_{j_2}} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n (x_i - \alpha_i) \frac{d\vartheta_i}{d\alpha_i} \frac{d\alpha_i}{d\xi_i} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}}, \end{aligned}$$

respectivamente, para $j_1 = 0, 1, \dots, p$ e $j_2 = 0, 1, \dots, q$, enquanto $\mathbf{U}(\gamma)$, $\mathbf{U}(\phi)$ e $\mathbf{U}(\varphi)$ pode ser reescrito como

$$\begin{aligned} \frac{\partial \ell(\cdot)}{\partial \gamma} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n (x_i - \alpha_i) \frac{d\vartheta_i}{d\alpha_i} \frac{d\alpha_i}{d\xi_i} \frac{d\xi_i}{\partial \gamma}, \\ \frac{\partial \ell(\cdot)}{\partial \phi} &= -\frac{a'(\phi)}{a(\phi)^2} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c'(y_i, \phi) \quad \text{e} \quad \frac{\partial \ell(\cdot)}{\partial \varphi} = -\frac{a^{*\prime}(\varphi)}{a^*(\varphi)^2} \sum_{i=1}^n [x_i \vartheta_i - b^*(\vartheta_i)] + \sum_{i=1}^n c^{*\prime}(x_i, \varphi), \end{aligned}$$

respectivamente, em que $a'(\phi) = \frac{d[a(\phi)]}{d\phi}$, $a^{*\prime}(\varphi) = \frac{d[a^*(\varphi)]}{d\varphi}$, $c'(y_i, \phi) = \frac{d[c(y_i, \phi)]}{d\phi}$ e $c^{*\prime}(x_i, \varphi) = \frac{d[c^*(x_i, \varphi)]}{d\varphi}$.

A derivada de segunda ordem da função log-verossimilhança com respeito a $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, γ , ϕ e

