Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

Yuri Alves de Araujo

Modelos Não Lineares Parciais Generalizados Superdispersados

Recife
2017

Yuri Alves de Araujo

Modelos Não Lineares Parciais Generalizados Superdispersados

Dissertação apresentada ao programa de Pós-graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Estatística. Área de Concentração: Estatística Aplicada.

ORIENTADOR: **Profa. Dra. Audrey Helen Mariz de Aquino Cysneiros**

Recife
2017

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

A663m Araujo, Yuri Alves de
Modelos não lineares parciais generalizados superdispersados / Yuri Alves de Araujo. – 2017.
72 f.: il., fig., tab.

Orientadora: Audrey Helen Mariz de Aquino Cysneiros.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências e apêndices.

1. Análise de regressão. 2. Modelos de regressão. I. Cysneiros, Audrey Helen Mariz de Aquino (orientadora). II. Título.

519.536 CDD (23. ed.) UFPE- MEI 2017-75

YURI ALVES DE ARAÚJO

MODELOS NÃO LINEARES PARCIAIS GENERALIZADOS SUPERDISPERSADOS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 20 de fevereiro de 2017.

BANCA EXAMINADORA

Prof.ª Audrey Helen Mariz de Aquino Cysneiros
UFPE

Prof. Francisco José de Azevêdo Cysneiros
UFPE

Prof. Roberto Ferreira Manghi
UFC

Dedico esse trabalho

Em memória do meu pai Wanderley, por todo amor, carinho, dedicação, companheirismo e amizade.
Saudades eternas.

Agradecimentos

Gostaria de agradecer primeiramente a Deus, por me amparar nos momentos difíceis, me dar forças interior para superar as dificuldades, mostrar os caminhos nas horas incertas e me suprir em todas as minhas necessidades.

A meus pais, Wanderley e Elianice, meu infinito agradecimento. Sempre acreditaram em minha capacidade, isso só me fortaleceu ao longo dos anos e fez com que eu chegasse onde estou. Agradeço-lhes a dedicação nos meus estudos, o apoio em minhas decisões, me orientando e incentivando a sempre buscar o melhor. Obrigado pelo amor incondicional.

Ao meu irmão Ygor e à toda minha família, a qual tenho o mais sincero respeito, carinho e admiração. Obrigado pela força de todos e por sempre me ajudarem, vocês tem grande peso em minha vida.

Quero agradecer a minha noiva Rethiele, por está sempre ao meu lado me encorajando a seguir meus sonhos e alcançar meus objetivos. Sou extremamente grato a todo o seu companheirismo, carinho e compreensão, sem seu apoio não seria possível superar todas as dificuldades que encontrei no caminho.

À minha orientadora Profa. Audrey Helen Mariz de Aquino Cysneiros, por acreditar no meu potencial, sempre dedicada a me incentivar e proporcionar conhecimentos valiosos. Por todo esse tempo sempre esteve disponível e disposta a ajudar, sendo certamente uma referência profissional e pessoal para meu crescimento. Muito obrigado pela dedicação e por todos os seus ensinamentos.

Agradeço ao prof. Juvêncio Santos Nobre, por todo apoio e orientação ao longo da minha graduação, sempre me motivou a ingressar na pós-graduação e a crescer profissionalmente. Meu muito obrigado!

À Katia e sua família, por todo carinho e cuidados nos momentos em que mais precisei, obrigado por me acolherem. Sou extremamente grato por todo apoio e paciência que tiveram comigo ao longo desses 2 anos.

Agradeço a todos os professores do CCEN, e em especial aos Professores Alex Dias Ramos, Audrey Helen Mariz de Aquino Cysneiros, Francisco Cribari Neto, Francisco José de Azevêdo Cysneiros e Gauss Moutinho Cordeiro. Sem essa grande equipe de profissionais o meu esforço não teria se concretizado, vocês contribuíram muito para minha formação.

Aos meus amigos de Mestrado, pelo companheirismo durante toda essa fase da minha vida e por todos os momentos divididos juntos, especialmente à Vinícius, Olivia, Alejandro, Laura, Hildemar, Jucelino e Rodney. Espero que nossas amizades durem por muito tempo e que sempre tenhamos boas estórias para recordar. Foi muito bom poder contar com vocês!

Quero agradecer também a todos meus amigos de longa data e conterrâneos de Quixadá, principalmente ao Pedro Janssen (PJ), Ramirez, Paulo Rogério (Chinês) e Ernand, por só quererem o meu bem e me valorizarem tanto como pessoa e como amigo. Obrigado a todos pela amizade de todas as horas.

Aos servidores da Universidade Federal de Pernambuco (UFPE), principalmente a Valéria Bittencourt e Elaine Mota, pela paciência e cuidado ao longo desses dois anos de mestrado.

Quero agradecer também à UFPE por me oferecer ensino de qualidade e todo o suporte necessário para a conclusão dessa dissertação e ao apoio financeiro da Capes durante todo desenvolvimento deste trabalho.

Por fim, quero agradecer à todas as pessoas que contribuíram diretamente e indiretamente para meu sucesso e crescimento pessoal e profissional.

"Life is what happens to you while you're busy making other plans."

John Lennon

"Happiness is only real when shared."

Christopher Mccandless

Resumo

Os modelos de regressão são amplamente utilizados quando desejamos avaliar o comportamento de uma ou mais características de interesse (variáveis respostas), em função de outras características observadas (variáveis explicativas). No entanto, os modelos usuais em geral são bastante restritivos e, naturalmente, ocorre uma busca por modelos cada vez mais flexíveis. Neste contexto, Dey et al. (1997) propõem uma classe de modelos lineares generalizados superdispersados, os quais tem a capacidade de controlar a variabilidade modelando também sua dispersão de forma independente de sua média. Por outro lado, classes de modelos semiparamétricos estão cada vez mais relevantes na literatura, visto que estes apresentam grande flexibilidade na relação entre a variável resposta e suas correspondentes variáveis explicativas. Nesta dissertação estendemos a classe de modelos superdispersados propostos por Dey et al. (1997) para o âmbito semiparamétrico, ao considerar que a média e a dispersão da variável resposta dependem de componentes paramétricos não lineares e de componentes não paramétricos. Propomos um processo de estimação conjunto dos parâmetros do modelo, e adicionalmente, um critério para a seleção dos parâmetros associados à suavidade das funções não paramétricas. Desenvolvemos técnicas de diagnóstico baseadas em medidas de alavancagem, análise de resíduos e influência local. Na análise de influência local, foram considerados três esquemas de perturbação: perturbação na variável resposta, perturbação nos preditores e ponderação de casos. Por fim, foram realizadas implementações computacionais das técnicas de diagnóstico com o auxilio do *software* `R`, as quais são relacionadas com propostas de aplicações práticas envolvendo análise de dados reais.

Palavras-chave: Influência local. Métodos de diagnóstico. Modelos com superdispersão. Modelos de regressão semiparamétricos.

Abstract

Regression models are widely used when we want to evaluate the behavior of one or more characteristics of interest (response variables), according to other observed characteristics (explanatory variables). However, usual models are so restrictive and, naturally, a search for models is becoming increasingly flexible. In this context, Dey et al. (1997) proposed a class of overdispersed generalized linear models, which has the capacity to controls variability the also modeling a dispersion independently of the mean. On the other hand, are increasingly relevant in literature, since they have great flexibility in the relationship between the response variable and their corresponding explanatory variables. In this work, we extend to the class of models proposed by Dey et al. (1997) for semiparametric context, considering that the mean and the dispersion for response variable depend on nonlinear parametric components and nonparametric components. We propose the joint parameter estimation process, and in addition, a selection criteria for the smooth parameters associated with nonparametric functions. We develop diagnostic techniques based on leverage measures, residuals analisys and local influence under different perturbation schemes. Finally, applications to real data are presented.

Keywords: Diagnostic methods. Local influence. Overdispersion models. Semiparametric regression models.

Lista de Figuras

Lista de Tabelas

Sumário

CAPÍTULO 1

Introdução

1.1 Motivação

Comumente nos deparamos com situações práticas nas quais temos o interesse em avaliar o comportamento de uma ou mais variáveis respostas em função de outras variáveis observadas (variáveis explicativas), isto é, modelar uma possível relação entre variáveis de interesse. Em geral, esse comportamento é estudado por meio da média da distribuição condicional da variável resposta dado as variáveis explicativas. Essa esperança é então associada a um preditor, que em grande parte das vezes é representado em função de parâmetros. Vale ressaltar que é usual utilizar o modelo linear clássico, o qual considera uma relação linear entre a variável resposta e as explicativas e também que os erros associados ao modelo segue uma distribuição normal. No entanto, por diversos fatores, esse tipo de modelagem pode não ser viável, e consequentemente causar interpretações errôneas. Uma forma de contornar a situação é utilizar técnicas para transformação de variáveis em busca da normalidade, mas eventualmente torna o modelo mais difícil de interpretar ou acaba não implicando na normalidade da variável transformada.

Por esse fato, alguns pesquisadores da área vêm propondo metodologias mais sofisticadas e que generalizam essa modelagem clássica, gerando classes de modelos bem mais flexíveis. Uma dessas classes na qual merece destaque são os Modelos Lineares Generalizados (MLGs), proposta por Nelder e Wedderburn (1972), baseando-se na suposição que a componente aleatória pertence a família de distribuições denominada família exponencial. Além disso, a relação entre a variável resposta e suas correspondentes variáveis explicativas não necessariamente devem ser lineares, basta apenas que seja modelada através de uma estrutura linearizável. Discussões mais abrangentes acerca dessa classe de modelos podem ser encontrados em Paula (2004), Cordeiro e Demetrio (1986) e Turkman e Silva (2000), por exemplo.

Por outro lado, mesmo nessa classe que possui um leque de opções bem amplo, há a presença de problemas do ponto de vista prático. Um problema ocasional em análise de regressão se dá quando a variabilidade predita pelo ajuste do modelo é maior do que sua variabilidade teórica, ou em casos que até mesmo a sua variabilidade não seja constante. A esse evento dá-se o nome de superdispersão, o qual vem sendo largamente discutido por inúmeros autores. Formas de resolver esse problema são discutidas por Efron (1986), Smyth (1989) e Dey et al. (1997), considerando uma modelagem também para o parâmetro de escala do modelo em questão. Especificamente, Dey et al. (1997) apresentam uma generalização à família exponencial dupla introduzida por Efron (1986), para inserir um parâmetro que controla a variância do modelo, e assim obter uma modelagem para a média e variância de forma independente.

Uma outra discussão importante é no que diz respeito à estrutura funcional do modelo, ou seja, de que forma a variável resposta está relacionada às suas covariáveis. Em muitos casos, atribuir uma forma linear ou linearizável a essa relação pode não ser satisfatória o suficiente. Uma alternativa é considerar funções não lineares, porém esse caso ainda é bastante restritivo e impõe um conhecimento prévio da relação funcional. Em contrapartida, o uso de funções de suavização para a construção de modelos de regressão não paramétricos permite estudar relações sem necessariamente conhecer sua estrutura funcional. Do ponto de vista prático, a relação entre algumas variáveis podem ser simples de identificar e outras não, motivando o uso de métodos distintos relacionados a cada variável. Como forma de unir essas duas metodologias de análise na busca para a melhoria do ajuste, é possível considerar a inclusão dessas relações não paramétricas em modelos paramétricos, permitindo uma maior flexibilidade para estimar as funções regressoras estendendo-se à classe de modelos semiparamétricos. Dentre essa classe de modelos, podemos destacar os modelos aditivos generalizados (MAGs), com importante utilidade por representar uma extensão semiparamétrica dos MLGs. Como ressalta Conceição et al. (2001), para o uso desses modelos não é necessário sequer ter algum conhecimento prévio da relação de regressão, bastando apenas estimá-la através dos dados.

Além disso, no contexto de modelos de regressão um tópico crucial é a avaliação do ajuste do modelo, tanto na verificação das suas suposições como em relação à presença de observações influentes. Essa etapa é conhecida na literatura como análise de diagnóstico e teve início nos modelos lineares clássicos. Desde então, seu conceito vem sendo estendido a diversas classes de modelos na forma de análise de resíduos, alavancagem, influência global e influência local.

Em Terra (2013) são abordados os modelos não lineares generalizados superdispersados. Diante disso, a ideia principal dessa dissertação é considerar essa classe de modelos e introduzir componentes não paramétricos, obtendo assim um modelo semiparamétrico que seja extremamente versátil do ponto de vista prático e abordar técnicas de diagnóstico específicas para essa classe de modelos.

1.2 Objetivos

Essa dissertação tem como foco principal três objetivos. O primeiro objetivo é fazer uma revisão literária acerca de modelos de regressão com superdispersão e do uso de funções de suavização na construção de modelos semiparamétricos. O segundo é propor uma classe de modelos bem geral, denominada de modelos de regressão não lineares parciais generalizados superdispersados (MNLPGS). Por fim, abordar métodos de diagnóstico nessa nova classe de modelos e, consequentemente, utilizá-los na avaliação da qualidade e sensibilidade dos ajustes. Visto isso, temos especificamente os seguintes objetivos em destaque:

1. Definir a função de verossimilhança penalizada para os modelos não lineares parciais generalizados superdispersados, bem como sua função escore e matriz de informação de Fisher para seus parâmetros.

2. Propor um método de estimação, baseado em um processo escore e um algoritmo *backfitting*, para os coeficientes de regressão e as funções não paramétricas.

3. Obter resultados importantes para a detecção de pontos influentes e de alavanca, além da definição de resíduos especificos para a média e dispersão.

4. Julgar a flexibilidade do modelo e a eficácia das técnicas de diagnóstico para detectar problemas no ajuste, feito por meio de aplicações práticas.

1.3 Organização da dissertação

Esta dissertação está organizada da seguinte forma. No Capítulo 2 é abordada uma explicação prévia de técnicas para a construção de modelos semiparamétricos, bem como algumas definições necessárias para tal. Sendo assim, o foco do capítulo é construir uma classe de modelos semiparamétricos que procuram modelar simultaneamente a média e a dispersão usando uma classe de distribuições bastante abrangente. Especificamente, é definida a função de verossimilhança e, a partir daí, são obtidas a função escore e a matriz de informação de Fisher, produzindo então um processo iterativo para a estimação dos parâmetros do modelo. Ainda nesse capítulo são discutidos os erros padrão dos estimadores e uma alternativa para seleção entre modelos concorrentes baseado nos métodos para modelos paramétricos.

No Capítulo 3, propomos técnicas de diagnóstico destinadas especificamente à classe de modelos não lineares parciais generalizados superdispersados, tais como: resíduos, alavancagem generalizada e influência local. Em resumo, as técnicas abordadas nesse capítulo são utilizadas para avaliar a qualidade do ajuste e a sensibilidade do modelo à observações que exercem pesos desproporcionais na estimação dos parâmetros. Isso é definido através de métodos bem gerais amplamente usados na literatura, sendo obtidos particularmente para os modelos em estudo.

No Capítulo 4, motivamos o uso da classe de modelos aqui proposta, resolvendo um problema de contexto prático, considerado a partir de dados reais. É ressaltado como a flexibilidade dos MNLPGS é bastante útil para resolver os mais variados problemas reais, visto que temos uma generalização de diversos modelos, com grande aplicabilidade.

Por fim, no Capítulo 5 revisamos os resultados obtidos e os objetivos atingidos com a execução deste trabalho. Em especial, são apresentadas algumas conclusões, tanto do ponto de vista teórico como prático, e perspectivas para possíveis trabalhos futuros.

CAPÍTULO 2

Modelos Não Lineares Parciais Generalizados Superdispersados

Neste capítulo será introduzida uma família de modelos semiparamétricos, obtida ao inserir uma função de suavização aos preditores na classe de modelos não lineares generalizados superdispersados. Inicialmente será discutida uma motivação ao uso dessa técnica, bem como alguns trabalhos relacionados aos modelos semiparamétricos e suas aplicações. Com isso, especificaremos o modelo e a utilização de uma função de penalização, na qual é usada para obter uma forma penalizada da função de verossimilhança. Em seguida, são calculadas a função escore e a matriz de informação de Fisher, obtidas a partir da derivação da função de verossimilhança penalizada, e então discute-se a estimação dos parâmetros desse modelo. Por fim, é abordada uma analogia aos critérios para seleção de modelos usuais, levando em consideração a função de penalização.

2.1 Introdução

Em estudos evolvendo análise de regressão, um dos interesses é avaliar o comportamento de uma ou mais variáveis respostas em função de outras variáveis observadas. Os modelos mais clássicos assumem que essa relação pode ser medida por uma estrutura funcional bem simples, representando a variável resposta como uma função linear de suas correspondentes variáveis explicativas. No entanto, há situações em que não é razoável estabelecer uma relação linear entre as variáveis, e consequentemente, uma opção é considerar classes de modelos que são capazes de contemplar estruturas mais complexas, como é o caso dos modelos não lineares. Esses modelos, embora sejam bem mais flexíveis na forma estrutural da relação de regressão, parte do pressuposto que essa apresenta uma forma funcional conhecida.

Ocasionalmente, identificar e especificar com precisão a relação de regressão não é algo tão simples, o que pode resultar em um mal ajuste. Por exemplo, estudos que levam em

consideração variáveis temporais costumam apresentar comportamentos bem peculiares, com relações representadas por curvas, obtidas a partir de técnicas não paramétricas conhecidas por métodos de suavização. Nesse contexto, os modelos não paramétricos vêm se tornando uma importante alternativa aos modelos paramétricos, por serem menos restritivos e apresentarem maior flexibilidade na sua estrutura funcional. Essa classe de modelos, além de exigir menos suposições, não obriga conhecimento prévio da relação entre as variáveis estudadas, podendo remeter à ajuste de curvas bem mais complexas e de forma precisa. De modo geral, esses métodos sugerem que, em vez de modelar a variável resposta utilizando uma função paramétrica de determinada variável explicativa, considerar uma função não paramétrica. Para mais detalhes da formalização e estudo dos modelos não paramétricos, veja Wu e Zhang (2006) e Simonoff (1996).

Uma outra classe de modelos que vem apresentando crescentes estudos e grande aplicabilidade, denominada de modelos semiparamétricos, é caracterizada por unir a modelagem paramétrica e não paramétrica. A motivação está em flexibilizar ainda mais a relação entre as variáveis explicativas e a variável resposta, ao adicionar funções de suavização às estruturas paramétricas. Particularmente, uma boa aplicabilidade é em casos nos quais deseja-se estabelecer uma relação paramétrica e ao mesmo tempo controlar o efeito da variável resposta através de determinadas características usando funções não paramétricas. Na literatura, diversos autores vem estendendo esse conceito para os mais diversificados modelos de regressão, como por exemplo, Hastie e Tibshirani (1990) que propuseram os modelos aditivos generalizados (MAGs), uma classe de modelos que, ao incluir funções não paramétricas aos preditores, representam uma importante extensão semiparamétrica dos modelos lineares generalizados.

Nas seguintes seções deste capítulo, será introduzido o conceito de modelos semiparamétricos, utilizando-o para obter uma extensão dos modelos não lineares generalizados superdispersados propostos por Dey et al. (1997). Esses modelos com superdispersão são capazes de modelar, de forma simultânea, a média e dispersão da variável de interesse, ao introduzir um preditor adicional para a dispersão de forma equivalente à média. Os preditores, por sua vez, podem assumir estruturas não lineares e, como caso particular, funções lineares mais simples. Com isso, nosso objetivo será incluir uma única função de suavização em cada um dos preditores, e então, discutir de que forma isso irá afetar o processo de estimação e a validação do modelo. Em outras palavras, o modelo será especificado por uma estrutura paramétrica de determinadas covariáveis e de uma função não paramétrica que depende apenas de uma variável explicativa em cada um dos preditores, tornando bem mais flexível a relação da variável resposta com suas covariáveis.

2.2 Conceitos preliminares

Considere inicialmente y_i $(i = 1, \ldots, n)$ como sendo a i-ésima observação de $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, o vetor de respostas. Dessa forma, a ideia base dos modelos semiparamétri-

cos é considerar uma regressão com estrutura geral definida por:

$$y_i = h(\boldsymbol{x}_i; \boldsymbol{\beta}) + f(t_i) + e_i, \tag{2.1}$$

em que $h(\boldsymbol{X}; \boldsymbol{\beta})$ é uma função paramétrica, $\boldsymbol{x}_i$ um vetor $(p \times 1)$ de variáveis explicativas, $\boldsymbol{\beta}$ um vetor de parâmetros $(p \times 1)$, $f(t_i)$ uma função de suavização duas vezes diferenciável que depende da variável $\boldsymbol{t}$ e e_i representando os erros aleatórios.

Se, por exemplo, adotarmos uma estrutura linear para a componente paramétrica e uma determinada função de suavização, então o modelo pode ser expresso pela seguinte forma matricial:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}\boldsymbol{f} + \boldsymbol{e}, \tag{2.2}$$

em que $\boldsymbol{X}$ é a matriz modelo $(n \times p)$ referente a parcela paramétrica, $\boldsymbol{N}$ é uma matriz de incidências $(n \times r)$ com elementos dados por $\mathbb{1}(t_i = t_k^0)$ $(i = 1, \ldots, n \ ; \ k = 1, \ldots, r)$ com t_k^0 sendo os valores ordenados e sem repetições da variável $\boldsymbol{t}$, $\boldsymbol{f} = (f(t_1^0), \ldots, f(t_r^0))^\top$ é o vetor $(r \times 1)$ referente a função de suavização e $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$ é um vetor de erros aleatório $(n \times 1)$.

Nesse caso, se considerarmos que os erros são normais, então o modelo apresentado em (2.2) é uma classe específica de modelos semiparamétricos, conhecida na literatura como modelos lineares parciais (MLP), os quais são discutidos por Heckman (1986) e Green e Silverman (1993), a título de exemplo. Claramente, dependendo de como é definido o componente sistemático e o componente aleatório, a estrutura geral dada em (2.1) pode abranger distintas classes de modelos, as quais tem sido estudadas por diversos autores. Alguns exemplos de casos particulares de (2.1) podem ser encontrados em Pulgar (2009).

Outro fator fundamental a ser considerado é a escolha da função de suavização, a qual vai caracterizar diretamente a forma da relação não paramétrica entre a variável resposta $\boldsymbol{y}$ e a variável explicativa $\boldsymbol{t}$. Assumindo apenas que essa função está em um espaço sob determinadas restrições, o objetivo é obter uma combinação linear de funções desse espaço que melhor aproxime a função de interesse, ou seja, assumimos que a função f pode ser escrita da seguinte forma:

$$f(\boldsymbol{t}) \ = \ \sum_{i=1}^{q} \alpha_i b_i(\boldsymbol{t}),$$

em que $b_i(\cdot)$ é uma função base e α_i são coeficientes de suavização.

No entanto, dentre essas inúmeras possibilidades da escolha de $f(\cdot)$ nos restringiremos a utilização apenas de splines cúbicos naturais, por apresentarem uma grande flexibilidade e alta eficiência computacional. De modo geral, splines são ferramentas amplamente utilizadas para a aproximação de funções, apresentando um alto poder adaptativo para a estimação de curvas. Por definição, um spline é uma curva contínua definida inicialmente por dois ou mais pontos de controle. Nesse caso, o domínio da função é dividida em

intervalos menores através de pontos de cortes, denominados de nós. Logo, a ideia geral do spline, é atribuir funções polinomiais dentro de cada um desses intervalos, formando assim curvas que passe por todos esses pontos de controle ou se aproxime o máximo possível de todos eles. Desse modo, uma função spline de ordem m com k intervalos é qualquer função que possua a seguinte forma:

$$f(t) \quad = \quad \sum_{s=0}^{m} \alpha_s t^s + \sum_{l=1}^{k} \alpha_{m+l}(t-\tau_l)_+^m.$$

em que $w_+ = \max(0, w)$ e $\tau_0, \tau_1, \tau_2, ..., \tau_k, \tau_{k+1}$ representam os nós do spline.

Os exemplos de tipos de splines mais comuns são as funções splines lineares, polinomiais e naturais. Dentre todas as possíbilidades de funções, os splines cúbicos naturais consistem em estabelecer polinômios de terceiro grau em cada intervalo definido pelos nós. Portanto, de acordo com Reinsch (1967) e Silverman (1985), estas funções são caracterizadas pelas seguintes propriedades:

- $f(\cdot)$ é um polinômio dividido em partes de terceira ordem em qualquer subintervalo $[\tau_i, \tau_{i+1})$
- Para cada ponto τ_i, a curva e suas duas primeiras derivadas são contínuas, embora sua terceira derivada possa ser descontínua.
- Nos intervalos $(-\infty, t_1]$ e $[t_{k+1}, \infty, t_1)$ a segunda derivada é zero, de mode que $f(\cdot)$ é uma função linear fora do domínio dos dados.

Abordagens mais detalhada sobre a especificação de $f(\cdot)$ e de sua construção para modelos regressão não paramétricos podem ser encontradas em Simonoff (1996), Wu e Zhang (2006) e Noda (2013), por exemplo. É importante ressaltar que, a estrutura descrita na equação (2.2) nos possibilita estimar a parcela não paramétrica através de um vetor de parâmetros desconhecido, de forma equivalente ao processo de estimação para a parcela paramétrica, facilitando assim todo o processo inferencial do modelo.

É valido notar que, o modelo definido em (2.2) possibilita o uso de procedimentos numéricos de interpolação para ajustar uma função de suavização, que por sua vez, necessita de algo para controlar a suavidade da curva estimada. Naturalmente, se a curva é muito suave, a estimação da função tende a ser menos precisa e acaba por não explicar corretamente o comportamento real dos dados. Por outro lado, se a estimação da curva tende a interpolar muitas observações obtendo pouca suavidade, isso ocasionará num ótimo ajuste, mas com alta variabilidade. Em outras palavras, processos de estimação que consideram a maximização direta da função de verossimilhança podem ocasionar problemas de super ajustes na parcela não paramétrica e, consequentemente, causar problemas de identificabilidade ao modelo.

Formas para solucionar esse problema de identificabilidade são propostas por inúmeros autores, e dentre esses métodos são destacados critérios que incorporam algum tipo de

penalidade na função de verossimilhança. Handscomb (1966) introduziu um critério de penalização no qual consiste em avaliar a integral em todo domínio de $\boldsymbol{t}$ de alguma derivada de $f(\boldsymbol{t})$, isto é, usando a seguinte função de penalização:

$$\mathrm{J} \;=\; \lambda \int_a^b [f^{(l)}(t)]^2 dt,$$

em que λ é o parâmetro de suavização e $f^{(l)}(\boldsymbol{t})$ a l-ésima derivada de $f(\boldsymbol{t})$.

Para o caso dos splines cúbicos naturais, a penalização que iremos adotar é através da segunda derivada de $f(\boldsymbol{t})$, a qual Green e Silverman (1993) demostram que podem assumir a seguinte forma quadrática:

$$\mathrm{J} \;=\; \lambda \int_a^b [f^{(2)}(t)]^2 dt = \lambda \boldsymbol{f}^\top \boldsymbol{K} \boldsymbol{f}, \tag{2.3}$$

em que $\boldsymbol{K} = \boldsymbol{Q}^\top \boldsymbol{R}^{-1} \boldsymbol{Q}$ é uma matriz $(n \times n)$ positiva definida.

As matrizes $\boldsymbol{Q}$ e $\boldsymbol{R}$ são obtidas através das distâncias h_i entre as observações de $\boldsymbol{t}$, ou seja, $h_i = t_{i+1} - t_i$. Logo, os elementos da matriz $\boldsymbol{Q}$ de dimensão $n \times (n-2)$ são dados por:

$$q_{(j-1),j} = h_{j-1}^{-1}, \qquad q_{j,j} = -h_{j-1}^{-1} - h_j^{-1} \quad \text{e} \quad q_{(j+1),j} = h_j^{-1},$$

com $q_{i,j} = 0$ para $|i-j| > 1$, $j = 2, \ldots, n-1$.

De forma semelhante, os elementos da matriz $\boldsymbol{R}$ são descritos pelas seguintes expressões:

$$\begin{aligned} r_{i,i} &= \frac{1}{3}(h_{i-1} - h_i), \quad i = 1, \ldots, n-1, \\ r_{i,(i+1)} &= q_{(i+1),i} = \frac{1}{6} h_i, \quad i = 1, \ldots, n-2, \end{aligned}$$

com $r_{ij} = 0$ para $|i-j| > 1$.

Em vista disso, esse contexto geral será utilizado para introduzir componentes sistemáticos semiparamétricos na construção dos modelos superdispersados, com o objetivo de tornar ainda mais flexível a relação entre a variável resposta e suas respectivas variáveis explicativas.

2.3 Especificação do modelo

Os modelos lineares generalizados superdispersados propostos por Dey et al. (1997) representam uma extensão aos modelos lineares generalizados, os quais foram introduzido por Nelder e Wedderburn (1972). Esses modelos, por sua vez, são caracterizados por três quantidades fundamentais: o componente aleatório, o componente sistemático e a função

de ligação. Do ponto de vista prático, ao fazer uso dessa teoria é necessário definir bem o que cada um desses termos irá representar no estudo em questão.

Visto isso, a fim de especificar essa classe de modelos, precisamos inicialmente definir o seu componente aleatório e seus dois componentes sistemáticos, um relacionado à média e outro a dispersão. Naturalmente, assumiremos que o componente aleatório do modelo tem distribuição pertencente a família de distribuições definida por Dey et al. (1997), generalizando a exponencial dupla proposta por Efron (1986), com função de densidade dada por:

$$\pi(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\phi}) \;=\; A(\boldsymbol{y})\exp\left\{(\boldsymbol{y}-\boldsymbol{\mu})\Psi^{(1,0)}(\boldsymbol{\mu},\boldsymbol{\phi})+\phi T(\boldsymbol{y})+\Psi(\boldsymbol{\mu},\boldsymbol{\phi})\right\}, \tag{2.4}$$

em que $A(\boldsymbol{y})$, $T(\cdot)$ e $\Psi(\cdot,\cdot)$ são funções conhecidas e $\Psi^{(r,s)}(\boldsymbol{\mu},\boldsymbol{\phi}) = \partial^{r+s}\Psi(\boldsymbol{\mu},\boldsymbol{\phi})/\partial\mu^r\partial\phi^s$, r,$s \geq 0$.

Considerando-se que $\boldsymbol{y}$ tem função de densidade dada por (2.4), então seguem as seguintes relações: $\mathbb{E}(y) = \mu$, $\mathrm{Var}(y) = \Psi^{(2,0)}(\boldsymbol{\mu},\boldsymbol{\phi})$, $\mathbb{E}[T(y)] = -\Psi^{(0,1)}(\boldsymbol{\mu},\boldsymbol{\phi})$ e $\mathrm{Var}[T(y)] = -\Psi^{(0,2)}(\boldsymbol{\mu},\boldsymbol{\phi})$. Já os componentes sistemáticos serão especificados por meio de funções não lineares das variáveis explicativas e de uma função não paramétrica associada a uma variável específica. Seja $\boldsymbol{X}$ a matriz modelo associada a média e $\boldsymbol{S}$ a matriz modelo associada a dispersão, então os componentes sistemáticos assumem a seguinte estrutura:

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \varphi_1(\boldsymbol{X};\boldsymbol{\beta}) + \boldsymbol{N}_\mu \boldsymbol{f}_{\boldsymbol{\mu}}, \tag{2.5}$$

e

$$h(\boldsymbol{\phi}) = \boldsymbol{\tau} = \varphi_2(\boldsymbol{S};\boldsymbol{\gamma}) + \boldsymbol{N}_\phi \boldsymbol{f}_{\boldsymbol{\phi}}, \tag{2.6}$$

com $g(\cdot)$ e $h(\cdot)$ representando as funções de ligação da média e da dispersão, respectivamente.

Nesse caso, as funções $g(\cdot)$ e $h(\cdot)$ deverão ser conhecidas, bijetoras e as funções $f_1(\boldsymbol{X};\boldsymbol{\beta})$ e $f_2(\boldsymbol{S};\boldsymbol{\gamma})$ serão funções não lineares contínuas e diferenciáveis com respeito ao vetor de parâmetros, tal que as matrizes de derivadas $\tilde{\boldsymbol{X}} = \partial\varphi_1(\boldsymbol{X};\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ e $\tilde{\boldsymbol{S}} = \partial\varphi_1(\boldsymbol{S};\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$ tenham posto completo. Com isso, o modelo discutido nesse capítulo é completamente especificado por meio do componente aleatório definido em (2.4) e dos componentes sistemáticos (2.5) e (2.6), para a média e dispersão, respectivamente.

2.4 Distribuições pertencentes à classe de modelos superdispersados

Algumas distribuições pertencentes à classe dos MNLPGS são apresentadas a seguir. Para mais detalhes da obtenção dessas distribuições como caso particular dos MNLPGS, veja Terra (2013) e Previdelli (2005), por exemplo.

2.4.1 Inversa gaussiana generalizada

Essa distribuição foi proposta por Good (1953) e representa um alternativa tri paramétrica para competir com a distribuição gama, por exemplo. Dessa forma, de acordo Johnson et al. (1994) a função de densidade da inversa gaussiana generalizada é dada por:

$$\pi(y;\alpha,\delta,r) \;=\; \frac{\frac{\alpha}{\delta}^{\frac{r}{2}}}{2K_r(\sqrt{\alpha\delta})}y^{r-1}\exp\left\{-\frac{1}{2}(\delta y^{-1}+\alpha y)\right\}, y,\alpha,\delta>0,$$

em que $K_r(\cdot)$ é uma função de Bessel modificada de terceira ordem.

Dessa forma, utilizando a classe dos MNLPGS na sua forma de origem apresentada por Gelfand e Dalal (1990), ou seja,

$$\pi(y;\mu,\phi) \;=\; A(y)\exp\{\theta y + rT(y) - \rho(\theta,\phi)\}.$$

Logo, a função de densidade da inversa gaussiana generalizada pode ser reescrita na seguinte forma:

$$\pi(y;\alpha,\delta,r) \;=\; y^{r-1}\exp\left\{-\frac{1}{2}\delta y^{-1} - \frac{1}{2}\alpha y + \log\left[\frac{(\frac{\alpha}{\delta})^{\frac{r}{2}}}{2K_r(\sqrt{\alpha\delta})}\right]\right\}.$$

Com isso, considerando r constante podemos definir as quantidades $A(y) = y^{r-1}$, $T(y) = -y^{-1}$, $\theta = \Psi^{(1,0)}(\mu,\phi) = -\alpha/2$ e $\phi = -\delta/2$. Adicionalmente, tome $\alpha\delta = 4\theta\phi$, então temos que:

$$\rho(\theta,\phi) \;=\; -\frac{r}{2}\log\left(\frac{\theta}{\phi}\right) + \log 2 + \log K_r(2\sqrt{\theta\phi}).$$

No entanto, a especificação de $\Psi(\mu,\phi)$ é considerada apenas por meio da manipulação da seguinte equação:

$$\Psi(\mu,\phi) \;=\; \int \theta = \int \frac{\partial\Psi(\mu,\phi)}{\partial\mu},$$

a qual só pode ser obtida de forma numérica, ou seja, não possui forma fechada.

Portanto, a distribuição inversa gaussiana generalizada é um caso particular da família com densidade definida em (2.4), e consequentemente generaliza todos os casos especiais ao variar os parâmetros α, δ e r dessa subfamília, como por exemplo a distribuição inversa gaussiana (se $r = -1/2$), gama (se $\delta = 0$ e $r > 0$) e recíproca gama (se $r = 0$).

2.4.2 Família exponencial dupla

De acordo com Efron (1986) e seguindo a notação de Gelfand e Dalal (1990), a família exponencial dupla pode ser definida através dos parâmetros θ,ρ e a com seguinte função

de densidade:

$$\begin{aligned}\pi(y;\theta,\rho,a) &= c(\theta,\rho,a)\rho^{1/2}\{\exp\{a\alpha(y\theta-\chi^2(\theta))\}h(y;a)\}^{\rho} \\ &\quad \{\exp\{a\alpha(y\theta-\chi^2(\theta))\}h(y;a)\}^{1-\rho}[dh(y;a)], \\ &= c(\theta,\rho,a)\rho^{1/2}\exp\{a\alpha(y\theta-\chi^2(\theta))+a(1-\rho)(y\theta(y)-\chi^2(\theta(y)))\}h(y;a),\end{aligned}$$

em que $\Psi(\mu)=\chi^2(\theta)$, $h(y;a)$ uma função conhecida, $c(\theta,\rho,a)$ uma constante aproximadamente igual a zero, $\theta=\mathbb{E}(y)=\mu$ e $\text{Var}(y)=V(\theta)/a\theta$.

Com isso, através de manipulações algébricas, como expõe Terra (2013), podemos reescrever essa densidade na seguinte forma:

$$\begin{aligned}\pi(y;\theta,\rho,a) &= c(\theta,\rho,a)\rho^{1/2}\exp\{\alpha\rho y\theta-\alpha\rho\chi^2(\theta)+a(1-\rho)T(y)h(y;a)\}, \\ &= c(\theta,\rho,a)\rho^{1/2}\exp\{\alpha y+T(y)-\rho(\alpha,\phi)h(y;a)\},\end{aligned}$$

em que $T(y)=y\theta(y)-\chi^2(\theta(y))$ e $\phi=a(1-\rho)$.

Sendo assim, a família de distribuições definida por Dey et al. (1997) englobam a família exponencial dupla exposta por Efron (1986), por meio da representação da família exponencial de dois parâmetros definida por Gelfand e Dalal (1990). Consequentemente, a classe de modelos superdispersados englobam as famílias distribuições binomial, gama e poisson dupla, por exemplo. Em seguida é apresentado o caso particular da densidade da poisson dupla, uma vez que está será utilizado para a aplicação prática em dados reais discutida no capítulo 4.

2.4.3 Família poisson dupla

Essa família é baseada na distribuição poisson, a qual é principalmente utilizada em dados de contagem onde a variável resposta é inteiro não negativo. Segundo Efron (1986), assume-se que a variável resposta Y $(y=0,1,2,\ldots)$ tem função densidade de probabilidade dada por:

$$\pi(y;\mu,\phi) = \left(\phi^{\frac{1}{2}}e^{-\phi\mu}\right)\left(\frac{e^{-y}y^y}{y!}\right)\left(\frac{e\mu}{y}\right)^{\phi y},$$

em que $\mathbb{E}(y)=\mu$, $\text{Var}(y)=\mu/\phi$.

Desse modo, através de manipulações algébricas é possível reescrever esta função de densidade sob a seguinte forma:

$$\pi(y;\mu,\phi) = \left(\frac{e^{-y}y^y}{y!}\right)\exp\left\{(y-\mu)[\phi+\phi\log\mu]-\phi y\log y+\phi y\log\mu+\frac{1}{2}\log\phi\right\}.$$

Logo, podemos obter a densidade da poisson dupla sob a notação definida em (2.4) com $A(y)=y^{r-1}$, $T(y)=-y^{-1}$, $\theta=\Psi^{(1,0)}(\mu,\phi)=-\alpha/2$ e $\phi=-\delta/2$ $A(y)=\left(\frac{e^{-y}y^y}{y!}\right)$, $T(y)=-y\log y$, $\Psi(\mu,\phi)=\phi\mu\log\mu+1/2\log\phi$ e $\Psi^{(1,0)}(\mu,\phi)=\phi+\phi\log\mu$. Portanto,

a distribuição poisson dupla também representa um caso particular da classe de modelos superdispersados, sendo assim um possível competidor aos modelos com distribuição binomial negativa.