φ , em forma matricial, é dada pelas equações (2.11),

$$\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \beta^\top} = \mathbf{S}_1^\top \left\{ \frac{1}{a(\phi)} [-\mathbf{W} + \mathbf{D}_{(y-\mu_Y)} (\mathbf{A}_\theta \mathbf{W} \mathbf{V} + \mathbf{V}^{-1} \mathbf{A}_{\mu_Y})] + \frac{1}{a^*(\varphi)} [-\mathbf{R} \mathbf{P}_Y^2 \mathbf{W} \mathbf{V} + \mathbf{D}_{(x-\alpha)} (\mathbf{A}_\theta \mathbf{R} \mathbf{Q} \mathbf{P}_Y^2 \mathbf{W} \mathbf{V} + \mathbf{Q}^{-1} (\mathbf{A}_\alpha \mathbf{P}_Y^2 \mathbf{W} \mathbf{V} + \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} (d\mathbf{P}_Y \mathbf{W} \mathbf{V} + \mathbf{P}_Y \mathbf{A}_{\mu_Y})))] \right\} \mathbf{S}_1 + \frac{1}{a(\phi)} \mathbf{A}^* + \frac{1}{a^*(\varphi)} \mathbf{A}^{**}, \quad (\text{A.1})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \Delta^\top} = \frac{1}{a^*(\varphi)} \mathbf{S}_2^\top \left\{ -\mathbf{R} \mathbf{P}_X^2 \mathbf{O}^2 + \mathbf{D}_{(x-\alpha)} [\mathbf{A}_\theta \mathbf{R} \mathbf{Q} \mathbf{P}_X^2 \mathbf{O}^2 + \mathbf{Q}^{-1} (\mathbf{A}_\alpha \mathbf{P}_X^2 \mathbf{O}^2 + \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} (d\mathbf{P}_X \mathbf{O}^2 + \mathbf{P}_X \mathbf{A}_{\mu_X}))] \right\} \mathbf{S}_2 + \frac{1}{a^*(\varphi)} \mathbf{A}^{***}, \quad (\text{A.2})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \gamma^2} = \frac{1}{a^*(\varphi)} \left\{ \mathbf{P}_\gamma^\top [-\mathbf{R} + \mathbf{D}_{(x-\alpha)} (\mathbf{A}_\theta \mathbf{R} \mathbf{Q} + \mathbf{Q}^{-1} \mathbf{A}_\alpha)] \mathbf{J}_\gamma + (x - \alpha)^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} d\mathbf{J}_\gamma \right\}, \quad (\text{A.3})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \phi^2} = \left\{ \frac{2}{a(\phi)^3} [a'(\phi)]^2 - \frac{a''(\phi)}{a(\phi)^2} \right\} (\mathbf{y}^\top \boldsymbol{\theta} - \mathbf{b} \mathbf{1}) + \mathbf{c}' \mathbf{1}, \quad (\text{A.4})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \varphi^2} = \left\{ \frac{2}{a^*(\varphi)^3} [a^*(\varphi)]^2 - \frac{a^{**}(\varphi)}{a^*(\varphi)^2} \right\} (\mathbf{x}^\top \boldsymbol{\vartheta} - \mathbf{b}^* \mathbf{1}) + \mathbf{c}^* \mathbf{1}, \quad (\text{A.5})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \Delta^\top} = \frac{1}{a^*(\varphi)} \mathbf{S}_1^\top \left\{ -\mathbf{R} \mathbf{P}_X \mathbf{O} \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} + \mathbf{D}_{(x-\alpha)} [\mathbf{A}_\theta \mathbf{R} \mathbf{Q} \mathbf{P}_X \mathbf{O} \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} + \mathbf{Q}^{-1} (\mathbf{A}_\alpha \mathbf{P}_X \mathbf{O} \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} + \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} (d\mathbf{J}_{XY} \mathbf{O} \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}}))] \right\} \mathbf{S}_2, \quad (\text{A.6})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \gamma} = \frac{1}{a^*(\varphi)} \left\{ \mathbf{S}_1^\top [-\mathbf{R} \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} + \mathbf{D}_{(x-\alpha)} (\mathbf{A}_\theta \mathbf{R} \mathbf{Q} \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} + \mathbf{Q}^{-1} \mathbf{A}_\alpha \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}})] \mathbf{J}_\gamma + \mathbf{S}_1^\top \mathbf{D}_{(x-\alpha)} \mathbf{Q}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} d\mathbf{J}_{\gamma Y} \right\}, \quad (\text{A.7})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \phi} = -\frac{1}{a(\phi)^2} a'(\phi) \mathbf{S}_1^\top \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \mu_Y), \quad (\text{A.8})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \varphi} = -\frac{1}{a^*(\varphi)^2} a^*(\varphi) \mathbf{S}_1^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} \mathbf{P}_Y \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{\frac{1}{2}} (\mathbf{x} - \alpha), \quad (\text{A.9})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \gamma} = \frac{1}{a^*(\varphi)} \left\{ \mathbf{S}_2^\top [-\mathbf{R} \mathbf{P}_X \mathbf{O} + \mathbf{D}_{(x-\alpha)} (\mathbf{A}_\theta \mathbf{R} \mathbf{Q} \mathbf{P}_X \mathbf{O} + \mathbf{Q}^{-1} \mathbf{A}_\alpha \mathbf{P}_X \mathbf{O})] \mathbf{J}_\gamma + \mathbf{S}_2^\top \mathbf{D}_{(x-\alpha)} \mathbf{Q}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{\frac{1}{2}} \mathbf{O} d\mathbf{J}_{\gamma X} \right\}, \quad (\text{A.10})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \varphi} = -\frac{1}{a^*(\varphi)^2} a^*(\varphi) \mathbf{S}_2^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} \mathbf{P}_X \mathbf{O} (\mathbf{x} - \alpha), \quad (\text{A.11})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \gamma \partial \varphi} = -\frac{1}{a^*(\varphi)^2} a^*(\varphi) (\mathbf{x} - \alpha)^\top \mathbf{R}^{\frac{1}{2}} \mathbf{Q}^{-\frac{1}{2}} \mathbf{J}_\gamma, \quad (\text{A.12})$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \phi} = \mathbf{0}_{(q+1) \times 1} \quad \text{e} \quad \frac{\partial^2 \ell(\cdot)}{\partial \gamma \partial \phi} = \frac{\partial^2 \ell(\cdot)}{\partial \phi \partial \varphi} = 0, \quad (\text{A.13})$$