2.5 Função de verossimilhança penalizada

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{f}_\mu^\top, \boldsymbol{f}_\phi^\top)^\top$ o vetor de parâmetros do modelo e $\boldsymbol{y}$ o vetor com as observações da variável resposta, então o logaritmo da função verossimilhança tem a seguinte forma:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left\{ (y_i - \mu_i)\Psi^{(1,0)}(\mu_i, \phi_i) + \phi_i T(y_i) + \Psi(\mu_i, \phi_i) \right\} + \sum_{i}^{n} \log A(y_i).$$

Como foi mencionado anteriormente, existe a necessidade de inserir penalizações na verossimilhança para cada uma das funções de suavização, ou seja, uma para a média e outra para dispersão. Dessa forma, considerando um spline cúbico natural e sua respectiva penalização dada em (2.3), o logaritmo da função de verossimilhança penalizada é dada por:

$$\begin{aligned} l_p(\boldsymbol{\theta}) = & \sum_{i=1}^{n} \left\{ (y_i - \mu_i)\Psi^{(1,0)}(\mu_i, \phi_i) + \phi_i T(y_i) + \Psi(\mu_i, \phi_i) \right\} + \sum_{i}^{n} \log A(y_i) \\ & - \frac{1}{2}\lambda_\mu^* \boldsymbol{f}_\mu^\top \boldsymbol{K}_\mu \boldsymbol{f}_\mu - \frac{1}{2}\lambda_\phi^* \boldsymbol{f}_\phi^\top \boldsymbol{K}_\phi \boldsymbol{f}_\phi. \end{aligned} \tag{2.7}$$

Desse modo, utilizamos a função definida em (2.7) para obter os estimadores de máxima verossimilhança (EMV) penalizados, os quais serão discutidos com mais detalhes na Seção 2.7.

2.6 Função escore e matriz de informação de Fisher penalizadas

Para a obtenção dessas funções, assumimos que λ_μ e λ_ϕ são valores fixos. Logo, a função escore penalizada é definida pela seguinte estrutura matricial:

$$U_p(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} U_p(\boldsymbol{\beta}) \\ U_p(\boldsymbol{f_\mu}) \\ U_p(\boldsymbol{\gamma}) \\ U_p(\boldsymbol{f_\phi}) \end{pmatrix},$$

em que

$$\begin{aligned} U_p(\boldsymbol{\beta}) &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}), \\ U_p(\boldsymbol{f}_\mu) &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}) - \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}} \boldsymbol{f}_{\boldsymbol{\mu}}, \\ U_p(\boldsymbol{\gamma}) &= \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 (\boldsymbol{y} - \boldsymbol{\mu}) + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Phi}_1 T(\boldsymbol{y}) + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1 \end{aligned}$$

e

$$U_p(\boldsymbol{f}_{\boldsymbol{\phi}}) = \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 (\boldsymbol{y} - \boldsymbol{\mu}) + \boldsymbol{N}_{\phi}^\top \boldsymbol{\Phi}_1 T(\boldsymbol{y}) + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1,$$

sendo $\boldsymbol{M}_1 = diag(a_1, \ldots, a_n)$, com $a_i = \partial \mu_i / \partial \eta$ e $\boldsymbol{\Phi_1} = diag(b_1, \ldots, b_n)$, com $b_i = \partial \phi_i / \partial \tau_i$, $i = 1, \ldots, n$.

A matriz de informação de Fisher penalizada é definida através da estrutura matricial dada:

$$\mathrm{I}_p(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{I}_p^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{\beta}\boldsymbol{f}_\mu}(\boldsymbol{\theta}) & \mathbf{0}_{p_1 \times p_2} & \mathbf{0}_{p_1 \times r_2} \\ \mathrm{I}_p^{\boldsymbol{f}_\mu\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{f}_\mu\boldsymbol{f}_\mu}(\boldsymbol{\theta}) & \mathbf{0}_{r_1 \times p_2} & \mathbf{0}_{r_1 \times r_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathbf{0}_{p_2 \times r_1} & \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{f}_\phi}(\boldsymbol{\theta}) \\ \mathbf{0}_{r_2 \times p_1} & \mathbf{0}_{r_2 \times r_1} & \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{f}_\phi}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{f}_\phi\boldsymbol{f}_\phi}(\boldsymbol{\theta}) \end{pmatrix}.$$

em que

$$\begin{aligned} \mathrm{I}_p^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}}, \\ \mathrm{I}_p^{\boldsymbol{\beta}\boldsymbol{f}_\mu}(\boldsymbol{\theta}) &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}}, \\ \mathrm{I}_p^{\boldsymbol{f}_\mu\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}}, \\ \mathrm{I}_p^{\boldsymbol{f}_\mu\boldsymbol{f}_\mu}(\boldsymbol{\theta}) &= \tilde{\boldsymbol{N}_{\boldsymbol{\mu}}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}}, \\ \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= -\tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \tilde{\boldsymbol{S}}, \\ \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{f}_\phi}(\boldsymbol{\theta}) &= -\tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{N}_{\boldsymbol{\phi}}, \\ \mathrm{I}_p^{\boldsymbol{f}_\phi\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= -\boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \tilde{\boldsymbol{S}} \end{aligned}$$

e

$$\mathrm{I}_p^{\boldsymbol{f}_\phi\boldsymbol{f}_\phi}(\boldsymbol{\theta}) = -\tilde{\boldsymbol{N}_{\boldsymbol{\phi}}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\phi}} + \lambda_{\boldsymbol{\phi}} \boldsymbol{K}_{\boldsymbol{\phi}}.$$

com $\boldsymbol{M}_1^2 = diag(a_1^2, \ldots, a_n^2)$ e $\boldsymbol{\Phi}_1^2 = diag(b_1^2, \ldots, b_n^2)$.

Note que, os parâmetros relacionados a média ($\boldsymbol{\beta}$ e $\boldsymbol{f}_\mu$) são ortogonais aos parâmetros associados a dispersão ($\boldsymbol{\gamma}$ e $\boldsymbol{f}_\phi$), indicando que a estimação desses parâmetros pode ser feita de forma independente.

2.7 Método de Estimação

O EMV penalizado para o vetor de parâmetros $\boldsymbol{\theta}$, dado $\boldsymbol{\lambda} = (\lambda_{\boldsymbol{\mu}}, \lambda_{\boldsymbol{\phi}})$ é o valor que maximiza (2.7) sobre todo o espaço paramétrico associado a $\boldsymbol{\theta}$, ou seja,

$$l_p(\hat{\boldsymbol{\theta}}; \lambda_{\boldsymbol{\mu}}, \lambda_{\boldsymbol{\phi}}) \;\; \geq \;\; \sup_{\theta \in \Theta} l_p(\boldsymbol{\theta}; \lambda_{\boldsymbol{\mu}}, \lambda_{\boldsymbol{\phi}}).$$

O processo de estimação desse modelo será divido em etapas. Inicialmente, será considerado que os parâmetros de suavização são valores fixos e assim obtemos um procedimento para a estimação dos parâmetros associados a média e a dispersão através da combinação dos algoritmos Scoring de Fisher e *backfitting*. Em seguida, será considerado um processo de estimação separadamente para os parâmetros de suavização usando métodos de validação cruzada, como veremos a seguir.

2.7.1 Estimação dos parâmetros da média

Considerando que $\lambda_{\boldsymbol{\mu}}$ assume um valor fixo, então o processo de estimação para os parâmetros da média são obtidos igualando-se as respectivas funções escore a zero, como temos abaixo:

$$U_p(\hat{\boldsymbol{\beta}}) \;\; = \;\; \tilde{\boldsymbol{X}}^{\top} \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}) = 0$$

e

$$U_p(\hat{\boldsymbol{f}}_{\boldsymbol{\mu}}) \;\; = \;\; \boldsymbol{N}_{\boldsymbol{\mu}}^{\top} \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}) - \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}} \hat{\boldsymbol{f}}_{\boldsymbol{\mu}} = 0.$$

No entanto, não é possível resolver esse sistema de equações de forma direta, sendo necessário o uso de métodos numéricos para a maximização de (2.7) em função de $\boldsymbol{\beta}$ e $\boldsymbol{f}_{\mu}$. Nesse caso, utilizando o algoritmo escore de Fisher, podemos definir o seguinte processo iterativo:

$$\begin{pmatrix} \mathrm{I}_p^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{\beta}\boldsymbol{f}_{\boldsymbol{\mu}}}(\boldsymbol{\theta}) \\ \mathrm{I}_p^{\boldsymbol{f}_{\boldsymbol{\mu}}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{f}_{\boldsymbol{\mu}}\boldsymbol{f}_{\boldsymbol{\mu}}}(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)} \\ \boldsymbol{f}_{\mu}^{(m+1)} - \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} \end{pmatrix} = \begin{pmatrix} U_p(\boldsymbol{\beta}) \\ U_p(\boldsymbol{f}_{\boldsymbol{\mu}}) \end{pmatrix}.$$

Logo, após algumas manipulações algébricas, as quais podem ser encontradas no Apêndice A, obtemos o seguinte processo iterativo para os parâmetros da média:

$$\begin{pmatrix} \boldsymbol{\beta}^{(m+1)} \\ \boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_0^{(m)} [\boldsymbol{r}_{\boldsymbol{\beta}}^{(m,m+1)} + \boldsymbol{\eta}^{(m)}] \\ \boldsymbol{D}_1^{(m)} [\boldsymbol{r}_{\boldsymbol{f}_{\boldsymbol{\mu}}}^{(m,m+1)} + \boldsymbol{\eta}^{(m)}] \end{pmatrix}, \tag{2.8}$$

em que

$$\boldsymbol{D}_j^{(m)} = \begin{cases} (\tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2, & \text{se } j = 0 \\ (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1} \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2, & \text{se } j = 1 \end{cases}$$

e

$$\boldsymbol{r}_{\boldsymbol{\delta}}^{(m,m+1)} = \begin{cases} \boldsymbol{M}_1^{-1}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)}) - \boldsymbol{N}_\mu \boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)}, & \text{se } \boldsymbol{\delta} = \boldsymbol{\beta} \\ \boldsymbol{M}_1^{-1}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)}) - \tilde{\boldsymbol{X}} \boldsymbol{\beta}^{(m+1)}, & \text{se } \boldsymbol{\delta} = \boldsymbol{f}_\mu \end{cases}$$

Por outro lado, é possível reescrever as equações definidas em (2.8) sobre uma forma alternativa utilizando substituições, apresentando certa utilidade em análises futuras. Logo, considerando $\boldsymbol{z} = \boldsymbol{\eta} + \boldsymbol{M}_1^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$ e substituindo a expressão referente a $\boldsymbol{f}_\mu$ em $\boldsymbol{\beta}$, obtemos a seguinte expressão:

$$\hat{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}^\top \boldsymbol{W}_{\boldsymbol{\beta}} \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{W}_{\boldsymbol{\beta}} \boldsymbol{z},$$

em que $\boldsymbol{W}_\beta = \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 - \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1} \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2$.

De forma equivalente, substituindo a expressão obtida para $\boldsymbol{\beta}$ em $\boldsymbol{f}_\mu$, temos o seguinte resultado:

$$\hat{\boldsymbol{f}}_{\boldsymbol{\mu}} = (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{W}_{\boldsymbol{\mu}} \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1} \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{W}_{\boldsymbol{\mu}} \boldsymbol{z},$$

em que $\boldsymbol{W}_{\boldsymbol{\mu}} = \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 - \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} (\tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2$.

Dessa forma, utilizando as expressões de $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{f}}_\mu$ e após algumas manipulações algébricas, o vetor $\hat{\boldsymbol{\eta}}$ de valores ajustados fica dada por:

$$\hat{\boldsymbol{\eta}} = \tilde{\boldsymbol{X}} \hat{\boldsymbol{\beta}} + \boldsymbol{N}_{\boldsymbol{\mu}} \hat{\boldsymbol{f}}_{\boldsymbol{\mu}} = \mathbf{H}_{\boldsymbol{\mu}} \hat{\boldsymbol{z}} = \begin{pmatrix} \tilde{\boldsymbol{X}} & \boldsymbol{N}_{\boldsymbol{\mu}} \end{pmatrix} \boldsymbol{C}_{\boldsymbol{\mu}}^{-1} \begin{pmatrix} \tilde{\boldsymbol{X}}^\top \\ \boldsymbol{N}_{\boldsymbol{\mu}}^\top \end{pmatrix} \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 \hat{\boldsymbol{z}}, \tag{2.9}$$

sendo a matriz $\boldsymbol{C}_{\boldsymbol{\mu}}$ dada por:

$$\boldsymbol{C}_{\boldsymbol{\mu}} = \begin{pmatrix} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} & \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} \\ \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} & (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_\mu \boldsymbol{K}_{\boldsymbol{\mu}}) \end{pmatrix}.$$

A matriz $\mathbf{H}_{\boldsymbol{\mu}}$ em (2.9), representa uma matriz *hat* associada a estimação da média, que embora não possua a propriedade de idempotência, é semelhante ao caso do modelo de regressão clássico. Em outras palavras, a matriz $\mathbf{H}_\mu$ tem intepretação e utilidade de forma equivalente a matriz *hat* definida em modelos puramente paramétricos, se destacando com certa relevância em técnicas de diagnóstico, na qual será discutida mais a frente.

2.7.2 Estimação dos parâmetros da dispersão

Considerando que $\lambda_{\boldsymbol{\phi}}$ assume um valor fixo, então o processo de estimação para os parâmetros de dispersão são obtidos igualando-se as funções escore de $\boldsymbol{\gamma}$ e $\boldsymbol{f_\phi}$ a zero, da seguinte forma:

$$U_p(\hat{\boldsymbol{\gamma}}) \;=\; \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 (\boldsymbol{y}-\boldsymbol{\mu}) + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Phi}_1 T(\boldsymbol{y}) + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1 = 0$$

e

$$U_p(\hat{\boldsymbol{f}}_{\boldsymbol{\phi}}) \;=\; \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 (\boldsymbol{y}-\boldsymbol{\mu}) + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Phi}_1 T(\boldsymbol{y}) + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1 = 0.$$

Novamente, não é possível obter uma expressão analítica para a solução desse sistema de equações, devendo-se recorrer ao uso de métodos numéricos para a maximização de (2.7) em função de $\boldsymbol{\gamma}$ e $\boldsymbol{f_\phi}$. Assim, o processo iterativo baseado no algoritmo escore de Fisher fica expresso por:

$$\begin{pmatrix} \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{f_\phi}}(\boldsymbol{\gamma}) \\ \mathrm{I}_p^{\boldsymbol{f_\phi}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{f_\phi}\boldsymbol{f_\phi}}(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\gamma}^{(m+1)} - \boldsymbol{\gamma}^{(m)} \\ \boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} - \boldsymbol{f}_{\boldsymbol{\phi}}^{(m)} \end{pmatrix} = \begin{pmatrix} U_p(\boldsymbol{\beta}) \\ U_p(\boldsymbol{f_\phi}) \end{pmatrix}.$$

Com isso, realizando algumas manipulações algébricas, novamente explicitadas no Apêndice A, chegamos ao seguinte processo iterativo para os parâmetros de dispersão:

$$\begin{pmatrix} \boldsymbol{\gamma}^{(m+1)} \\ \boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{D}_2^{(m)}[\boldsymbol{r}_{\boldsymbol{\gamma}}^{(m,m+1)} + \boldsymbol{\tau}^{(m)}] \\ \boldsymbol{D}_3^{(m)}[\boldsymbol{r}_{\boldsymbol{f_\phi}}^{(m,m+1)} + \boldsymbol{\tau}^{(m)}] \end{pmatrix}. \tag{2.10}$$

em que

$$\boldsymbol{D}_j^{(m)} \;=\; \begin{cases} (\tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \tilde{\boldsymbol{S}})^{-1} \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2, \text{ se } j = 2 \\ (\boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Psi}_1^2 \tilde{\boldsymbol{S}} + \lambda_{\boldsymbol{\phi}} \boldsymbol{K}_{\boldsymbol{\phi}})^{-1} \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2, \text{ se } j = 3 \end{cases}$$

e

$$\boldsymbol{r}_{\boldsymbol{\delta}}^{(m,m+1)} \;=\; \begin{cases} -(\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1)^{-1} \boldsymbol{\Psi}^{(1,1)} - \boldsymbol{N}_{\boldsymbol{\phi}} \boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} - (\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1)^{-1} \boldsymbol{\Psi}^{(1,1)} (\boldsymbol{y}-\boldsymbol{\mu}) - (\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1)^{-1} T(\boldsymbol{y}), \text{ se } \delta \\ -(\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1)^{-1} \boldsymbol{\Psi}^{(1,1)} - \tilde{\boldsymbol{S}} \boldsymbol{\gamma}^{(m+1)} - (\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1)^{-1} \boldsymbol{\Psi}^{(1,1)} (\boldsymbol{y}-\boldsymbol{\mu}) \\ -(\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1)^{-1} T(\boldsymbol{y}), \text{ se } \boldsymbol{\delta} = \boldsymbol{f}_\phi \end{cases}$$

Reescrevendo as equações definidas em (2.10) sobre uma forma alternativa, considerando $\boldsymbol{s} = [\boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2]^{-1} [\boldsymbol{\tau} - \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1^2 (\boldsymbol{y}-\boldsymbol{\mu}) - \boldsymbol{\Phi}_1 T(\boldsymbol{y}) - \boldsymbol{\Psi}^{(0,1)}]$ e substituindo a expressão referente a $\boldsymbol{f_\phi}$ em $\boldsymbol{\gamma}$, temos que:

$$\hat{\boldsymbol{\gamma}} \;=\; (\tilde{\boldsymbol{S}}^\top \boldsymbol{W}_\gamma \tilde{\boldsymbol{S}})^{-1} \tilde{\boldsymbol{S}}^\top \boldsymbol{W}_\gamma \boldsymbol{s},$$

em que $\boldsymbol{W}_\gamma = \boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2 - \boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}(\boldsymbol{N}_{\boldsymbol{\phi}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}} + \lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}})^{-1}\boldsymbol{N}_{\boldsymbol{\phi}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2$.

De forma equivalente, substituíndo a expressão referente a $\boldsymbol{\gamma}$ em $\boldsymbol{f}_\phi$, obtemos a seguinte expressão:

$$\hat{\boldsymbol{f}}_{\boldsymbol{\phi}} \;=\; (\boldsymbol{N}_{\boldsymbol{\phi}}^\top\boldsymbol{W}_{\boldsymbol{\phi}}\boldsymbol{N}_{\boldsymbol{\phi}} + \lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}})^{-1}\boldsymbol{N}_{\boldsymbol{\phi}}^\top\boldsymbol{W}_{\boldsymbol{\phi}}\boldsymbol{t},$$

em que $\boldsymbol{W}_{\boldsymbol{\phi}} = \boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2 - \boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}}(\tilde{\boldsymbol{S}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}})^{-1}\tilde{\boldsymbol{S}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2$.

Dessa forma, utilizando as expressões de $\hat{\boldsymbol{\gamma}}$ e $\hat{\boldsymbol{f}}_{\boldsymbol{\phi}}$ e após algumas manipulações algébricas, o vetor $\hat{\boldsymbol{\tau}}$ de valores ajustados fica dada por:

$$\hat{\boldsymbol{\tau}} \;=\; \tilde{\boldsymbol{S}}\hat{\boldsymbol{\gamma}} + \boldsymbol{N}_{\boldsymbol{\phi}}\hat{\boldsymbol{f}}_{\boldsymbol{\phi}} = \mathbf{H}_{\boldsymbol{\phi}}\hat{\boldsymbol{s}} = \begin{pmatrix} \tilde{\boldsymbol{X}} & \boldsymbol{N}_{\boldsymbol{\phi}} \end{pmatrix}\boldsymbol{C}_{\boldsymbol{\phi}}^{-1}\begin{pmatrix} \tilde{\boldsymbol{X}}^\top \\ \boldsymbol{N}_{\boldsymbol{\mu}}^\top \end{pmatrix}\boldsymbol{\Psi}^{(2,0)}\boldsymbol{M}_1\hat{\boldsymbol{s}}, \tag{2.11}$$

sendo a matriz $\boldsymbol{C}_{\boldsymbol{\phi}}$ dada por:

$$\boldsymbol{C}_{\boldsymbol{\phi}} = \begin{pmatrix} \tilde{\boldsymbol{S}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}} & \tilde{\boldsymbol{S}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}} \\ \boldsymbol{N}_{\boldsymbol{\phi}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}} & (\boldsymbol{N}_{\boldsymbol{\phi}}^\top\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}} + \lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}}) \end{pmatrix}.$$

A matriz $\boldsymbol{H}_{\boldsymbol{\phi}}$ em (2.11), representa uma matriz *hat* associada a estimação da dispersão, equivalente a matriz $\boldsymbol{H}_{\boldsymbol{\mu}}$ obtida na seção anterior, possuindo interpretação similar.