em que

$$\begin{aligned}
\mathbf{A}_\theta &= \text{diag} \left(\frac{d^2\theta_1}{d\mu_{1Y}^2}, \dots, \frac{d^2\theta_n}{d\mu_{nY}^2} \right), \quad \mathbf{A}_{\mu_Y} = \text{diag} \left(\frac{d^2\mu_{1Y}}{d\eta_{1Y}^2}, \dots, \frac{d^2\mu_{nY}}{d\eta_{nY}^2} \right), \quad \mathbf{A}_\alpha = \text{diag} \left(\frac{d^2\alpha_1}{d\alpha_1^2}, \dots, \frac{d^2\alpha_n}{d\alpha_n^2} \right), \\
\mathbf{A}_\alpha &= \text{diag} \left(\frac{d^2\alpha_1}{d\xi_1^2}, \dots, \frac{d^2\alpha_n}{d\xi_n^2} \right), \quad \mathbf{A}_{\mu_X} = \text{diag} \left(\frac{d^2\mu_{1X}}{d\eta_{1X}^2}, \dots, \frac{d^2\mu_{nX}}{d\eta_{nX}^2} \right), \quad \mathbf{dP}_Y = \text{diag} \left(\frac{d^2\xi_1}{d\mu_{1Y}^2}, \dots, \frac{d^2\xi_n}{d\mu_{nY}^2} \right), \\
\mathbf{dP}_X &= \text{diag} \left(\frac{d^2\xi_1}{d\mu_{1X}^2}, \dots, \frac{d^2\xi_n}{d\mu_{nX}^2} \right), \quad \mathbf{dJ}_{XY} = \text{diag} \left(\frac{d^2\xi_1}{d\mu_{1X}d\mu_{1Y}}, \dots, \frac{d^2\xi_n}{d\mu_{nX}d\mu_{nY}} \right), \quad \mathbf{R}^{\frac{1}{2}} = \text{diag}(r_1^{\frac{1}{2}}, \dots, r_n^{\frac{1}{2}}), \\
\mathbf{W}^{\frac{1}{2}} &= \text{diag}(w_1^{\frac{1}{2}}, \dots, w_n^{\frac{1}{2}}), \quad \mathbf{V}^{\frac{1}{2}} = \text{diag}(V_1^{\frac{1}{2}}, \dots, V_n^{\frac{1}{2}}), \quad \mathbf{Q}^{\frac{1}{2}} = \text{diag}(Q_1^{\frac{1}{2}}, \dots, Q_n^{\frac{1}{2}}), \quad \mathbf{V}^{-\frac{1}{2}} = \text{diag}(V_1^{-\frac{1}{2}}, \dots, V_n^{-\frac{1}{2}}), \\
\mathbf{Q}^{-\frac{1}{2}} &= \text{diag}(Q_1^{-\frac{1}{2}}, \dots, Q_n^{-\frac{1}{2}}) \text{ e } \mathbf{Q}^{-1} = \text{diag}(Q_1^{-1}, \dots, Q_n^{-1}),
\end{aligned}$$

são matrizes diagonal $n \times n$, $\mathbf{D}(\mathbf{y} - \boldsymbol{\mu}_Y)$ e $\mathbf{D}(\mathbf{x} - \boldsymbol{\alpha})$ são matrizes diagonal cujos elementos são formados, respectivamente pelos elementos dos vetores $(\mathbf{y} - \boldsymbol{\mu}_Y)$ e $(\mathbf{x} - \boldsymbol{\alpha})$,

$$\begin{aligned}
\mathbf{dJ}_{\gamma Y} &= \left(\frac{d^2\xi_1}{d\gamma d\mu_{1Y}}, \dots, \frac{d^2\xi_n}{d\gamma d\mu_{nY}} \right)^\top, \quad \mathbf{dJ}_{\gamma X} = \left(\frac{d^2\xi_1}{d\gamma d\mu_{1X}}, \dots, \frac{d^2\xi_n}{d\gamma d\mu_{nX}} \right)^\top, \quad \mathbf{dJ}_\gamma = \left(\frac{d^2\xi_1}{d\gamma^2}, \dots, \frac{d^2\xi_n}{d\gamma^2} \right)^\top, \\
\mathbf{c}' &= (c''(y_1, \phi), \dots, c''(y_n, \phi)) \text{ e } \mathbf{c}^{*\prime} = (c^{*''}(x_1, \varphi), \dots, c^{*''}(x_n, \varphi)),
\end{aligned}$$

são vetores n -dimensional,

$$\mathbf{A}^* = \begin{pmatrix} A_{00}^* & A_{01}^* & \dots & A_{0p}^* \\ A_{10}^* & A_{11}^* & \dots & A_{1p}^* \\ \vdots & \vdots & \ddots & \vdots \\ A_{p0}^* & A_{p1}^* & \dots & A_{pp}^* \end{pmatrix}, \quad \mathbf{A}^{**} = \begin{pmatrix} A_{00}^{**} & A_{01}^{**} & \dots & A_{0p}^{**} \\ A_{10}^{**} & A_{11}^{**} & \dots & A_{1p}^{**} \\ \vdots & \vdots & \ddots & \vdots \\ A_{p0}^{**} & A_{p1}^{**} & \dots & A_{pp}^{**} \end{pmatrix}, \quad \mathbf{A}^{***} = \begin{pmatrix} A_{00}^{***} & A_{01}^{***} & \dots & A_{0q}^{***} \\ A_{10}^{***} & A_{11}^{***} & \dots & A_{1q}^{***} \\ \vdots & \vdots & \ddots & \vdots \\ A_{q0}^{***} & A_{q1}^{***} & \dots & A_{qq}^{***} \end{pmatrix},$$

são matrizes em que

$$\begin{aligned}
A_{j_1 k_1}^* &= \sum_{i=1}^n (y_i - \mu_{iY}) \sqrt{\frac{w_i}{V_i}} \frac{\partial^2 \eta_{iY}}{\partial \beta_{j_1} \partial \beta_{k_1}}, \quad A_{j_1 k_1}^{**} = \sum_{i=1}^n (x_i - \alpha_i) \sqrt{\frac{r_i w_i V_i}{Q_i}} \frac{d\xi_i}{d\mu_{iY}} \frac{\partial^2 \eta_{iY}}{\partial \beta_{j_1} \partial \beta_{k_1}}, \quad \text{para } j_1, k_1 = 0, 1, \dots, p \text{ e} \\
A_{j_2 k_2}^{***} &= \sum_{i=1}^n (x_i - \alpha_i) \sqrt{\frac{r_i}{Q_i}} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial^2 \eta_{iX}}{\partial \delta_{j_2} \partial \delta_{k_2}}, \quad \text{para } j_2, k_2 = 0, 1, \dots, q,
\end{aligned}$$

$$a''(\phi) = \frac{d^2[a(\phi)]}{d\phi^2}, \quad c''(y_i, \phi) = \frac{d^2[c(y_i, \phi)]}{d\phi^2}, \quad a^{*''}(\varphi) = \frac{d^2[a^*(\varphi)]}{d\varphi^2} \text{ e } c^{*''}(x_i, \varphi) = \frac{d^2[c^*(x_i, \varphi)]}{d\varphi^2}.$$

Note que $\mathbf{A}^* = \mathbf{A}^{**} = \mathbf{0}$ e $\mathbf{A}^{***} = \mathbf{0}$ se η_{iY} e η_{iX} são preditores lineares, respectivamente.

Os elementos de $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \beta^\top}$, $\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \Delta^\top}$ e $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \Delta^\top}$ são matrizes dadas respectivamente por

$$\begin{aligned} \frac{\partial^2 \ell(\cdot)}{\partial \beta_{j_1} \partial \beta_{k_1}} &= \frac{1}{a(\phi)} \sum_{i=1}^n \left\{ -\frac{d\theta_i}{d\mu_{iY}} \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + (y_i - \mu_{iY}) \left[\frac{d^2 \theta_i}{d\mu_{iY}^2} \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \right. \\ &\quad \left. \left. + \frac{d\theta_i}{d\mu_{iY}} \frac{d^2 \mu_{iY}}{d\eta_{iY}^2} \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + \frac{d\theta_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial^2 \eta_{iY}}{\partial \beta_{k_1} \partial \beta_{j_1}} \right] \right\} \\ &\quad + \frac{1}{a^*(\varphi)} \sum_{i=1}^n \left\{ -\frac{d\vartheta_i}{d\alpha_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \left(\frac{d\xi_i}{d\mu_{iY}} \right)^2 \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \\ &\quad \left. + (x_i - \alpha_i) \left\{ \frac{d^2 \vartheta_i}{d\alpha_i^2} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \left(\frac{d\xi_i}{d\mu_{iY}} \right)^2 \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \right. \\ &\quad \left. \left. + \frac{d\vartheta_i}{d\alpha_i} \left[\frac{d^2 \alpha_i}{d\xi_i^2} \left(\frac{d\xi_i}{d\mu_{iY}} \right)^2 \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + \frac{d\alpha_i}{d\xi_i} \left(\frac{d^2 \xi_i}{d\mu_i^2} \left(\frac{d\mu_{iY}}{d\eta_{iY}} \right)^2 \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{d\xi_i}{d\mu_{iY}} \left[\frac{d^2 \mu_{iY}}{d\eta_{iY}^2} \frac{\partial \eta_{iY}}{\partial \beta_{k_1}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial^2 \eta_{iY}}{\partial \beta_{k_1} \partial \beta_{j_1}} \right] \right] \right] \right\} \right\}, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\cdot)}{\partial \delta_{j_2} \partial \delta_{k_2}} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n \left\{ -\frac{d\vartheta_i}{d\alpha_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \left(\frac{d\xi_i}{d\mu_{iX}} \right)^2 \left(\frac{d\mu_{iX}}{d\eta_{iX}} \right)^2 \frac{\partial \eta_{iX}}{\partial \delta_{k_2}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \right. \\ &\quad \left. + (x_i - \alpha_i) \left\{ \frac{d^2 \vartheta_i}{d\alpha_i^2} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \left(\frac{d\xi_i}{d\mu_{iX}} \right)^2 \left(\frac{d\mu_{iX}}{d\eta_{iX}} \right)^2 \frac{\partial \eta_{iX}}{\partial \delta_{k_2}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \right. \right. \\ &\quad \left. \left. + \frac{d\vartheta_i}{d\alpha_i} \left(\frac{d^2 \alpha_i}{d\xi_i^2} \left(\frac{d\xi_i}{d\mu_{iX}} \right)^2 \left(\frac{d\mu_{iX}}{d\eta_{iX}} \right)^2 \frac{\partial \eta_{iX}}{\partial \delta_{k_2}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} + \frac{d\alpha_i}{d\xi_i} \left(\frac{d^2 \xi_i}{d\mu_{iX}^2} \left(\frac{d\mu_{iX}}{d\eta_{iX}} \right)^2 \frac{\partial \eta_{iX}}{\partial \delta_{k_2}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \right. \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{d\xi_i}{d\mu_{iX}} \left(\frac{d^2 \mu_{iX}}{d\eta_{iX}^2} \frac{\partial \eta_{iX}}{\partial \delta_{k_2}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} + \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial^2 \eta_{iX}}{\partial \delta_{k_2} \partial \delta_{j_2}} \right) \right] \right] \right\} \right\} \text{ e} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \ell(\cdot)}{\partial \beta_{j_1} \partial \delta_{j_2}} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n \left\{ -\frac{d\vartheta_i}{d\alpha_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \\ &\quad \left. + (x_i - \alpha_i) \left\{ \frac{d^2 \vartheta_i}{d\alpha_i^2} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \right. \\ &\quad \left. \left. + \frac{d\vartheta_i}{d\alpha_i} \left[\frac{d^2 \alpha_i}{d\xi_i^2} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + \frac{d\alpha_i}{d\xi_i} \left(\frac{d^2 \xi_i}{d\mu_{iX} d\mu_{iY}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right) \right] \right\} \right\}, \end{aligned}$$