2.7.3 Chute inicial

Como a estimação dos parâmetros do modelo é obtida usando o algoritmo escore, então um ponto crucial a se discutir é a respeito dos valores que inicializarão esse processo, os quais estão fortemente ligados ao sucesso na convergência do método. Em muitos casos, se os chutes iniciais não estiverem próximos do valor real, o processo pode ter problemas com máximos locais, ou até mesmo não obter convergência alguma.

Uma possível alternativa para os chutes iniciais é considerar valores iniciais para os preditores baseados nos dados. Nesse contexto, Gijbels et al. (2010) sugerem que, para a classe de modelos exponenciais duplos não paramétricos, os chutes iniciais para as funções da média e dispersão são considerados constantes e representados por seus respectivos valores amostrais baseados no modelo exponencial correspondente. De forma equivalente, considerando que há uma relação entre os preditores com a média e dispersão, então os chutes iniciais para $\boldsymbol{\eta}$ e $\boldsymbol{\tau}$, respectivamente, são os vetores com as constantes $g^{-1}(\overline{\boldsymbol{y}})$ e $h^{-1}(\phi_n)$, com ϕ_n representando o valor conhecido para ϕ no modelo exponencial correspondente ao considerado nessa classe com superdispersão. Consequentemente, os valores iniciais dos vetores $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ referente a parcela paramétrica são definidos diretamente da forma funcional para os preditores. Por outro lado, os valores iniciais para $\boldsymbol{f}_{\boldsymbol{\mu}}$ e $\boldsymbol{f}_{\boldsymbol{\phi}}$ são obtidos a partir da funções dos valores de $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ definidos anteriormente.

2.7.4 Algoritmo de *backfitting*

As EMV via método escore de Fisher para os parâmetros da média e da dispersão são obtidas resolvendo, de forma iterativa, as expressões (2.8) e (2.10). No entanto, não é aconselhável resolver as equações obtidas em (2.8) e (2.10) de forma direta. Na classe de modelos semiparamétricos, é comum aproximar o processo de estimação através da combinação do algoritmo *backfitting* e do processo escore de Fisher. Como ressalta Pulgar (2009), "embora o algoritmo *backfitting* seja uma técnica iterativa que fornece dificuldades adicionais no desenvolvimento da teoria assintótica, o método tem sido refinado e estendido para modelos mais complexos". Para mais detalhes a respeito das condições de convergência desse método de estimação e suas propriedades assintóticas veja Hastie e Tibshirani (1987), Buja et al. (1989) e Pulgar (2009).

Com isso, considere que m representa a iteração do processo escore de Fisher e m^* a etapa do algoritmo *backfitting*, então o processo de estimação para $\boldsymbol{\theta}$ se dá por meio dos passos apresentados logo a seguir.

- **Passo 1:** Inicie o processo com $\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m,0)}$, $\boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} = \boldsymbol{f}_{\boldsymbol{\mu}}^{(0,0)}$, $\boldsymbol{\beta}^{(m)} = \boldsymbol{\gamma}^{(m,0)}$ e $\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)} = \boldsymbol{f}_{\boldsymbol{\phi}}^{(0,0)}$.

- **Passo 2:** Para m, $m^* = 0, 1, 2, \ldots$ calcular:

  a - processo de estimação da média:

$$\begin{aligned}
\boldsymbol{r}_{\boldsymbol{\beta}}^{(m,m*)} &= \boldsymbol{M}_1^{-1}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)}) - \boldsymbol{N}_{\mu}\boldsymbol{f}_{\boldsymbol{\mu}}^{(m,m*)}, \\
\boldsymbol{\beta}^{(m+1,m*+1)} &= \boldsymbol{D}_0^{(m)}\left[\boldsymbol{r}_{\boldsymbol{\beta}}^{(m,m*)} + \boldsymbol{\eta}^{(m)}\right], \\
\boldsymbol{r}_{\boldsymbol{f}_{\boldsymbol{\mu}}}^{(m,m*)} &= \boldsymbol{M}_1^{-1}(\boldsymbol{y} - \boldsymbol{\mu}^{(m)}) - \tilde{\boldsymbol{X}}\boldsymbol{\beta}^{(m+1,m*+1)}
\end{aligned}$$

  e

$$\boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1,m*+1)} = \boldsymbol{D}_0^{(m)}\left[\boldsymbol{r}_{\boldsymbol{f}_{\mu}}^{(m,m*)} + \boldsymbol{\eta}^{(m)}\right].$$

  b - processo de estimação da dispersão:

$$\begin{aligned}
\boldsymbol{r}_{\boldsymbol{\gamma}}^{(m,m*)} &= -\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}\boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} + (-\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Psi}^{(1,1)}(\boldsymbol{y} - \boldsymbol{\mu}) + (\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}T(\boldsymbol{y}) \\
&\quad +(\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Psi}^{(1,1)}], \\
\boldsymbol{\gamma}^{(m+1,m*+1)} &= \boldsymbol{D}_2^{(m)}\left[\boldsymbol{r}_{\boldsymbol{\gamma}}^{(m,m*)} + \boldsymbol{\tau}^{(m)}\right], \\
\boldsymbol{r}_{\boldsymbol{f}_{\boldsymbol{\phi}}}^{(m,m*)} &= -\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}}\boldsymbol{\gamma}^{(m+1)} + (-\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Psi}^{(1,1)}(\boldsymbol{y} - \boldsymbol{\mu}) + (\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}T(\boldsymbol{y}) \\
&\quad +(\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Psi}^{(1,1)}]
\end{aligned}$$

  e

$$\boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1,m*+1)} = \boldsymbol{D}_3^{(m)}\left[\boldsymbol{r}_{\boldsymbol{f}_{\boldsymbol{\phi}}}^{(m,m*)} + \boldsymbol{\tau}^{(m)}\right].$$

- **Passo 3:** Repita os passos (1) e (2) até obter convergência.

Naturalmente é necessário estabelecer um critério de parada para esse processo iterativo baseado em alguma medida, como por exemplo, quando a distância relativa entre as iterações consecutivas for menor que um valor pré-fixado ϵ, ou seja,

$$\frac{\left\|\hat{\boldsymbol{\theta}}^{(m+1)} - \hat{\boldsymbol{\theta}}^{(m)}\right\|}{\left\|\hat{\boldsymbol{\theta}}^{(m)}\right\|} \leq \epsilon. \tag{2.12}$$

2.7.5 Estimação dos parâmetro de suavização

A escolha dos parâmetros de suavização representa um ponto crucial no processo de estimação, pois estes determinam o grau de suavidade das curvas estimadas referente às funções não paramétricas. Visto isso, algumas propostas para a estimação dessas quantidades são sugeridas na literatura. Iremos apresentar os métodos de validação cruzada e validação cruzada generalizada.

De um modo geral, esses métodos consistem em fixar o parâmetro de suavização em um valor λ e em seguida eliminar a i-ésima observação, para então usar o modelo ajustado para predizer o valor dessa observação. Dessa forma, o objetivo desse método é encontrar o valor do parâmetro de suavização que minimiza esse erro de predição, que pode ser representado em função dos parâmetros. Para o modelo em questão, as funções de validação cruzada para λ_μ e λ_ϕ podem ser escritas como:

$$\mathrm{VC}(\lambda_{\boldsymbol{\mu}}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\Psi^{(2,0)}(\hat{\mu}_i, \phi_i)\left[y_i - \hat{\mu}_i\right]^2}{\left[1 - \mathrm{h}_{ii}(\boldsymbol{\mu})\right]^2}\right\}$$

e

$$\mathrm{VC}(\lambda_{\boldsymbol{\phi}}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\Psi^{(0,2)}(\hat{\mu}_i, \hat{\phi}_i)\left[\Psi^{(1,1)}(\hat{\mu}_i, \hat{\phi}_i)(y_i - \hat{\mu}_i) + T(y_i) + \Psi^{(0,1)}(\hat{\mu}_i, \hat{\phi}_i)\right]^2}{\left[1 - \mathrm{h}_{ii}(\boldsymbol{\phi})\right]^2}\right\},$$

em que $\mathrm{h}_{ii}(\boldsymbol{\mu})$ representa o i-ésimos elemento da diagonal da matriz $\mathbf{H}_{\boldsymbol{\mu}}$ e $\mathrm{h}_{ii}(\boldsymbol{\phi})$ o i-ésimos elemento da diagonal da matriz $\mathbf{H}_{\boldsymbol{\phi}}$.

Com isso, os estimadores dos parâmetros de suavização são obtidos minimizando as funções $\mathrm{VC}(\lambda_{\boldsymbol{\mu}})$ e $\mathrm{VC}(\lambda_{\boldsymbol{\phi}})$ em $\lambda_{\boldsymbol{\mu}}$ e $\lambda_{\boldsymbol{\phi}}$, respectivamente. No entanto, esse procedimento pode ser custoso do ponto de vista computacional, pois o mesmo envolve muitas operações matemáticas. Uma alternativa é substituir $h_{ii}(\boldsymbol{\mu})$ e $h_{ii}(\boldsymbol{\phi})$ por suas respectivas aproximações, $n^{-1}\mathrm{tr}(\boldsymbol{H}_{\boldsymbol{\mu}})$ e $n^{-1}\mathrm{tr}(\boldsymbol{H}_{\boldsymbol{\phi}})$, e então obter expressões simplificadas, ou seja,

$$\mathrm{VCG}(\lambda_{\boldsymbol{\mu}}) = \frac{\left\|\hat{\boldsymbol{\Psi}}^{(2,0)}\left[\boldsymbol{y} - \hat{\boldsymbol{\mu}}\right]^2\right\|}{\left[n - \mathrm{tr}\{\boldsymbol{H}_{\phi}\}\right]^2}$$

e

$$\text{VCG}(\lambda_{\boldsymbol{\phi}}) \quad = \quad \frac{\left\| \hat{\boldsymbol{\Psi}}^{(0,2)} \left[\hat{\boldsymbol{\Psi}}^{(1,1)} (\boldsymbol{y} - \hat{\boldsymbol{\mu}}_i) + T(\boldsymbol{y}) + \hat{\boldsymbol{\Psi}}^{(0,1)} \right]^2 \right\|}{[n - \text{tr}\{\mathbf{H}_{\boldsymbol{\phi}}\}]^2}.$$

Este método é conhecido como validação cruzada generalizada, apresentando um custo computacional bem menor. Segundo este método, os estimadores dos parâmetros de suavização são obtidos minimizando as $\text{VCG}(\lambda_{\boldsymbol{\mu}})$ e $\text{VCG}(\lambda_{\boldsymbol{\phi}})$ em função de $\lambda_{\boldsymbol{\mu}}$ e $\lambda_{\boldsymbol{\phi}}$, respectivamente. Para mais detalhes a respeito dos métodos de validação cruzada e validação cruzada generalizada, ver Wahba (1990) e Simonoff (1996).

2.8 Erros padrão aproximados

Na teoria de regressão, no que diz respeito a modelos paramétricos, é bem comum considerar que os estimadores obtidos por máxima verossimilhança são assintoticamente normais e que sua matriz de covariância é o inverso da informação de Fisher. No entanto, no caso de modelos semiparamétricos vários autores vêm discutindo uma forma de obter a matriz de variância e covariância, porém ainda não há nada bem definido. Wahba (1978) mostra que, para o modelo aditivo clássico a matriz de variâncias e covariâncias podem ser estimada usando argumentos bayesianos, e que, a mesma coincide com a inversa da informação de Fisher penalizada. De forma semelhante, Wood (2006) conclui que, na classe de modelos aditivos generalizados, a matriz de variâncias e covariâncias assintótica também é estimada pelo inverso da informação de Fisher. Naturalmente, esse conceito tem se extendido nas diversas classes de modelos semiparamétricos como uma forma de aproximar os erros padrão associados as suas estimativas, como vemos em Pulgar (2009), Relvas (2013) e Manghi (2016), por exemplo.

Com isso, estendendo esse conceito é razoável supor que uma estimativa para matriz de variâncias e covariâncias é obtida por meio da inversão da matriz de informação de Fisher. Sabendo-se que os parâmetros da média são ortogonais aos da dispersão e utilizando propriedades de matrizes em blocos, então a inversa da informação de Fisher nos MNLPGS pode ser representada pela seguinte estrutura matricial:

$$\boldsymbol{K}_p(\boldsymbol{\theta}) = \begin{pmatrix} K_p^{\boldsymbol{\mu}}(\boldsymbol{\theta}) & 0 \\ 0 & K_p^{\boldsymbol{\phi}}(\boldsymbol{\theta}) \end{pmatrix},$$

em que

$$\boldsymbol{K}_p^{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \begin{pmatrix} (\tilde{\boldsymbol{X}}^\top \boldsymbol{W}_\beta \tilde{\boldsymbol{X}})^{-1} & -\boldsymbol{E}_{\boldsymbol{\mu}} \\ -\boldsymbol{E}_{\boldsymbol{\mu}}^\top & (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{W}_{\boldsymbol{\mu}} \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1} \end{pmatrix},$$

com $\boldsymbol{E}_{\boldsymbol{\mu}} = (\tilde{\boldsymbol{X}}^\top \boldsymbol{W}_\beta \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1}$ e

$$K_p^{\boldsymbol{\phi}}(\boldsymbol{\theta}) = \begin{pmatrix} -(\tilde{\boldsymbol{S}}^\top \boldsymbol{W}_\gamma \tilde{\boldsymbol{S}})^{-1} & \boldsymbol{E}_{\boldsymbol{\phi}} \\ \boldsymbol{E}_{\boldsymbol{\phi}}^\top & -(\boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{W}_{\boldsymbol{\phi}} \boldsymbol{N}_{\boldsymbol{\phi}} - \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1} \end{pmatrix},$$

com $\boldsymbol{E}_{\boldsymbol{\phi}} = (\tilde{\boldsymbol{S}}^\top \boldsymbol{W}_\gamma \tilde{\boldsymbol{S}})^{-1} \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{N}_{\boldsymbol{\phi}} (\boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{N}_{\boldsymbol{\phi}} - \lambda_{\boldsymbol{\phi}} \boldsymbol{K}_{\boldsymbol{\phi}})^{-1}$.

Logo, as matrizes de variâncias e covariâncias assintóticas dos estimadores, $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$, são dadas, respectivamente, por:

$$\mathrm{Cov}_{\mathrm{A}}(\hat{\boldsymbol{\beta}}) \;=\; (\tilde{\boldsymbol{X}}^\top \boldsymbol{W}_{\boldsymbol{\beta}} \tilde{\boldsymbol{X}})^{-1}$$

e

$$\mathrm{Cov}_{\mathrm{A}}(\hat{\boldsymbol{\gamma}}) \;=\; -(\tilde{\boldsymbol{S}}^\top \boldsymbol{W}_{\boldsymbol{\gamma}} \tilde{\boldsymbol{S}})^{-1}.$$

Consequentemente, as matrizes de variâncias e covariâncias assintóticas relacionadas às componentes não paramétricas estimadas, $\hat{\boldsymbol{f}}_{\boldsymbol{\mu}}$ e $\hat{\boldsymbol{f}}_{\boldsymbol{\mu}}$, são expressas, respectivamentes, por:

$$\mathrm{Cov}_{\mathrm{A}}(\hat{\boldsymbol{f}_{\boldsymbol{\mu}}}) \;=\; (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{W}_{\boldsymbol{f}_{\boldsymbol{\mu}}} \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1}$$

e

$$\mathrm{Cov}_{\mathrm{A}}(\hat{\boldsymbol{f}_{\boldsymbol{\phi}}}) \;=\; (\boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{W}_{\boldsymbol{f}_{\boldsymbol{\phi}}} \boldsymbol{N}_{\boldsymbol{\phi}} + \lambda_{\boldsymbol{\phi}} \boldsymbol{K}_{\boldsymbol{\phi}})^{-1}.$$

2.9 Seleção de modelos

Como é amplamente discutido em modelos paramétricos, a comparação entre modelos concorrentes (com o objetivo de selecionar o melhor modelo) é algo de suma importância. Dentre os métodos para a seleção de modelos, destacam-se o critério de informação de Akaike, proposto por Akaike (1974), e o critério de informação de Schwarz, apresentado em Schwarz et al. (1978). Essas medidas são fundamentadas em pontuar o modelo por meio da sua verossimilhança e penalizar por sua quantidade de parâmetros, servindo então de indicativo para o modelo mais parcimonioso.

Naturalmente, autores tem considerado uma analogia aos critérios de seleção nos modelos paramétricos para o caso de modelos semiparamétricos, vide Pulgar (2009), Relvas (2013), Noda (2013) e Manghi (2016). Nesse caso, é substituída a função de verossimilhança por sua forma penalizada, visto que todo processo de estimação esta em termos dessa penalização. Além disso, devemos levar em consideração uma medida para o número efetivo de parâmetros no modelo, pois diferentemente dos modelos paramétricos, há uma dificuldade adicional para mensurarmos isso. Uma quantidade viável para indicar esse número é denominada de graus de liberdade efetivos (GLE), e para os MNLPGS, pode

ser obtida da seguinte forma:

$$\text{GLE} \quad = \quad \text{tr}\{\mathbf{H}_{\boldsymbol{\mu}}\} + \text{tr}\{\mathbf{H}_{\boldsymbol{\phi}}\}.$$

Visto isso, o critério da informação de Akaike penalizado (AICp) para os modelos não lineares parciais generalizados superdispersados, pode ser definido como:

$$\text{AIC}_p(\hat{\boldsymbol{\theta}}) \quad = \quad -2l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2\text{tr}\{\mathbf{H}_{\boldsymbol{\mu}}\} + 2\text{tr}\{\mathbf{H}_{\boldsymbol{\phi}}\},$$

em que $l_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ representa a função de verossimilhança penalizada avaliada na EMVp de $\boldsymbol{\theta}$.

Já o critério de informação bayesiano penalizado (BICp) para essa classe de modelos, pode ser definido pela seguinte expressão:

$$\text{BIC}_p(\hat{\boldsymbol{\theta}}) \quad = \quad -2l_p(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{f}}_{\boldsymbol{\mu}}; \lambda_\mu) + \log(n)\text{tr}\{\mathbf{H}(\boldsymbol{\mu})\} + \log(n)\text{tr}\{\mathbf{H}(\boldsymbol{\phi})\}.$$

Note que, em ambos os critérios, o interesse é em escolher o modelo que apresente o menor valor para o critério utilizado, dentre todos os possíveis modelos em estudo.

CAPÍTULO 3

Métodos de diagnóstico em MNLPGS

As técnicas de diagnóstico são amplamente utilizadas para avaliar a qualidade de ajuste do modelo e a sua sensibilidade. Assim, na avaliação da qualidade de ajuste, é necessário verificar se há possíveis afastamentos das suposições feitas ao modelo. Já na metodologia da análise de sensibilidade, devem-se avaliar as variações dos resultados quando o modelo sofre ligeiras modificações na sua formulação inicial. Caso estas alterações modifiquem significativamente as conclusões do modelo, então é necessário ter certos cuidados nas conclusões do modelo ou optar por uma proposta alternativa de modelo que melhor se adeque aos dados. Por outro lado, dentro dessas duas metodologias de análise, são considerados diversos métodos e quantidades para o diagnóstico do modelo sob estudo.