para $j_1, k_1 = 0, 1, \dots, p$ e $j_2, k_2 = 0, 1, \dots, q$, enquanto os elementos de $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \gamma}$, $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \phi}$, $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \varphi}$,

$\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \gamma}$ e $\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \varphi}$ são vetores dados respectivamente por

$$\begin{aligned} \frac{\partial^2 \ell(\cdot)}{\partial \beta_{j_1} \partial \gamma} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n \left\{ -\frac{d\vartheta_i}{d\alpha_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \frac{d\xi_i}{d\gamma} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + (x_i - \alpha_i) \left\{ \frac{d^2 \vartheta_i}{d\alpha_i^2} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \frac{d\xi_i}{d\gamma} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right. \right. \\ &\quad \left. \left. + \frac{d\vartheta_i}{d\alpha_i} \left[\frac{d^2 \alpha_i}{d\xi_i^2} \frac{d\xi_i}{d\gamma} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} + \frac{d\alpha_i}{d\xi_i} \left(\frac{d^2 \xi_i}{d\gamma d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}} \right) \right] \right\} \right\}, \\ \frac{\partial^2 \ell(\cdot)}{\partial \beta_{j_1} \partial \phi} &= -\frac{1}{a(\phi)^2} a'(\phi) \sum_{i=1}^n (y_i - \mu_{iY}) \frac{d\theta_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}}, \\ \frac{\partial^2 \ell(\cdot)}{\partial \beta_{j_1} \partial \varphi} &= -\frac{1}{a^*(\varphi)^2} a^{*\prime}(\varphi) \sum_{i=1}^n (x_i - \alpha_i) \frac{d\vartheta_i}{d\alpha_i} \frac{d\alpha_i}{d\xi_i} \frac{d\xi_i}{d\mu_{iY}} \frac{d\mu_{iY}}{d\eta_{iY}} \frac{\partial \eta_{iY}}{\partial \beta_{j_1}}, \\ \frac{\partial^2 \ell(\cdot)}{\partial \delta_{j_2} \partial \gamma} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n \left\{ -\frac{d\vartheta_i}{d\alpha_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \frac{d\xi_i}{d\gamma} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} + (x_i - \alpha_i) \left\{ \frac{d^2 \vartheta_i}{d\alpha_i^2} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \frac{d\xi_i}{d\gamma} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \right. \right. \\ &\quad \left. \left. + \frac{d\vartheta_i}{d\alpha_i} \left[\frac{d^2 \alpha_i}{d\xi_i^2} \frac{d\xi_i}{d\gamma} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} + \frac{d\alpha_i}{d\xi_i} \left(\frac{d^2 \xi_i}{d\gamma d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}} \right) \right] \right\} \right\} \text{ e} \\ \frac{\partial^2 \ell(\cdot)}{\partial \delta_{j_2} \partial \varphi} &= -\frac{1}{a^*(\varphi)^2} a^{*\prime}(\varphi) \sum_{i=1}^n (x_i - \alpha_i) \frac{d\vartheta_i}{d\alpha_i} \frac{d\alpha_i}{d\xi_i} \frac{d\xi_i}{d\mu_{iX}} \frac{d\mu_{iX}}{d\eta_{iX}} \frac{\partial \eta_{iX}}{\partial \delta_{j_2}}, \end{aligned}$$