Inicialmente, uma forma de avaliar a qualidade de ajuste do modelo proposto é verificando se as suas suposições relacionadas a componente aleatória são satisfeitas. Nesse ponto de vista, a análise de resíduos visa verificar a discrepância entre os valores ajustados e as observações dos dados, a qual apresenta a capacidade em identificar a presença de pontos aberrantes, ou seja, pontos que interferem de maneira desproporcional nos valores ajustados. Além disso, essa análise permite verificar adequação distribucional da variável resposta e a independência dos erros.

Uma outra análise de grande importância em diagnóstico de modelos é a verificação da presença de observações que exercem pesos desproporcionais nos valores preditos do modelo, denominadas de pontos de alavanca. Para identificar esses pontos existem propostas baseadas em medidas de alavancagem, que podem ser obtidas por meio da matriz de projeção (matriz *hat*) ou por meio de uma medida bem mais geral, conhecida como alavancagem generalizada. De forma geral, a alavancagem generalizada é medida através da amplitude da derivada dos valores preditos em relação aos valores da variável resposta, ou seja, $d\hat{y}_i/dy_i$. No caso em que modela-se simultaneamente a média e a dispersão, há a necessidade de obter medidas de alavancagem para cada uma das componentes sis-

temáticas. Além disso, embora a matriz *hat* definida no contexto semiparamétrico não seja necessariamente projetora, autores a consideram como uma medida de alavancagem viável.

Outro tópico importante e que vem recebendo crescente destaque no que diz respeito a técnicas de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Em outras palavras, uma observação é dita ser influente para determinado parâmetro se, com a sua exclusão, obtemos alterações significativas na análise do modelo. Nesse contexto, as medidas de influência global surgiram com uma forma de mensurar quais observações apresentam grandes influências na estimação dos parâmetros do modelo, obtidas especificamente ao estimar esses parâmetros após a retirada individual de cada uma das observações. Por outro lado, o estudo de influência de um ponto de vista bem mais geral é denominado na literatura como análise de influência local, onde o interesse é em verificar a influência baseado em diferentes perturbações, ao invés de verificarmos apenas o impacto na retirada individual de observações. Essas perturbações representam pequenas alterações incorporadas no modelo ou nos dados, com o intuito de identificar a presença de diferentes fatores que podem influenciar o modelo. No presente trabalho será considerado apenas o estudo de influência local em termos da função de verossimilhança penalizada, visto que este generaliza também o caso da análise de influência global.

Dessa forma, no presente capítulo são abordadas algumas propostas para a análise de diagnóstico, especificamente direcionado para a classe de modelos não lineares parciais generalizados superdispersados. A princípio, é obtida a matriz de informação de Fisher observada derivando-se a função de verossimilhança penalizada. Em seguida são discutidos resultados acerca das medidas de alavancagem associadas ao modelo proposto e a análise de resíduos são discutidos. Finalmente, consideramos uma breve abordagem de influência local, apresentada sob alguns esquemas específicos de perturbações.

3.1 Informação de Fisher Observada

Essa matriz hessiana é fundamental para a obtenção de medidas de diagnóstico, como será visto nas próximas seções deste capítulo, sendo utilizada tanto para construir uma medida de alavancagem bem geral como também no contexto de influência local. A matriz de informação de Fisher observada $\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}}$ é definida pela segunda derivada de $l_p(\boldsymbol{\theta})$ em relação a $\boldsymbol{\theta}$, ou seja,

$$\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}} = \frac{\partial^2 l_p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \tag{3.1}$$

Dessa forma, para os modelos não lineares parciais generalizados superdispersados,

essa matriz assume a seguinte representação matricial:

$$\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}} = \begin{pmatrix} \ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta\beta}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta f_\mu}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta\gamma}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta f_\phi}} \\ \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu\beta}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu f_\mu}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu\gamma}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu f_\phi}} \\ \ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma\beta}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma f_\mu}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma\gamma}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma f_\phi}} \\ \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi\beta}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi f_\mu}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi\gamma}} & \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi f_\phi}} \end{pmatrix},$$

em que

$$\begin{aligned}
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta\beta}} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(3,0)} \boldsymbol{M}_1^2 \boldsymbol{U} \tilde{\boldsymbol{X}} + \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_2 \boldsymbol{U} \tilde{\boldsymbol{X}} + [\boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu})^\top][\tilde{\tilde{\boldsymbol{X}}}] \\
&\quad - \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta f_\mu}} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(3,0)} \boldsymbol{M}_1^2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}} + \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}} - \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta\gamma}} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,1)} \boldsymbol{M}_1 \boldsymbol{\Phi}_1 \boldsymbol{U} \tilde{\boldsymbol{S}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta f_\phi}} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,1)} \boldsymbol{M}_1 \boldsymbol{\Phi}_1 \boldsymbol{N}_{\boldsymbol{\phi}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu\beta}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(3,0)} \boldsymbol{M}_1^2 \boldsymbol{U} \tilde{\boldsymbol{X}} + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_2 \boldsymbol{U} \tilde{\boldsymbol{X}} - \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu f_\mu}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(3,0)} \boldsymbol{M}_1^2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}} + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}} - \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} - \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu\gamma}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,1)} \boldsymbol{M}_1 \boldsymbol{\Phi}_1 \boldsymbol{U} \tilde{\boldsymbol{S}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu f_\phi}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,1)} \boldsymbol{M}_1 \boldsymbol{\Phi}_1 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma\beta}} &= \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(2,1)} \boldsymbol{M}_1 \boldsymbol{\Phi}_1 \boldsymbol{U} \tilde{\boldsymbol{X}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma f_\mu}} &= \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(2,1)} \boldsymbol{M}_1 \boldsymbol{\Phi}_1 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma\gamma}} &= \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{U} \tilde{\boldsymbol{S}} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_2 \boldsymbol{U} \tilde{\boldsymbol{S}} + [\boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 (\boldsymbol{y} - \boldsymbol{\mu})^\top][\tilde{\tilde{\boldsymbol{S}}}] + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Phi}_2 \boldsymbol{T} \tilde{\boldsymbol{S}} \\
&\quad + [\boldsymbol{T}) \boldsymbol{\Phi}_1][\tilde{\tilde{\boldsymbol{S}}}] + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \tilde{\boldsymbol{S}} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_2 \tilde{\boldsymbol{S}} + [\boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1][\tilde{\tilde{\boldsymbol{S}}}], \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma f_\phi}} &= \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\phi}} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\phi}} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Phi}_2 \boldsymbol{T} \boldsymbol{N}_{\boldsymbol{\phi}} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{N}_{\boldsymbol{\phi}} \\
&\quad + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_2 \boldsymbol{N}_{\boldsymbol{\phi}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi\beta}} &= \boldsymbol{N}_{\boldsymbol{\phi}} \boldsymbol{\Psi}^{(2,1)} \boldsymbol{\Phi}_1 \boldsymbol{M}_1 \boldsymbol{U} \tilde{\boldsymbol{X}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi f_\mu}} &= \boldsymbol{N}_{\boldsymbol{\phi}} \boldsymbol{\Psi}^{(2,1)} \boldsymbol{\Phi}_1 \boldsymbol{M}_1 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\mu}}, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi\gamma}} &= \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{U} \tilde{\boldsymbol{S}} + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_2 \boldsymbol{U} \tilde{\boldsymbol{S}} + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Phi}_2 \boldsymbol{T} \tilde{\boldsymbol{S}} + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \tilde{\boldsymbol{S}} \\
&\quad + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_2 \tilde{\boldsymbol{S}}
\end{aligned}$$

e

$$\begin{aligned}
\ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi f_\phi}} &= \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\phi}} + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_2 \boldsymbol{U} \boldsymbol{N}_{\boldsymbol{\phi}} + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Phi}_2 \boldsymbol{T} \boldsymbol{N}_{\boldsymbol{\phi}} + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{N}_{\boldsymbol{\phi}} \\
&\quad + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_2 \boldsymbol{N}_{\boldsymbol{\phi}} - \lambda_{\boldsymbol{\phi}} \boldsymbol{K}_{\boldsymbol{\phi}}.
\end{aligned}$$

com $\boldsymbol{U} = \text{diag}\{(y_1 - \mu_1), \ldots, (y_n - \mu_n)\}$, $\boldsymbol{T} = \text{diag}\{T(y_1), \ldots, T(y_n)\}$, $\tilde{\tilde{X}} = \partial^2 \boldsymbol{\eta} / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$ e $\tilde{\tilde{S}} = \partial^2 \boldsymbol{\tau} / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top$.

Note que, $\tilde{\tilde{\boldsymbol{S}}}$ representa uma *array* de dimensão $n \times p \times p$ e, consequentemente, a notação [.][.] simboliza a multiplicação de uma matriz por uma array, a qual é denominada de produto colchete e sua definição pode ser encontrada em Wei et al. (1998).

3.2 Alavancagem

De modo geral, a análise de alavancagem consiste em medir a variação dos valores preditos quando há um acréscimo de um infinitésimo em seus respectivos valores observados. É possível detectar pontos de alavanca estudando a diagonal principal das matrizes *hat* $\boldsymbol{H_\mu}$ e $\boldsymbol{H_\phi}$, obtidas no capítulo anterior. A detecção é feita de tal modo que, os valores da diagonal principal da matriz $\boldsymbol{H_\mu}$ que forem razoavelmente grandes, indicam que sua respectiva observação seja um possível ponto de alavanca no modelo da média. De maneira similar, altos valores na diagonal principal de $\boldsymbol{H_\phi}$ idicam possíveis pontos de alavanca no modelo para dispersão. Sabendo-se que os elementos da diagonal principal da matriz *hat* representam o efeito de cada observação na estimação da média ou da dispersão, então uma medida de corte comumente utilizada é duas vezes o valor médio de da diagonal principal da matriz *hat*. Para a estimação da média, esse valor é representado por $2\text{tr}\{\boldsymbol{H_\mu}\}/n$ e, de forma equivalente, no que diz respeito a dispersão o valor $2\text{tr}\{\boldsymbol{H_\phi}\}/n$ é usado como referência.

Uma outra forma de detectar pontos de alavanca é abordada por Wei et al. (1998), apresentando uma medida bem geral, denominada de alavancagem generalizada. Nesse caso, uma maneira de medir esse efeito é através da amplitude da derivada dos valores preditos em relação aos valores da variável resposta, ou seja, $d\hat{y}_i/dy_i$, que segundo Wei et al. (1998), pode ser obtida da seguinte forma:

$$\text{GL}(\boldsymbol{\theta}) = \frac{\partial \hat{\boldsymbol{y}}}{\partial \boldsymbol{y}} \;=\; \left\{ \boldsymbol{D_\theta} (-\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}})^{-1} \ddot{\boldsymbol{L}}_p^{\boldsymbol{y\theta}} \right\} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

em que $\boldsymbol{D_\theta} = \partial \boldsymbol{\mu}/\partial \boldsymbol{\theta}^\top$ e $\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta y}} = \partial^2 L_p(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{y}$.

Como já foram calculadas as expressões para $\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}}$, devemos agora obter as matrizes $\boldsymbol{D_\theta}$ e $\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta y}}$, derivando-se a função de verossimilhança penalizada. Com isso, obtemos o vetor $\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta y}}$ que pode ser representado pela seguinte partição:

$$\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta y}} \;=\; \begin{pmatrix} \ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta y}} \\ \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\mu y}} \\ \ddot{\boldsymbol{L}}_p^{\boldsymbol{\gamma y}} \\ \ddot{\boldsymbol{L}}_p^{\boldsymbol{f_\phi y}} \end{pmatrix}.$$

Usando a notação $\boldsymbol{T}' = \partial \boldsymbol{T}/\partial \boldsymbol{y}$, então os elementos desse vetor podem ser expressos

por meio das seguintes equações matriciais:

$$\begin{aligned}
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta y}} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta y}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1, \\
\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta y}} &= \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Phi}_1 \boldsymbol{T}'
\end{aligned}$$

e

$$\ddot{\boldsymbol{L}}_p^{\boldsymbol{\beta y}} = \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Phi}_1 \boldsymbol{T}'.$$

Considere agora que o vetor $\boldsymbol{D}_{\boldsymbol{\theta}}$ de derivadas da verossimilhança possa ser particionado da seguinte maneira:

$$\boldsymbol{D}_{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{D}_{\boldsymbol{\beta}} \\ \boldsymbol{D}_{\boldsymbol{f_\mu}} \\ \boldsymbol{D}_{\boldsymbol{\gamma}} \\ \boldsymbol{D}_{\boldsymbol{f_\phi}} \end{pmatrix}.$$

Sob a modelagem da média, obtemos as seguintes expressões:

$$\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\beta}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = \boldsymbol{M}_1 \tilde{\boldsymbol{X}}, \\
\boldsymbol{D}_{\boldsymbol{f_\mu}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{f_\mu}} = \boldsymbol{M}_1 \boldsymbol{N}_{\boldsymbol{\mu}}, \\
\boldsymbol{D}_{\boldsymbol{\gamma}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\gamma}} = 0
\end{aligned}$$

e

$$\boldsymbol{D}_{\boldsymbol{f_\phi}} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{f_\phi}} = 0.$$

Analogamente, para o modelo da dispersão, temos as seguintes expressões:

$$\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\beta}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = 0, \\
\boldsymbol{D}_{\boldsymbol{f_\mu}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{f_\mu}} = 0, \\
\boldsymbol{D}_{\boldsymbol{\gamma}} &= \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\gamma}} = \boldsymbol{\Phi}_1 \tilde{\boldsymbol{S}}
\end{aligned}$$

e

$$\boldsymbol{D}_{\boldsymbol{f_\phi}} = \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{f_\phi}} = \boldsymbol{\Phi}_1 \boldsymbol{N}_{\boldsymbol{\phi}}.$$

Logo, utilizando a matriz de informação de Fisher definida em (3.1), $\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta y}}$ e $\boldsymbol{D_\theta}$, obtemos então uma medida de alavancagem que na prática pode ser considerada na detecção de pontos que exercem influência desproporcionais nos valores estimados. Visto isso, uma forma de detectar essas observações é verificar quais valores da diagonal principal de GL ultrapassam o limiar de $2\text{tr}(\text{GL})/n$, isto é, os casos que $\text{GL}_{ii} \geq 2\text{tr}(\text{GL})/n$ implica que a i-ésima observação é um possível ponto de alavanca.

3.3 Análise de resíduos

Os resíduos têm grande utilidade na análise de regressão para validação de determinadas suposições de modelos estatísticos. É possível utilizá-los para verificar a homocedasticidade, existência de pontos discrepantes, a adequação da distribuição proposta para a variável resposta e independência dos erros. Sucintamente, os resíduos são os valores que medem a discrepância entre os valores observados da variável resposta e seus correspondentes valores ajustados. Na literatura, há uma ampla discussão e inúmeras medidas que visam uma forma de avaliar essa discrepância, no entanto vamos tratar apenas dos resíduos ordinários e uma versão padronizada dos mesmos.

No caso dos modelos não lineares parciais generalizados superdispersados, nossa proposta inicial é considerar o conceito de resíduos ordinários. Esses, por sua vez, são obtidos ao reescrever os processos escore, apresentado no capítulo anterior, sobre a forma da solução de mínimos quadrados reponderados. Desse modo, levando em consideração o processo escore de Fisher usado para estimação dos parâmetros da média, então o vetor de resíduos ordinários para a média é definido por:

$$\boldsymbol{r_\mu} \;=\; \left[\hat{\boldsymbol{\Psi}}^{(2,0)}\boldsymbol{M}_1^2\right]^{1/2}(\hat{\boldsymbol{z}}-\hat{\boldsymbol{\eta}}) = \left[\hat{\boldsymbol{\Psi}}^{(2,0)}\right]^{1/2}(\boldsymbol{y}-\hat{\boldsymbol{\mu}}).$$

em que $\boldsymbol{M}_1^2$ e $\boldsymbol{\Psi}^{(2,0)}$ representam as mesmas quantidades definidas no Capítulo 2.

Por outro lado, como cada resíduo pode possuir uma variância diferente, é mais adequado expressá-los em uma forma padronizada. Uma padronização desse resíduo é feita por meio da sua respectiva medida de alavancagem, que nesse caso, pode ser obtida pela matriz $\boldsymbol{H}_\mu$, discutida anteriormente. Considere $h_{\boldsymbol{\mu}ii}$ o i-ésimo elemento da diagonal principal de $\boldsymbol{H}_\mu$, então o i-ésimo resíduo padronizado para a média é dado por:

$$r^*_{\boldsymbol{\mu}i} \;=\; \frac{(y_i-\hat{\mu}_i)}{[\Psi^{(2,0)}(\hat{\mu}_i,\hat{\phi}_i)]^{-1/2}\sqrt{1-h_{\boldsymbol{\mu}ii}}}.$$

Para avaliar a presença de observações discrepantes e alguma tendência não linear dos resíduos, são sugeridos os gráficos dos resíduos versus índices e resíduos versus valores ajustados, de acordo com o resíduo utilizado. Além disso, para verificar a adequacidade da distribuição usada o gráfico de quantis-quantis com envelopes simulados é aconselhável.

3.4 Influência local

Motivado em avaliar a presença de observações que exercem mudanças na inferência do modelo, Cook (1986) propõe avaliar a influência conjunta das observações sob pequenas mudanças no modelo e/ou nos dados. Originalmente elaborada para o modelo normal clássico, essa metodologia foi denominada de influência local e seu conceito vem sendo estendido para diversas classes de modelos, tornando-se um dos métodos mais sofisticados de diagnóstico. De forma resumida, a influência local é obtida modificando o modelo, por meio de algum tipo de perturbação, e assim quantificar o peso que determinadas observações exercem sobre o modelo, medida por meio da distância entre as verossimilhanças do modelo sob perturbação e do modelo sem perturbação. No caso de modelos semiparamétricos, é mais coerente usar a versão penalizada da função de verossimilhança, pois há possibilidade de incluir perturbações também nas suas penalizações.

Considere o modelo não linear parcial generalizado superdispersado, cuja função de verossimilhaça penalizada $l_p(\boldsymbol{\theta})$ é dada em (2.7), então $l_p(\boldsymbol{\theta}|\boldsymbol{w})$ irá representar essa verossimilhança ao incluir um vetor de perturbação $\boldsymbol{w}$. Assuma que existe algum vetor de não perturbação $\boldsymbol{w}_0$, satisfazendo $l_p(\boldsymbol{\theta}|\boldsymbol{w}) = l_p(\boldsymbol{\theta})$. Assim, temos o interesse em medir a distância entre essas verossimilhanças, ou seja,

$$\mathrm{LD}(\boldsymbol{w}) \;=\; 2\{l_p(\hat{\boldsymbol{\theta}}) - l_p(\hat{\boldsymbol{\theta}}_w)\} \geq 0,$$

em que $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}}_w$ representam as estimativas de máxima verossimilhança penalizada do modelo sem e com perturbação, respectivamente.

Dessa forma, verificamos o grau de influência do modelo através do afastamento entre as verossimilhanças. Em outras palavras, a ideia básica desse método é estudar o comportamento da função $LD(\boldsymbol{\delta})$ em torno de $\boldsymbol{\delta}_0$. O procedimento procura selecionar uma direção ao redor de $\boldsymbol{\delta}$, $\boldsymbol{\delta} + a\boldsymbol{l}$ ($||\boldsymbol{l}|| = 1$), e então fazer uma análise do gráfico de $LD(\boldsymbol{\delta}_0 + a\boldsymbol{l})$ em termos dos valores de a, sendo $a \in \mathbb{R}$. Esse estudo pode então ser associado com a curvatura normal $\boldsymbol{C}_l(\boldsymbol{\theta})$ na direção $\boldsymbol{l}$, dada por:

$$\boldsymbol{C}_l(\boldsymbol{\theta}) \;=\; 2\left|\boldsymbol{l}^\top \boldsymbol{\Delta}_p^\top (\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}})^{-1} \boldsymbol{\Delta}_p \boldsymbol{l}\right|, \tag{3.2}$$

em que $\boldsymbol{\Delta}_p = \partial^2 L_p(\boldsymbol{\theta}|\boldsymbol{w})/\partial\boldsymbol{\theta}\partial\boldsymbol{w}^\top$ é uma matriz de perturbações.

No entanto, há casos em que o interesse reside em avaliar a influência apenas relativa a um subvetor dos parâmetros $\boldsymbol{\theta}_1$, de tal forma que $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ é uma partição do vetor de parâmetros. Nesse caso, a curvatura normal fica dada pela seguinte equação:

$$\boldsymbol{C}_l(\boldsymbol{\theta}) \;=\; 2\left|\boldsymbol{l}^\top \boldsymbol{\Delta}_p^\top \left[(\ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta\theta}})^{-1} - \boldsymbol{B}_1^{-1}\right] \boldsymbol{\Delta}_p \boldsymbol{l}\right|, \tag{3.3}$$

em que $\boldsymbol{B}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\boldsymbol{L}}_p^{\boldsymbol{\theta_2\theta_2}} \end{pmatrix}$.

Logo, em ambos os casos, uma sugestão é de considerar a maior curvatura, na qual possui a direção $\boldsymbol{l}_{max}$, como forma de avaliar a influência local das observações nas estimativas dos parâmetros. É importante ressaltar que, a matriz de perturbações Δ_p é obtida sob diversos tipos de alterações no modelo, denominadas de perturbações. Serão abordadas no presente texto apenas três perturbações no modelo, a saber: aditiva na resposta, aditiva nos preditores e ponderações de casos. Note que, as perturbações irão interferir diretamente na função de verossimilhança do modelo, indicando então que a matriz Δ_p apresentará expressões distintas para cada tipo de perturbação, as quais serão explicitadas nas seções a seguir.

3.4.1 Perturbação aditiva na resposta

Esse esquema de perturbação consiste em verificar a sensibilidade do modelo ao inserir perturbações de forma aditiva na variável resposta, i.e., $y_{iw} = y_i + w$. Logo, considerando a forma perturbada da variável resposta $\boldsymbol{y}_w$, então o logaritmo da função de verossimilhança penalizada pode ser definida como:

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{w}) &= \sum_{i=1}^{n} \left\{ (y_{i\boldsymbol{w}} - \mu_i)\Psi^{(1,0)}(\mu_i, \phi_i) + \phi_i T(y_{i\boldsymbol{w}}) + \Psi(\mu_i, \phi_i) \right\} + \sum_{i}^{n} \log A(y_i). \\ &\quad -\frac{1}{2}\lambda_\mu^* \boldsymbol{f}_\mu^\top \boldsymbol{K}_\mu \boldsymbol{f}_\mu - \frac{1}{2}\lambda_\phi^* \boldsymbol{f}_\phi^\top \boldsymbol{K}_\phi \boldsymbol{f}_\phi. \end{aligned}$$

Calculando os elementos da matriz de perturbações através da derivação dessa função, obtemos as seguintes expressões:

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1, \\ \boldsymbol{\Delta}_{\boldsymbol{f_\mu}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1, \\ \boldsymbol{\Delta}_{\boldsymbol{\gamma}} &= \tilde{S}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 + \tilde{S}^\top \boldsymbol{\Phi}_1 \boldsymbol{T}_{\boldsymbol{w}}' \end{aligned}$$

e

$$\boldsymbol{\Delta}_{\boldsymbol{f_\phi}} = \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Phi}_1 \boldsymbol{T}_{\boldsymbol{w}}',$$

em que $\boldsymbol{T}_{\boldsymbol{w}}' = \partial \boldsymbol{T} / \partial \boldsymbol{y}_{\boldsymbol{w}}$.

3.4.2 Perturbação aditiva nos preditores

De forma semelhante, o intuito dessa perturbação é avaliar a sensibilidade do modelo ao introduzir modificações nas variáveis explicativas, podendo ser interpretada como um erro de medição. Desse modo, considere a perturbação aditiva para i-ésima observação da k-ésima variável explicativa na forma $x_{ikw} = x_{ik} + w_i$ e $s_{ikw} = s_{ikw} + w_i$, assim sua função

de verossimilhança penalizada é dada por:

$$\begin{aligned} l_p(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left\{ (y_i - \mu_{i\boldsymbol{w}})\Psi^{(1,0)}(\mu_{i\boldsymbol{w}}, \phi_{i\boldsymbol{w}}) + \phi_{i\boldsymbol{w}} T(y_i) + \Psi(\mu_{i\boldsymbol{w}}, \phi_{i\boldsymbol{w}}) \right\} + \sum_{i}^{n} \log A(y_i) \\ &\quad -\frac{1}{2}\lambda_\mu^* \boldsymbol{f}_\mu^\top \boldsymbol{K}_\mu \boldsymbol{f}_\mu - \frac{1}{2}\lambda_\phi^* \boldsymbol{f}_\phi^\top \boldsymbol{K}_\phi \boldsymbol{f}_\phi. \end{aligned}$$

Dessa maneira, obtendo os elementos da matriz de perturbação $\boldsymbol{\Delta}_p$, temos os seguintes resultados:

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \tilde{\boldsymbol{X}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(3,0)} \boldsymbol{M}_1^2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{F}_1 + \tilde{\boldsymbol{X}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{F}_1 + [\boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 \boldsymbol{U}_{\boldsymbol{w}}][\boldsymbol{F}_2] - \tilde{\boldsymbol{X}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{F}_1, \\ \boldsymbol{\Delta}_{\boldsymbol{f_\mu}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(3,0)} \boldsymbol{M}_1^2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{F}_1 + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{F}_1 - \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{F}_1, \\ \boldsymbol{\Delta}_{\boldsymbol{\gamma}} &= \tilde{\boldsymbol{S}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(1,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{G}_1 + \tilde{\boldsymbol{S}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{G}_1 + [\boldsymbol{\Psi}^{(1,1)} \Phi_1 \boldsymbol{U}][\boldsymbol{G}_2] + \tilde{\boldsymbol{S}}_{\boldsymbol{w}}^\top \boldsymbol{\Phi}_2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{T} \boldsymbol{G}_1 \\ &\quad + [\boldsymbol{T}\boldsymbol{\Phi}][\boldsymbol{G}_2] + \tilde{\boldsymbol{S}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{G}_1 + \tilde{\boldsymbol{S}}_{\boldsymbol{w}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_2 \boldsymbol{G}_1 + [\boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1][\boldsymbol{G}_2] \end{aligned}$$

e

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{f_\phi}} &= \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{G}_1 + \boldsymbol{N}_{\boldsymbol{\phi}}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{G}_1 + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Phi}_2 \boldsymbol{U}_{\boldsymbol{w}} \boldsymbol{T} \boldsymbol{G}_1 + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(0,2)} \boldsymbol{\Phi}_1^2 \boldsymbol{G}_1 \\ &\quad + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_2 \boldsymbol{G}_1. \end{aligned}$$

em que $\boldsymbol{U}_{\boldsymbol{w}} = \text{diag}\{(y_1 - \mu_{1\boldsymbol{w}}), \ldots, (y_n - \mu_{n\boldsymbol{w}})\}$.

3.4.3 Perturbação de casos ponderados

A ponderação de casos visa verificar a contribuição individual de cada observação no processo de estimação, indicando quais exercem contribuições desproporcionais ao modelo. Nesse caso, são atribuídos pesos diferentes para cada observação, e então é mensurada a sensibilidade do modelo ao sofrer essas modificações. Considere o vetor de não perturbação $\boldsymbol{w}_0 = (1, \ldots, 1)$ e o vetor de perturbação $\boldsymbol{w} = (w_1, \ldots, w_n)$, então inserindo essa ponderação o logaritmo da função de verossimilhança em sua forma penalizada fica expressa por:

$$\begin{aligned} l_p(\boldsymbol{\theta}|\boldsymbol{w}) &= \sum_{i=1}^{n} w_i \left\{ (y_i - \mu_i)\Psi^{(1,0)}(\mu_i, \phi_i) + \phi_i T(y_i) + \Psi(\mu_i, \phi_i) \right\} + \sum_{i}^{n} w_i \log A(y_i) \\ &\quad -\frac{1}{2}\lambda_\mu^* \boldsymbol{f}_\mu^\top \boldsymbol{K}_\mu \boldsymbol{f}_\mu - \frac{1}{2}\lambda_\phi^* \boldsymbol{f}_\phi^\top \boldsymbol{K}_\phi \boldsymbol{f}_\phi. \end{aligned}$$

Novamente, calculando-se as derivadas de $l_p(\boldsymbol{\theta}|\boldsymbol{w})$, então as expressões para $\boldsymbol{\Delta}_p$ ficam dadas por:

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{\beta}} &= \tilde{X}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 \boldsymbol{U}, \\ \boldsymbol{\Delta}_{\boldsymbol{f_\mu}} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 \boldsymbol{U}, \\ \boldsymbol{\Delta}_{\boldsymbol{\gamma}} &= \tilde{S}^\top \boldsymbol{\Psi}^{(1,1)} \boldsymbol{\Phi}_1 \boldsymbol{U} + \tilde{S}^\top \boldsymbol{\Phi}_1 \boldsymbol{T}'_{\boldsymbol{w}} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\Psi}^{(0,1)} \boldsymbol{\Phi}_1 \end{aligned}$$

e

$$\boldsymbol{\Delta_{f_\phi}} = \boldsymbol{N_\phi^\top \Psi^{(1,1)} \Phi_1 U} + \boldsymbol{N_\phi^\top \Phi_1 T_w'} + \boldsymbol{N_\phi^\top \Psi^{(0,1)} \Phi_1}.$$

Uma outra alternativa na ponderação de casos é incluir os pesos também nas penalizações, i.e., o logaritmo da função de verossimilhança penalizada apresenta a seguinte forma:

$$\begin{aligned} l_p(\boldsymbol{\theta}) = & \sum_{i=1}^{n} w_i \left\{ (y_i - \mu_i)\Psi^{(1,0)}(\mu_i, \phi_i) + \phi_i T(y_i) + \Psi(\mu_i, \phi_i) + \sum_{i}^{n} \log A(y_i) \right. \\ & \left. -\frac{1}{2}\lambda_\mu^* \boldsymbol{f}_\mu^\top \boldsymbol{K}_\mu \boldsymbol{f}_\mu - \frac{1}{2}\lambda_\phi^* \boldsymbol{f}_\phi^\top \boldsymbol{K}_\phi \boldsymbol{f}_\phi \right\}. \end{aligned}$$

Desse modo, as expressões de $\boldsymbol{\Delta_\beta}$ e $\boldsymbol{\Delta_\gamma}$ não sofrem alterações, porém $\boldsymbol{\Delta_{f_\mu}}$ e $\boldsymbol{\Delta_{f_\phi}}$ tem uma leve mudança em suas expressões, dadas por:

$$\boldsymbol{\Delta_{f_\mu}} = \tilde{X}^\top \boldsymbol{\Psi^{(2,0)} M_1 U} - \lambda_\mu \boldsymbol{K_\mu f_\mu}$$

e

$$\boldsymbol{\Delta_{f_\phi}} = \boldsymbol{N_\phi^\top \Psi^{(1,1)} \Phi_1 U} + \boldsymbol{N_\phi^\top \Phi_1 T_w'} + \boldsymbol{N_\phi^\top \Psi^{(0,1)} \Phi_1} - \lambda_\mu \boldsymbol{K_\mu f_\mu}.$$

É válido ressaltar que, esse esquema de perturbação generaliza o caso em que simplesmente é excluída a i-ésima observação, fazendo $w_i = 0$ e os demais pesos iguais a um. Em outras palavras, a ponderação de casos pode ser vista também como uma medida de influência global sem a necessidade de obter os estimadores dos parâmetros quando excluímos uma observação, como é o caso da distância de Cook, definida a princípio para modelos lineares clássicos em Cook e Weisberg (1982).

Visto os esquemas de perturbações e suas respectivas matrizes $\boldsymbol{\Delta}$, torna-se possível usar as expressões (3.2) e (3.3) para verificar a presença ou não de observações influentes em cada um desses casos. Logo, uma sugestão é analisar a diagonal principal da curvatura $\boldsymbol{C}_l$ associada a um vetor $\boldsymbol{l}_i$ de zeros com um na i-ésima observação, por meio do gráfico de $\boldsymbol{C}_l$ versus seus índices. Assim, considerando C_{ii} o i-ésimo valor da diagonal principal de $\boldsymbol{C}_l$, então destacam-se como possíveis pontos influentes aquelas observações que $C_{ii} > 2\bar{C}$, com $\bar{C}$ sendo o valor médio das curvaturas.

CAPÍTULO 4

Aplicação

Neste capítulo, ilustraremos a utilidade das técnicas de diagnóstico propostas nesta dissertação à classe dos MNLPGS atráves de uma aplicação prática envolvendo dados reais, o qual foi analisado e discutido em Zeviani et al. (2013) propondo uma modelagem não linear da média e da dispersão. Uma motivação a respeito do experimento em questão e uma análise descritiva do mesmo será apresentada. Um modelo específico será ajustado e serão avaliadas as suas suposições e qualidade desse ajuste. A implementação computacional foi feita através do *software* livre R, desenvolvido por R Core Team (2016). No ajuste do modelo consideramos o auxílio das funções disponibilizadas nos pacotes `gamlss` e `gamlss.nl`, propostos por Akantziliotou et al. (2002) e sucessivamente aprimorados em Stasinopoulos et al. (2007), direcionadas para os ajustes de modelos semiparamétricos específicos. As funções para análise de diagnóstico foram todas implementadas no *software* R e posteriormente aplicadas para o ajuste em questão.

4.1 Apresentação dos dados

Os dados utilizados nesta aplicação remetem a aspectos fisiológicos e anatômicos dos frutos de goiabeiras "Pedro Sato", inicialmente apresentados em Cabrini (2009) e posteriomente analisados por Zeviani et al. (2013), supondo uma modelagem não linear para média e controlando a dispersão do modelo através de pesos na estimação pelo métodos dos mínimos quadrados ponderados. Os dados utilizados são referentes a um estudo do Departamento de Fitotecnia da Universidade Federal de Viçosa, e é composto por 7 variáveis referentes a informações de 15 frutos colhidos, em determinadas datas, aleatoriamente de quatro plantas. Esta aplicação trata-se de um exemplo do ajuste de um modelo para dados positivos com a presença de superdispersão.

O objetivo do experimento é avaliar de que forma a massa fresca do fruto, medida

Tabela 4.1: Descrição das variáveis referente aos aspectos fisiológicos e anatômicos dos frutos de goiabeiras "Pedro Sato".

daa	Período após antese (dias);
coleta	Índice da coleta;
rep	Rep: Índice do fruto na coleta;
long	Comprimento longitudinal (em mm);
trans	Comprimento transversal (em mm);
peso	Peso do fruto (em gramas);
volume	Volume do fruto.

através do seu peso, se relaciona com o tempo após acontecer a antese da goiabeira e com suas medidas geométricas especificadas. A utilização dessas informações visa obter resultados para inferir melhor na colheita do fruto, indicando quais são as características que mais influenciam o seu peso. Com essa finalidade, será proposto um modelo de regressão para quantificar essas relações por meio de funções paramétricas e não paramétricas, as quais serão discutidas no decorrer desse capítulo. A variável volume será desconsiderada da análise, pois constatou-se ausência de dados nessa variável e em análises preliminares o volume não apresentou contribuições significativas ao problema proposto.

4.2 Análise descritiva

4.2.1 Análise individual das variáveis

Inicialmente são observados os comportamentos individuais das variáveis que serão utilizadas na modelagem da massa fresca dos frutos de goiaba, feita por meio de uma breve análise descritiva dessas variáveis. A seguir a Tabela 4.2 apresenta os valores descritivos para essas variáveis e na Figura 4.1 encontram-se os boxplots para cada variável sob estudo.

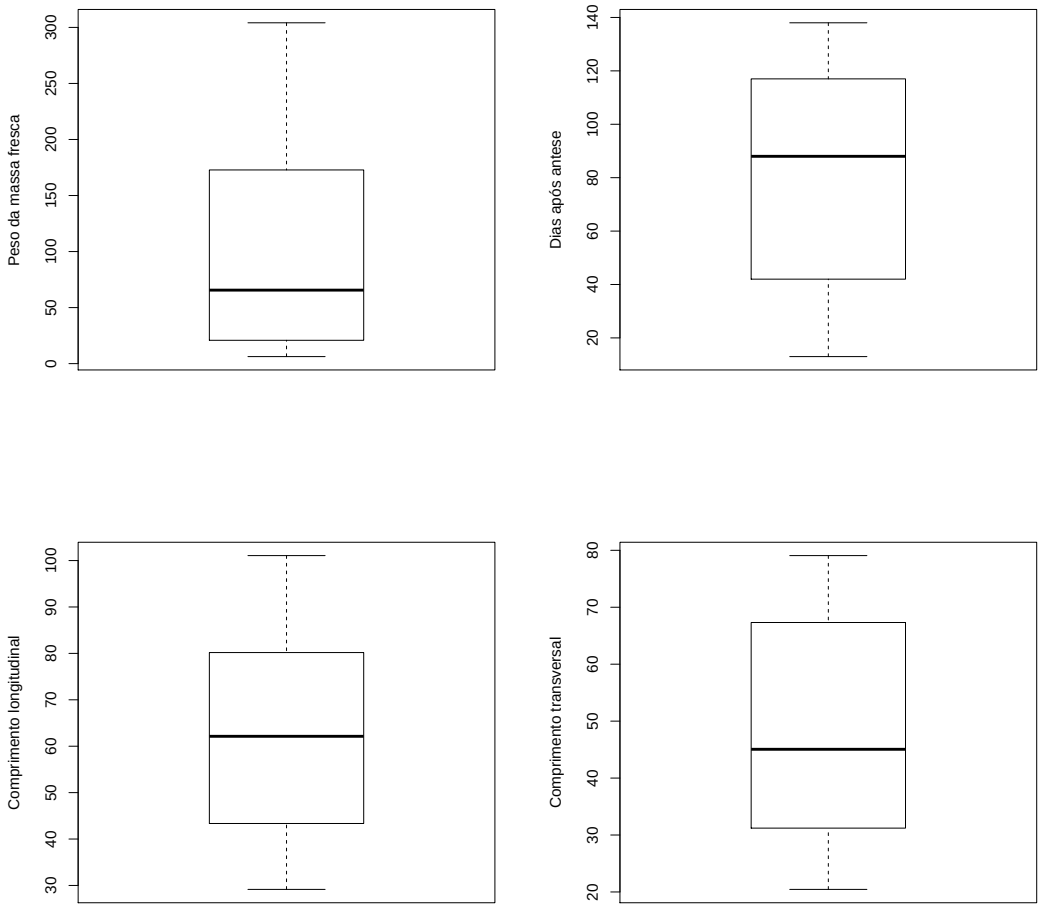
Tabela 4.2: Análise descritiva das variáveis referente dados fisiológicos e anatômicos dos frutos.

Variável	Mínimo	Mediana	Média	Máximo	CV	Assimetria	Curtose
peso	$6,27$	$65,58$	$98,02$	$304,10$	$0,8557$	$0,5143$	$-1,1387$
daa	$13,00$	$88,00$	$83,69$	$138,00$	$0,4771$	$-0,3856$	$-1,2119$
long	$29,15$	$62,14$	$61,71$	$101,10$	$0,3196$	$0,0096$	$-1,4158$
trans	$20,45$	$45,06$	$48,64$	$79,07$	$0,3905$	$0,0457$	$-1,5753$

Na Tabela 4.2, verificam-se medidas descritivas para cada uma das variáveis estudadas, tais como, quantis, média, coeficiente de variação, assimetria e curtose. Podemos observar que o peso do fruto apresenta alta variabilidade estando entre os valores 6,27 e 304,10 gramas. As medidas do fruto foram acompanhadas entre 13 e 138 dias após antese,

apresentando uma variabilidade razoável. As medidas de comprimento longitudinal e transversal têm uma variabilidade menor e possuem uma aparente simetria. É valido ressaltar que, por definição, a massa fresca do fruto trata-se de uma variável contínua estritamente positiva, e nota-se que sua média amostral não está tão próxima da mediana amostral e seu coeficiente de assimetria é positivo, indicando que a variável resposta tem uma possível assimetria à direita.