para $j_1 = 0, 1, \dots, p$ e $j_2 = 0, 1, \dots, q$. As equações (A.3), (A.4), (A.5) e (A.12) pode ser reescritas como

$$\begin{aligned} \frac{\partial^2 \ell(\cdot)}{\partial \gamma^2} &= \frac{1}{a^*(\varphi)} \sum_{i=1}^n \left\{ -\frac{d\vartheta_i}{d\alpha_i} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \left(\frac{d\xi_i}{d\gamma} \right)^2 + (x_i - \alpha_i) \left\{ \frac{d^2 \vartheta_i}{d\alpha_i^2} \left(\frac{d\alpha_i}{d\xi_i} \right)^2 \left(\frac{d\xi_i}{d\gamma} \right)^2 + \frac{d\vartheta_i}{d\alpha_i} \left(\frac{d^2 \alpha_i}{d\xi_i^2} \left(\frac{d\xi_i}{d\gamma} \right)^2 + \frac{d\alpha_i}{d\xi_i} \frac{d^2 \xi_i}{d\gamma^2} \right) \right\} \right\}, \\ \frac{\partial^2 \ell(\cdot)}{\partial \phi^2} &= \left\{ \frac{2}{a(\phi)^3} [a'(\phi)]^2 - \frac{a''(\phi)}{a(\phi)^2} \right\} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n [c''(y_i, \phi)], \\ \frac{\partial^2 \ell(\cdot)}{\partial \varphi^2} &= \left\{ \frac{2}{a^*(\varphi)^3} [a^{*\prime}(\varphi)]^2 - \frac{a^{*\prime\prime}(\varphi)}{a^*(\varphi)^2} \right\} \sum_{i=1}^n [x_i \vartheta_i - b^*(\vartheta_i)] + \sum_{i=1}^n [c^{*\prime\prime}(x_i, \varphi)] \text{ e} \\ \frac{\partial^2 \ell(\cdot)}{\partial \gamma \partial \varphi} &= -\frac{1}{a^*(\varphi)^2} a^{*\prime}(\varphi) \sum_{i=1}^n (x_i - \alpha_i) \frac{d\vartheta_i}{d\alpha_i} \frac{d\alpha_i}{d\xi_i} \frac{d\xi_i}{d\gamma}, \text{ respectively.} \end{aligned}$$

A matriz de informação observada é dada por

$$M = \begin{bmatrix} M_\beta & M_{\beta\Delta} & M_{\beta\gamma} & M_{\beta\phi} & M_{\beta\varphi} \\ M_{\beta\Delta}^\top & M_\Delta & M_{\Delta\gamma} & M_{\Delta\phi} & M_{\Delta\varphi} \\ M_{\beta\gamma}^\top & M_{\Delta\gamma}^\top & M_\gamma & M_{\gamma\phi} & M_{\gamma\varphi} \\ M_{\beta\phi}^\top & M_{\Delta\phi}^\top & M_{\gamma\phi}^\top & M_\phi & M_{\phi\varphi} \\ M_{\beta\varphi}^\top & M_{\Delta\varphi}^\top & M_{\gamma\varphi}^\top & M_{\phi\varphi}^\top & M_\varphi \end{bmatrix} = \begin{bmatrix} M_\beta & M_{\beta\Delta} & M_{\beta\gamma} & M_{\beta\phi} & M_{\beta\varphi} \\ M_{\beta\Delta}^\top & M_\Delta & M_{\Delta\gamma} & \mathbf{0}_{(q+1) \times 1} & M_{\Delta\varphi} \\ M_{\beta\gamma}^\top & M_{\Delta\gamma}^\top & M_\gamma & 0 & M_{\gamma\varphi} \\ M_{\beta\phi}^\top & \mathbf{0}_{1 \times (q+1)} & 0 & M_\phi & 0 \\ M_{\beta\varphi}^\top & M_{\Delta\varphi}^\top & M_{\gamma\varphi}^\top & 0 & M_\varphi \end{bmatrix},$$