Figura 4.1: Boxplot das variáveis sob análise, no estudo dos dados fisiológicos e anatômicos dos frutos.

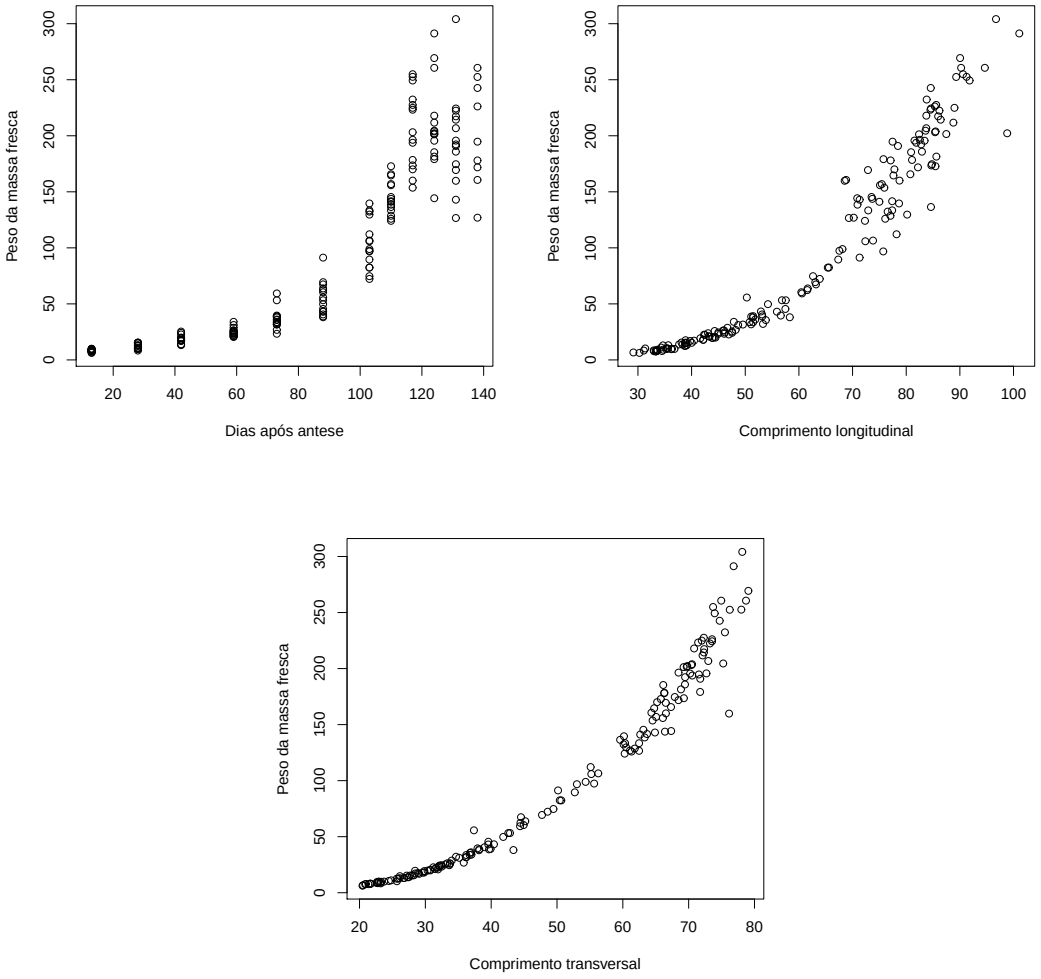


Na Figura 4.1 observa-se novamente uma forma aparentemente assimétrica no peso do fruto e também na distribuição dos dias após antese, no entanto não apresentam nenhum ponto discrepante. Já em relação aos comprimentos, observa-se nos boxplots que há comportamentos simétricos e não se destacam valores discrepantes. Portanto, como foi observada anteriormente, a variável resposta é aparentemente assimétrica, indicando assim que a massa magra da goiaba parece não ser normalmente distribuída, um indício de que o modelo linear clássico não seja adequado para o problema.

4.2.2 Avaliação da relação entre as variáveis

Com o intuito de estabelecer qual a forma funcional do modelo, deve-se avaliar a relação entre as variáveis estudadas por meio de análise gráfica e observando suas correlações. Logo, através dos dados obtidos, são contruídos os gráficos de dispersão para verificar o comportamento do peso dos frutos da goiabeira em função de cada uma das possíveis variáveis explicativas e seus respectivos coeficientes de correlação, conforme podemos observa na Figura 4.2.

Figura 4.2: Peso dos frutos da goiabeiras em termos de cada variável explicativa.



Pela Figura 4.2, podemos verificar que a relação do peso com os dias após antese apresenta uma estrutura não linear. Com isso, trabalhar essa relação por meio de funções de suavização possa ser uma opção mais viável. Além disso, podemos observar que o peso do fruto apresenta variabilidade maior para um maior número de dias após antese, indicando a presença de superdispersão nos dados. No caso das variáveis associadas ao comprimento do fruto apresentam uma clara forma não linear.

Devemos ressaltar que, é necessário avaliar, de forma cuidadosa, o modelo que seja

mais adequado, pois constatamos indícios claros de não linearidade e de dispersão variável. Desse modo, visando um ajuste melhor, considerou-se um modelo não linear parcial generalizado superdispersado, o qual estabelece uma relação não paramétrica com o número de dias após antese e estruturas não lineares aos comprimentos longitudinais e transversais com o peso médio. Para a modelagem da variabilidade do peso é definido apenas uma relação linear (ou linearizável), dependendo apenas da quantidade de dias após antese, visto que a variabilidade não pareça sofrer influência das demais variáveis. Em seguida formalizamos o modelo que irá descrever essa regressão e avaliamos suas suposições e a sua qualidade de ajuste.

4.3 Modelo poisson duplo

Dadas as características dos dados anteriormente apresentadas, naturalmente algumas propostas de modelos são apropriadas. Zeviani et al. (2013) propõe um modelo poisson considerando a média da massa fresca como uma função não linear dos dias após antese, obtido a partir da manipulação do modelo Gompertz. Como foi constatado heterocedasticidade natural do peso em realação aos dias após antese, Zeviani et al. (2013) sugeriu controlar a variabilidade do modelo atribuíndo ponderações no método de estimação por meio de quatro diferentes alternativas de formas funcionais. Dentre as funções para a variância do modelo, a que obteve melhor resultado foi a função potencia do logaritmo ($\sigma^2|\log daa|^{2\delta}$), no entanto houve indicativos que a heterocedasticidade do modelo não foi corrigida de maneira satisfatória. Como um proposta alternativa, iremos considerar uma modelagem simultânea da média e dispersão supondo um componente aleatório associado à distribuição poisson dupla. Por outro lado, uma análise prévia para atribuição do componente sistemático indicou uma possível modelagem através das seguintes formas funcionais para a média e para a dispersão, respectivamente:

$$\begin{aligned} \mu_i &= \beta_0 + \beta_1(\text{long}_i + \beta_2)^{\beta_3} + \beta_4(\text{trans}_i + \beta_5)^{\beta_6} + f(\text{daa}_i), \\ \log(\phi_i) &= \gamma_0 + \gamma_1 \text{daa}_i, \end{aligned} \tag{4.1}$$

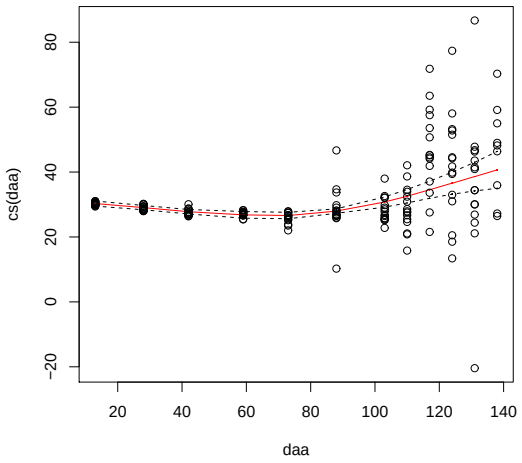
em que $\text{peso}_i \sim \mathcal{PD}(\mu_i, \phi_i)$.

Note que, como as relações entre o peso dos frutos e suas variáveis explicativas são descritas por formas não lineares, optamos então em usar a ligação identidade. No caso da dispersão, optou-se por assumir uma forma linearizável, obtida com o auxilio da ligação logaritmica para a relação entre esse preditor. Com base no modelo proposto, os valores dos métodos de seleção de modelos AIC e BIC são, respectivamente, $1017,41$ e $1058,49$. Logo, formalizando o modelo em sua forma inicial, devemos então proceder com o seu ajuste, interpretação e diagnóstico do modelo, com o intuito de verificar se o mesmo está bem ajustado aos dados referentes ao problema proposto.

4.3.1 Ajuste do modelo

Uma análise inferencial do modelo representa uma etapa de suma importância, levando em conta que é necessário avaliar se os parâmetros do modelo são representativos, indicando assim quais variáveis deverão permanecer na estrutura do modelo. Dessa forma, utilizando o modelo proposto em (4.1), verificamos quais são as inferências obtidas para os parâmetros do modelo e de sua função não paramétrica. No que diz respeito à função não paramétrica, pode-se verificar seus valores ajustados e seus erros padrões na Tabela 4.3 e o seu comportamento em função dos dias após antese é ilustrado na Figura 4.3. Além disso, na Tabela 4.3 observa-se também os valores ajustados dos parâmetros dos modelos, bem como seus respectivos erros padrões.

Figura 4.3: Gráfico da função não paramétrica dos dias após antese estimda sob o modelo Poisson duplo ajustado aos dados fisiológicos e anatômicos dos frutos.



Na figura 4.3 podemos verificar a relação não paramétrica estimada entre o preditor da média e os dias após a antese, bem como o seu intervalo de confiança para cada estimativa. É possível observar uma têndencia crescente da variabilidade dos dados com o passar dos dias após antese e que a curva estimada é uma representação razoável dessa relação.

Note que, pela Tabela 4.3 os erros padrão das estimativas do modelo são todos razoavelmente pequenos, indicando que a estimativa dos parâmetros apresentam pouca variabilidade, e consequentemente, o modelo está aparentemente bem ajustado. Observe também que, em relação a curva ajustada para a relação não paramétrica, seus erros padrão apresenta um comportamento crescente em função dos dias após antese, assim como foi observado na Figura 4.3. Dessa forma, de acordo com os valores apresentados na Tabela 4.3, verifica-se que o ajuste do MNLPGS com distribuição poisson dupla tem a seguinte forma funcional:

$$\begin{array}{rcl} \hat{\mu}_i & = & -37,22 + 0,000076(\text{long}_i + 10)^{2,83} + 0,002589(\text{trans}_i + 5)^{2,56} + \hat{f}(\text{daa}_i), \\ \log(\hat{\phi}_i) & = & -4,18 + 0,036\text{daa}_i. \end{array} \tag{4.2}$$

Tabela 4.3: Análise inferencial do modelo poisson duplo ajustado aos dados fisiológicos e anatômicos dos frutos.

Estimador	Estimativa	Erro Padrão	Estimador	Estimativa	Erro Padrão
$\hat{\beta}_0$	$-37,2200$	$0,1687$	$f(42)$	$-111,7225$	$0,3494$
$\hat{\beta}_1$	$0,0001$	$< 0,0001$	$f(59)$	$-112,6899$	$0,5224$
$\hat{\beta}_2$	10	$< 0,0001$	$f(73)$	$-112,8983$	$0,5058$
$\hat{\beta}_3$	$2,8300$	$< 0,0001$	$f(88)$	$-111,5210$	$0,3513$
$\hat{\beta}_4$	$0,0026$	$< 0,0001$	$f(100)$	$-106,9481$	$1,0547$
$\hat{\beta}_5$	5	$< 0,0001$	$f(103)$	$-108,7246$	$0,8045$
$\hat{\beta}_6$	$2,5600$	$< 0,0001$	$f(117)$	$-104,9637$	$1,3466$
$\hat{\gamma}_0$	$4,1838$	$0,2377$	$f(124)$	$-102,9254$	$1,7338$
$\hat{\gamma}_1$	$-0,0358$	$0,0026$	$f(131)$	$-100,8807$	$2,2487$
$f(13)$	$-109,2018$	$0,3992$	$f(138)$	$-98,8240$	$2,8983$
$f(28)$	$-110,4616$	$0,3381$			

Sendo assim, as equações definidas em (4.2) expressam o modelo na sua forma ajustada aos dados, o qual está especificando de que forma a média e a dispersão do peso das frutas estão relacionadas com as suas correspondentes variáveis explicativas. Além disso, observa-se que o modelo ajustado fornerceu um valor de $1017,42$ e $1058,49$, para os critérios de seleção de modelos AICp e BICp, respectivamente.

4.3.2 Diagnóstico do modelo

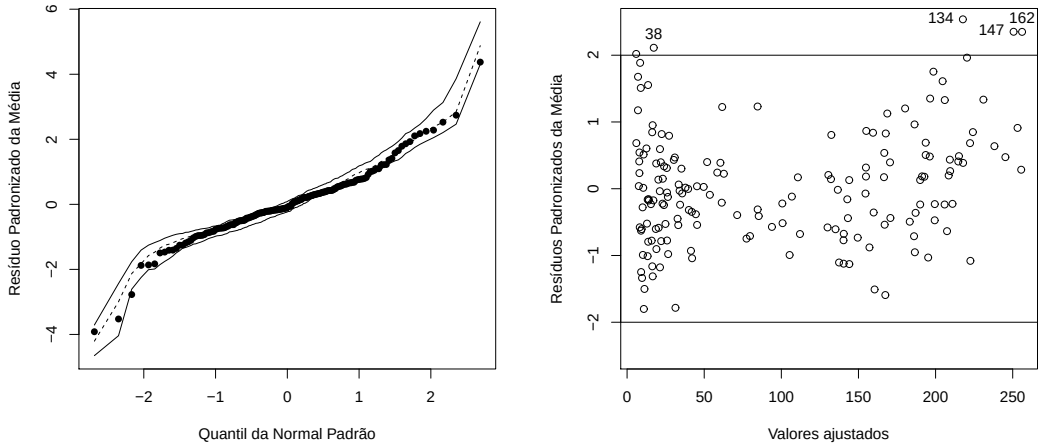
Como já foi mencionado, devemos verificar agora se o modelo adotado está realmente bem ajustado e representativo aos dados. Para isso, uma forma de avaliar o quão bom está o ajuste é através de uma análise das suposições do modelo e sua sensibilidade em relação às observações aberrantes e/ou influentes. Em seguida, caso o diagnóstico do modelo seja capaz de detectar observações que apresentam fortes influências no seu ajuste, então uma forma de confirmar a sensibilidade do modelo a essas observações é verificar as inferências do modelo com a retirada desses pontos. Se houver mudanças inferenciais significativas ao considerar essa retirada, então devemos ter cuidado nas conclusões do modelo. A seguir é abordada toda essa metodologia na validação do modelo definido.

Análise das suposições

Para investigar se o modelo foi bem ajustado, devemos fazer uma análise de resíduos e verificar se as suposições acerca do modelo estão sendo satisfeitas. Para verificar as suposições do modelo ajustado, podemos avaliar o comportamento dos gráficos de quantis-quantis dos resíduos padronizados, com o intuito de detectar possíveis pontos aberrantes e tendências não aleatórias. Na Figura 4.4 temos os gráficos dos resíduos para o modelo proposto.

Note que, pelo gráfico de resíduos padronizados para média versus valor ajustado, as

Figura 4.4: Gráficos para análise de resíduos referente ao modelo Poisson duplo ajustado aos dados fisiológicos e anatômicos dos frutos.



observações estão dispostas aleatoriamente com variação constante, indicando evidências de que as suposições da componente aleatória estão satisfeitas. No respectivo gráfico de quantis-quantis, observa-se um comportamento aleatório e apenas uma pequena quantidade de pontos no limiar das bandas de 95% de confiança. Além disso, destacam-se apenas as pontos #38, #134, #147 e #162 como possíveis valores aberrantes, no entanto apresentam uma leve discrepância.

Análise de Sensibilidade

A análise de sensibilidade visa avaliar o comportamento do ajuste de um modelo quando ele está sujeito a algum tipo de perturbação, isto é, verificar se há observações em que interferem significativamente na estimativa do modelo, implicando possíveis conclusões errôneas. Como uma forma inicial de avaliar a sensibilidade do modelo, deve-se verificar a presença de pontos de alavanca, os quais tendem a mudar as estimativas do modelo. Na Figura 4.5 são apresentados os gráficos da diagonal principal da matriz *hat* e alavancagem generalizada para a média.

Nesse caso, de acordo com as medidas de alavancagem, destacam-se as observações #147, #148 e #160 como candidatos a pontos de alavanca, os quais merecem uma atenção maior. Por outro lado, devemos considerar a análise de influência local sob os esquemas de perturbações anteriormente mencionado, a saber: perturbação aditiva na variável resposta, perturbarção aditiva nos preditores e ponderação de casos. Nas Figuras 4.6, 4.7 e 4.8 são apresentadas as curvaturas para o ajuste do modelo proposto.

Note que há indícios de que a observação #147 e #162 são pontos que causam forte influência ao modelo e, consequentemente, podem acarretar mudanças inferenciais significativas, comprometendo assim o ajuste do modelo. De modo geral, os pontos que podem estar comprometendo o ajuste do modelo são #147, #160 e #162, pois são pontos que teoricamente estão contradizendo o que foi proposto pelo modelo em questão.

Figura 4.5: Gráficos de alavancagem baseada na matriz *hat* e alavancagem generalizada para a média sob o modelo Poisson duplo referente aos dados fisiológicos e anatômicos dos frutos.

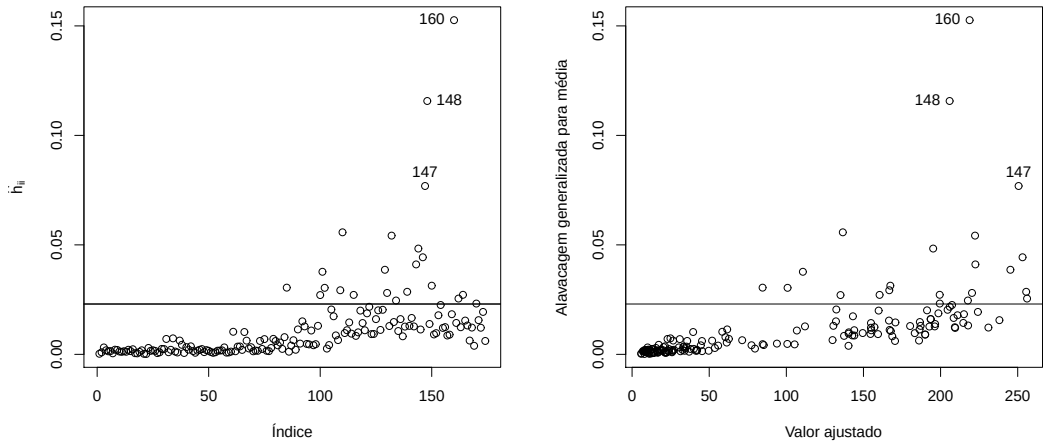
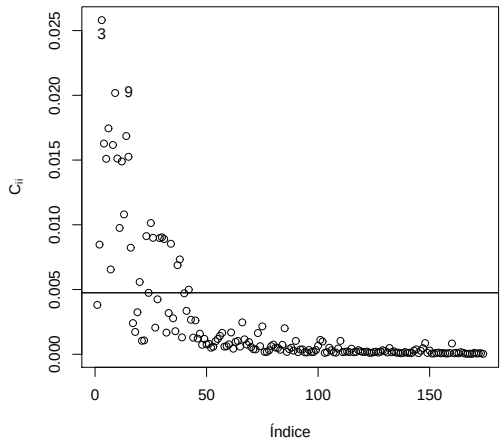


Figura 4.6: Gráficos de C_i com perturbação aditiva na resposta por seu índice sob o modelo Poisson duplo referente aos dados fisiológicos e anatômicos dos frutos.