em que

$$\begin{aligned}
M_{\beta} &= -\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \beta^{\top}}, \quad M_{\Delta} = -\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \Delta^{\top}}, \quad M_{\beta \Delta} = -\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \Delta^{\top}}, \quad M_{\beta \gamma} = -\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \gamma}, \quad M_{\beta \phi} = -\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \phi}, \\
M_{\beta \varphi} &= -\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \varphi}, \quad M_{\Delta \gamma} = -\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \gamma}, \quad M_{\Delta \varphi} = -\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \varphi}, \quad M_{\gamma} = -\frac{\partial^2 \ell(\cdot)}{\partial \gamma^2}, \quad M_{\phi} = -\frac{\partial^2 \ell(\cdot)}{\partial \phi^2}, \\
M_{\varphi} &= -\frac{\partial^2 \ell(\cdot)}{\partial \varphi^2} \text{ e } M_{\gamma \varphi} = -\frac{\partial^2 \ell(\cdot)}{\partial \gamma \partial \varphi},
\end{aligned} \tag{A.14}$$

com $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \beta^{\top}}$, $\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \Delta^{\top}}$, $\frac{\partial^2 \ell(\cdot)}{\partial \gamma^2}$, $\frac{\partial^2 \ell(\cdot)}{\partial \phi^2}$, $\frac{\partial^2 \ell(\cdot)}{\partial \varphi^2}$, $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \Delta^{\top}}$, $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \gamma}$, $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \phi}$, $\frac{\partial^2 \ell(\cdot)}{\partial \beta \partial \varphi}$, $\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \gamma}$, $\frac{\partial^2 \ell(\cdot)}{\partial \Delta \partial \varphi}$ e $\frac{\partial^2 \ell(\cdot)}{\partial \gamma \partial \varphi}$ dadas pelas equações (A.1), (A.2), (A.3), (A.4), (A.5), (A.6), (A.7), (A.8), (A.9), (A.10), (A.11) e (A.12), respectivamente.

Bibliografia

- Balakrishnan, N. & Lai, C.-D. (2009). *Continuous bivariate distributions*. Springer, Dordrecht, second edition.
- Beilei, W. (2013). *Contributions to Copula Modeling of Mixed Discrete-Continuous Outcomes*. Ph.D. thesis, University of Calgary.
- Berkhout, P. & Plug, E. (2004). A bivariate Poisson count data model using conditional probabilities. *Statistica Neerlandica*, **58**, 349–364.
- Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Buja, A., Hastie, T. & Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, **17**, 453–510.
- Catalano, P. J. & Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **87**, 651–658.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.

-
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B*, **48**, 133–169. With discussion.
- Cox, D. R. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society. Series C*, **21**, 113–120.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B*, **30**, 248–275.
- Cox, D. R. & Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika*, **79**, 441–461.
- Currie, I. D. & Durban, M. (2002). Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, **2**, 333–349.
- Currie, I. D., Durban, M. & Eilers, P. H. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 259–280.
- de Boor, C. (1977). Package for calculating with *B*-splines. **14**, 441–472.
- de Leon, A. R. (2005). Pairwise likelihood approach to grouped continuous model and its extension. **75**, 49–57.
- de Leon, A. R. & Carrière, K. C. (2007). General mixed-data model: extension of general location and grouped continuous models. **35**, 533–548.
- de Leon, A. R. & Chough, K. C., editors (2013). *Analysis of mixed data*. CRC Press, Boca Raton, FL. Methods & applications.
- de Leon, A. R. & Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. **30**, 175–185.

-
- Demetrio, C. (2002). Modelos lineares generalizados em experimentação agrícola. 1^o. ed. Piracicaba, SP: ESALQ/USP.
- Dierckx, P. (1993). *Curve and surface fitting with splines*. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications.
- Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236–244.
- Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B*, **62**, 355–366.
- Durbán, M. & Currie, I. D. (2003). A note on p-spline additive models with correlated errors. *Computational Statistics*, **18**, 251–262.
- Durbán, M., Currie, I. & Eilers, P. (2006). Mixed models, array methods and multidimensional density estimation. In *Proceedings of the 21st International Workshop of Statistical Modelling*. Citeseer.
- Eilers, P. H. C. & Marx, B. D. (1996). Flexible smoothing with *B*-splines and penalties. *Statist. Sci.*, **11**, 89–121. With comments and a rejoinder by the authors.
- Eilers, P. H. C. & Marx, B. D. (2002). Generalized linear additive smooth structures. *Journal of Computational and Graphical Statistics*, **11**, 758–783.
- Fitzmaurice, G. M. & Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association*, **90**, 845–852.
- Fitzmaurice, G. M. & Laird, N. M. (1997). Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics*, **53**, 110–122.

-
- George, E. O., Armstrong, D., Catalano, P. J. & Srivastava, D. K. (2007). Regression models for analyzing clustered binary and continuous outcomes under an assumption of exchangeability. *Journal of Statistical Planning and Inference*, **137**, 3462–3474.
- Gomes, E. M. d. C. (2007). *Análise de sensibilidade e resíduos em modelos de regressão com respostas bivariadas por meio de cópulas*. Ph.D. thesis, Universidade de São Paulo.
- Good, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics*, **11**, 823–831.
- Green, P. J. & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman & Hall, London.
- Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, **1**, 177–193.
- Gueorguieva, R. V. & Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, **96**, 1102–1112.
- Hastie, T. & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, **82**, 371–386.
- Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. Chapman and Hall, Ltd., London. ISBN 0-412-34390-8.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (2002). *Continuous Multivariate Distributions, volume 1, Models and Applications*. New York: John Wiley & Sons.
- Jung, R. C. & Winkelmann, R. (1993). Two aspects of labor mobility: a bivariate poisson regression approach. *Empirical Economics*, **18**, 543–556.

-
- Khafri, S., Kazemnejad, A. & Eskandari, F. (2008). Hierarchical bayesian analysis of bivariate poisson regression model. *World Applied Sciences Journal*, **4**, 667–675.
- Lauritzen, S. L. & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, **17**, 31–57.
- Lee, S.-Y., Lu, B. & Song, X.-Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics & Data Analysis*, **50**, 1356–1377.
- Lesaffre, E. & Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.
- Lin, L., Bandyopadhyay, D., Lipsitz, S. R. & Sinha, D. (2010). Association models for clustered data with binary and continuous responses. *Biometrics*, **66**, 287–293.
- Marx, B. D. & Eilers, P. H. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, **28**, 193–209.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall, London.
- McDonald, B. W. (1993). Estimating logistic regression parameters for bivariate binary data. *Journal of the Royal Statistical Society. Series B*, **55**, 391–397.
- Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A*, **135**, 370–384.
- Olkin, I. & Tate, R. (1961). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*, **32**, 448–465.

-
- Ortega, E. M., Bolfarine, H. & Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics & Data Analysis*, **42**, 165–186.
- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, **1**, 502–527.
- Paula, G. A. (2004). *Modelos de Regressão com apoio computacional*. IME-USP, São Paulo:.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J. & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, **27**, 85–96.
- Ramsay, T. O., Burnett, R. T. & Krewski, D. (2003). The effect of concavity in generalized additive models linking mortality to ambient particulate matter. *Epidemiology*, **14**, 18–23.
- Ruppert, D., Wand, M. P. & Carroll, R. J. (2003). *Semiparametric regression*. Cambridge University Press, Cambridge.
- Samani, E. B. & Ganjali, M. (2014). Mixed correlated bivariate ordinal and negative binomial longitudinal responses with nonignorable missing values. *Communications in Statistics-Theory and Methods*, **43**, 2659–2673.
- Scollnik, D. P. (2002). Regression models for bivariate loss data. *North American Actuarial Journal*, **6**, 67–80.
- Shults, J., Sun, W., Tu, X., Kim, H., Amsterdam, J., Hilbe, J. M. & Ten-Have, T. (2009). A comparison of several approaches for choosing between working correlation

-
- structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine*, **28**, 2338–2355.
- Song, J., Barnhart, H. X. & Lyles, R. H. (2004). A GEE approach for estimating correlation coefficients involving left-censored variables. *Journal of Data Science*, **2**, 245–257.
- Tang, N.-S., Wei, B.-C. & Zhang, W.-Z. (2006). Influence diagnostics in nonlinear reproductive dispersion mixed models. *Statistics*, **40**, 227–246.
- Tate, R. F. (1954). Correlation between a discrete and a continuous variable. Point-biserial correlation. *Annals of Mathematical Statistics*, **25**, 603–607.
- Thiébaud, R., Jacqmin-Gadda, H., Chêne, G., Leport, C. & Commenges, D. (2002). Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine*, **69**, 249–256.
- Thomas, W. & Cook, R. D. (1989). Assessing influence on regression coefficients in generalized linear models. *Biometrika*, **76**, 741–749.
- Thomas, W. & Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics*, **32**, 59–65.
- van Ophem, H. (1999). A general method to estimate correlated discrete random variables. *Econometric Theory*, **15**, 228–237.
- Verbeke, G. & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag, New York.
- Whittaker, E. T. (1922). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, **41**, 63–75.
- Williams, D. (1984). Residuals in generalized linear models. In *Proc. 12th International Biometrics Conference, Tokyo*, pages 59–68.

- Yang, Y. & Kang, J. (2010). Joint analysis of mixed poisson and continuous longitudinal data with nonignorable missing values. *Computational Statistics & Data Analysis*, **54**, 193–207.
- Yang, Y., Kang, J., Mao, K. & Zhang, J. (2007). Regression models for mixed poisson and continuous longitudinal data. *Statistics in Medicine*, **26**, 3782–3800.
- Zhu, H., Ibrahim, J. G., Lee, S., Zhang, H. *et al.* (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics*, **35**, 2565–2588.