Análise Confirmatória

Como já foi citado, após destacadas as possíveis observações influentes no ajuste do modelo proposto devemos verificar se o mesmo sofre alguma mudança inferencial ao retirarmos essas observações. Inicialmente estudamos o motivo pelo qual as observações #147, #160 e #162 se destacaram das demais, tornando o modelo possivelmente sensível, como temos a seguir:

#147: Apresenta o segundo maior peso (291,3g) e mesmo com altos valores para os comprimentos longitudinais (101,07mm) e transversais (76,83mm) e um número razoável de dias após antese (124 dias), este representa um possível valor discrepante na resposta.

Figura 4.7: Gráficos de C_i com perturbação aditiva no predito para comprimento longitudinal (a) e transversal (b) por seu índice sob o modelo Poisson duplo referente aos dados fisiológicos e anatômicos dos frutos.

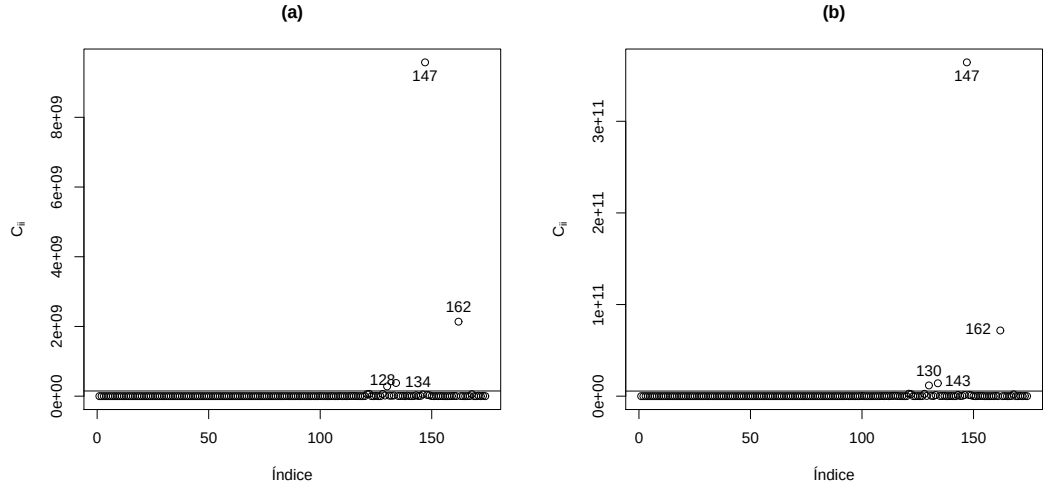
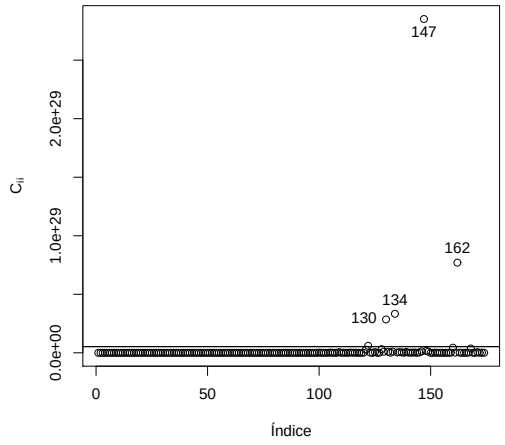


Figura 4.8: Gráficos de C_i com perturbação de casos ponderados por seu índice sob o modelo Poisson duplo referente aos dados fisiológicos e anatômicos dos frutos.



#**160:** Observa-se um peso (159,82g) do fruto menor que mais de 25% dos demais mesmo apresentando um alto número de dias após antese (131 dias), um dos maiores comprimentos transversais (76,82mm) e um comprimento longitudinal razoável (68,53mm).

#**162:** Essa observação representa o fruto com maior peso (304,11g), bem distantes dos demais, representando um valor discrepante na variável resposta. Vale observar que, embora esse tenha comprimento longitudinal (96,73mm) e transversal (78,14mm) grandes e um número significativo de dias após antese (131 dias), ainda representa um valor acima do que o ajuste calcula.

Dessa, a seguir é apresentada uma tabela com a análise confirmatória da influência no modelo, obtida reajustando o modelo sem as principais observações destacadas como

influentes e comparando com o modelo completo, verificando então o quanto esse valor influência na inferência do modelo. Na Tabela 4.4 é verifica-se a alteração nas estimativas dos parâmetros ao retirar as observações aparentemente mais influentes (#147, #160 e #162) em relação ao modelo completo (sem a retirada das observações). Note que, a tabela apresenta os valores de cada estimativa dos parâmetros e seus respectivos erros padrões, observando-se, entre parênteses, o impacto percentual na estimativa do parâmetro ao retirar a respectiva observação.

Tabela 4.4: Estimativas dos parâmetros ao retirar as observações influentes sob o modelo Poisson duplo referente aos dados fisiológicos e anatômicos dos frutos.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Modelo Completo	$-37,2200$	$0,0001$	$10,0000$	$2,8300$	$0,0026$
Excluindo #283	$-37,5288$	$0,0001$	$10,0000$	$2,8300$	$0,1376$
	(-0,83%)	(0,00%)	(0,00%)	(0,00%)	(0,00%)
Excluindo #372	$-35,6429$	$0,0001$	$10,0000$	$2,8300$	$0,1343$
	(4,24%)	(0,00%)	(0,00%)	(0,00%)	(0,00%)
Excluindo #283 e #372	$34,4609$	$0,0001$	$10,0000$	$2,8300$	$0,1280$
	(7,41%)	(0,00%)	(0,00%)	(0,00%)	(0,00%)
	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	
Modelo Completo	$5,000$	$2,5600$	$4,1839$	$-0,0359$	
Excluindo #283	$5,0000$	$2,5600$	$4,1903$	$-0,0358$	
	(0,00%)	(0,00%)	(-0,15%)	(0,28%)	
Excluindo #372	$5,0000$	$2,5600$	$3,9681$	$-0,0328$	
	(0,00%)	(0,00%)	(5,16%)	(8,64%)	
Excluindo #283 e #372	$5,0000$	$2,5600$	$4,1663$	$-0,0354$	
	(0,00%)	(0,00%)	(0,42%)	(1,39%)	

Como podemos observar na Figura 4.4, os ajustes ao excluir as observações apresentam mudanças substanciais entre as estimativas de alguns parâmetros, o que pode influenciar na predição do modelo. Por outro lado, notou-se que não houve mudanças substanciais nas inferências do modelo, em relação aos testes de significância dos parâmetros, tanto excluindo as observações individualmente como as excluindo simultaneamente. Dessa forma, concluímos que essas observações não representam fortes influências no modelo em questão, e que caso não se tenha interesse em predição de valores, esse ajuste é satisfatório. Dessa forma, embora algumas observações apresentem um pequeno impacto nas estimativas, em geral, o ajuste proposto se mostrou adequado aos dados.

CAPÍTULO 5

Considerações Finais

5.1 Discussão e contribuições

Na presente dissertação foi discutida uma nova classe de modelos intitulada de modelos não lineares parciais generalizados superdispersados, formalizando sua estrutura funcional, sua componente aleatória e o processo de estimação associado à função de verossimilhança penalizada. Essa classe foi obtida ao incluir funções não paramétricas e estruturas paramétricas não lineares à classe de modelos proposta por Dey et al. (1997). Com isso, esses modelos se mostram uma alternativa bastante atrativa para casos que a variabilidade não é constante, especialmente quando há presença de relações não lineares entre as variáveis estudadas. O acréscimo de uma função não paramétrica nos preditores possibilitam uma forma de lidar com alguma variável explicativa que possui uma relação funcional até mesmo desconhecida.

Como principais contribuições teóricas podemos destacar a proposta de um modelo bastante flexível, generalizando casos mais particulares de modelos com superdispersão ou sobredispersão. O estudo de diagnóstico dessa classe de modelos presente nessa dissertação representa uma importante contribuição, uma vez que essa é uma etapa fundamental para avaliar as suposições do modelo e sua sensibilidade à observações discrepantes ou influentes. Especificamente, foram obtidas as medidas de alavancagem e de influência local, sob alguns esquemas de perturbação.

No que diz respeito ao contexto prático, foi discutido um exemplo de aplicação com dados reais referente ao estudo dos aspectos fisiológicos e anatômicos dos frutos de determinada goiabeira. O modelo ajustado se mostrou adequado e o uso de um modelo semiparamétrico com superdispersão é uma alternativa viável para a discussão desse problema prático. A análise gráfica do modelo utilizando as medidas de alavancagem e influência local apresentou-se eficiente na detecção de observações influentes, ou seja, que causam

influências desproporcionais nas estimativas do modelo.

5.2 Implementação computacional

Como já foi mencionado anteriormente, o processo de estimação apresentado nesta dissertação para toda a aplicação prática foi realizado com o auxílio de pacotes presentes no *software* livre R, os quais possibilitaram o ajuste de um modelo específico presente na classe de distribuições abordada nesta dissertação, extraindo-se então os valores ajustados, os parâmetros estimados e seus respectivos erros padrões. No entanto, no que diz respeito a análise de diagnóstico, a princípio foi implementada uma função para obtenção da matriz de informação de Fisher observada. Além disso, foi considerada a implementação das funções referentes as medidas de alavancagem e análise de resíduos. Na análise de influência local, elaboramos três esquemas de perturbações: perturbação na variável resposta, perturbação nos preditores paramétricos e perturbação de casos ponderados. Ainda com o auxilio computacional presente no *software* R, foi realizada uma análise gráfica baseada nas medidas obtidas via técnicas de diagnóstico e então utilizados na identificação da sensibilidade do modelo.

5.3 Perspectivas de trabalhos futuros

Uma primeira perspectiva de trabalhos futuros está relacionada à construção de modelos aditivos a partir da classe de modelos superdispersados apresentado nesta dissertação. A classe de modelos apresentada aqui representa apenas o uso parcial de funções não paramétricas, incluindo essa forma funcional relacionada apenas a uma variável explicativa. Naturalmente, seria possível generalizar esse caso ao fazer uso de mais de uma função não paramétrica na estrutura definida nos preditores não lineares no modelo proposto. Essa abordagem tornaria o modelo ainda mais flexível, pois ocasionalmente há uma necessidade de estabelecer relações não paramétricas em mais de uma variável explicativa.

Outra perspectiva interessante seria em considerar uma análise inferencial bem mais ampla a esta classe de modelos, como por exemplo o desenvolvimento de testes de hipóteses e intervalos de confiança associados principalmente à parcela não paramétrica. O estudo inferencial do modelo representa uma grande importância na sua validação e na qualidade do ajuste, sendo então fundamental avaliar a significância das funções não paramétricas. Nesse ponto de vista, um fator fundamental seria testar a linearidade das funções não paramétricas, verificando assim se o uso desses processos computacionais apresentam um ganho real na estimativa do modelo.

Por fim, uma perspectiva viável é em relação à inclusão dos parâmetros de suavização à análise de diagnóstico referente ao modelo estudado, uma vez que o ajuste do modelo apresentado nesta dissertação leva em consideração que os valores desses parâmetros são fixos e obtidos separadamente do processo de estimação dos demais parâmetros do modelo.

Essa abordagem se mostra especialmente importante, visto que o modelo poderia apresentar algum tipo específico de sensibilidade ao introduzir perturbações em seus parâmetros de suavização. Dessa forma, como esses parâmetros afetam diretamente na suavidade da curva estimada, seria interessante pensar em algum esquema de perturbação que leve em consideração a sensibilidade do modelo aos parâmetros de suavização.

Referências

AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974.

AKANTZILIOTOU, C.; RIGBY, R.; STASINOPOULOS, D. The R implementation of generalized additive models for location, scale and shape. In: STATISTICAL MODELLING SOCIETY. *Statistical modelling in Society: Proceedings of the 17th International Workshop on statistical modelling*. Chania, 2002. p. 75–83.

BUJA, A.; HASTIE, T.; TIBSHIRANI, R. Linear smoothers and additive models. *The Annals of Statistics*, JSTOR, p. 453–510, 1989.

CABRINI, E. C. *Aspectos fisiológicos e anatômicos de goiaba "Pedro Sato" em desenvolvimento*. Tese (Doutorado) — Universidade Federal de Viçosa, 2009.

CONCEIÇÃO, G. M. d. S.; SALDIVA, P. H. N.; SINGER, J. d. M. Glm and gam model for analyzing the association between atmospheric pollution and morbidity-mortality markers: an introduction based on data from the city of são paulo. *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 4, n. 3, p. 206–219, 2001.

COOK, R. D. Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 133–169, 1986.

COOK, R. D.; WEISBERG, S. *Residuals and influence in regression*. New York: Chapman and Hall, 1982.

CORDEIRO, G. M.; DEMETRIO, C. *Modelos lineares generalizados*. Campinas: Universidade Estadual de Campinas. Dep. de estatistica Campinas, 1986.

DEY, D. K.; GELFAND, A. E.; PENG, F. Overdispersed generalized linear models. *Journal of Statistical Planning and Inference*, Elsevier, v. 64, n. 1, p. 93–107, 1997.

EFRON, B. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, Taylor & Francis, v. 81, n. 395, p. 709–721, 1986.

GELFAND, A.; DALAL, S. A note on overdispersed exponential families. *Biometrika*, Biometrika Trust, v. 77, n. 1, p. 55–64, 1990.

GIJBELS, I.; PROSDOCIMI, I.; CLAESKENS, G. Nonparametric estimation of mean and dispersion functions in extended generalized linear models. *Test*, Springer, v. 19, n. 3, p. 580–608, 2010.

GOOD, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika*, JSTOR, p. 237–264, 1953.

GREEN, P. J.; SILVERMAN, B. W. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Boca Raton: CRC Press, 1993.

HANDSCOMB, D. Spline functions. *Methods of Numerical Approximation. Oxford: Pergamon Press*, 1966.

HASTIE, T.; TIBSHIRANI, R. Generalized additive models: some applications. *Journal of the American Statistical Association*, Taylor & Francis, v. 82, n. 398, p. 371–386, 1987.

HASTIE, T. J.; TIBSHIRANI, R. J. Generalized additive models, volume 43 of monographs on statistics and applied probability. *Chapman & Hall*, Chapman & Hall, 1990.

HECKMAN, N. E. Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 244–248, 1986.

JOHNSON, N.; KOTZ, S.; BALAKRISHNAN, N. *Continuous Univariate Discrete Distributions, vol. 1*. [S.l.]: Wiley, New York, 1994.

MANGHI, R. F. *Técnicas de diagnóstico em modelos parcialmente lineares aditivos generalizados para dados correlacionados*. Tese (Doutorado) — Universidade Federal de Pernambuco, 2016.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society*, Série A, v. 135, p. 370-384, 1972.

NODA, G. R. *Análise de diagnóstico em modelos semiparamétricos normais*. Dissertação (Mestrado) — Universidade de São Paulo, 2013.

PAULA, G. A. *Modelos de regressão: com apoio computacional*. São Paulo: IME-USP São Paulo, 2004.

PREVIDELLI, I. T. S. *Estimadores corrigidos para modelos não-lineares generalizados superdispersados*. Tese (Doutorado) — Universidade Federal de Santa Catarina, 2005.

PULGAR, G. M. I. *Modelos mistos aditivos semiparamétricos de contornos elıpticos*. Tese (Doutorado) — Universidade de Sao Paulo, 2009.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <https://www.R-project.org/>.

REINSCH, C. H. Smoothing by spline functions. *Numerische mathematik*, Springer, v. 10, n. 3, p. 177–183, 1967.

RELVAS, C. E. M. *Modelos parcialmente lineares com erros simétricos autoregressivos de primeira ordem*. Dissertação (Mestrado) — Universidade de São Paulo, 2013.

SCHWARZ, G. et al. Estimating the dimension of a model. *The annals of statistics*, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

SILVERMAN, B. W. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–52, 1985.

SIMONOFF, J. Smoothing methods in statistics. *Springer Series in Statistics*, Springer Series in Statistics, 1996.

SMYTH, G. K. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 47–60, 1989.

STASINOPOULOS, D. M.; RIGBY, R. A. et al. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, v. 23, n. 7, p. 1–46, 2007.

TERRA, M. L. C. *Modelos Não Lineares Generalizados com Superdispersão*. Tese (Doutorado) — Universidade Federal de Pernambuco, 2013.

TURKMAN, M. A. A.; SILVA, G. L. Modelos lineares generalizados-da teoria à prática. In: *VIII Congresso Anual da Sociedade Portuguesa de Estatística*. Lisboa: Lisboa, 2000.

WAHBA, G. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 364–372, 1978.

WAHBA, G. *Spline models for observational data*. Philadelphia: Siam, 1990. v. 59.

WEI, B.-C.; HU, Y.-Q.; FUNG, W.-K. Generalized leverage and its applications. *Scandinavian Journal of statistics*, Wiley Online Library, v. 25, n. 1, p. 25–37, 1998.

WOOD, S. *Generalized additive models: an introduction with R*. Boca Raton: CRC press, 2006.

WU, H.; ZHANG, J.-T. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. Hoboken: John Wiley & Sons, 2006. v. 515.

ZEVIANI, W. M.; JR, P. J. R.; BONAT, W. H. *Modelos de regressão não linear*. 58ž. ed. Universidade Federal do Paraná, 2013.

APÊNDICE A

Cálculo do algoritmo escore de Fisher

Como bem sabemos, o algoritmo escore de Fisher para a obtenção dos estimadores de máxima verossimilhança pode ser expresso, de forma geral, por meio da equação dada por:

$$\mathrm{I}_p(\boldsymbol{\theta})(\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{(m)}) = U_p(\boldsymbol{\theta}).$$

Nesse caso, sabendo-se da ortogonalidade entre os parâmetros da média e da dispersão, torna-se possível particionar o processo de estimação através de dois algoritmo escore de Fisher distintos, os quais serão especificados logo em seguida.

Obtenção do processo escore para a média

Considere o processo de estimação direcionado para os parâmetros da média, ou seja, $\boldsymbol{\beta}$ e $\boldsymbol{f_\mu}$. Dessa forma, o seu algoritmo de estimação escore de Fisher é obtido resolvendo a seguinte equação matricial:

$$\begin{pmatrix} \mathrm{I}_p^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{\beta f_\mu}}(\boldsymbol{\theta}) \\ \mathrm{I}_p^{\boldsymbol{f_\mu \beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{f_\mu f_\mu}}(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)} \\ \boldsymbol{f}_\mu^{(m+1)} - \boldsymbol{f}_\mu^{(m)} \end{pmatrix} = \begin{pmatrix} U_p(\boldsymbol{\beta}) \\ U_p(\boldsymbol{f}_\mu) \end{pmatrix}.$$

Substituindo os valores do vetor de funções escores e da matriz de informação de Fisher

esperada, obtemos a seguinte estrutura:

$$\begin{pmatrix} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} & \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} \\ \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} & \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)} \\ \boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)} - \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}) \\ \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}) - \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}} \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} \end{pmatrix}.$$

Resolvendo o sistema de equações designado ao vetor de parâmetros $\boldsymbol{\beta}$ encontramos a igualdade dada por:

$$\begin{aligned} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} \boldsymbol{\beta}^{(m+1)} &= \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} \boldsymbol{\beta}^{(m)} + \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} (\boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)} - \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)}) \\ &\quad + \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}). \end{aligned}$$

Assim, o estimador de $\boldsymbol{\beta}$ para a $m+1$-ésimo iteração é expresso pela seguinte forma:

$$\begin{aligned} \boldsymbol{\beta}^{(m+1)} &= (\tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 [\tilde{\boldsymbol{X}} \boldsymbol{\beta}^{(m)} + \boldsymbol{N}_{\boldsymbol{\mu}} \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} - \boldsymbol{N}_{\boldsymbol{\mu}} \boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)} \\ &\quad + \boldsymbol{M}_1^{-1} (\boldsymbol{y} - \boldsymbol{\mu})]. \end{aligned}$$

De forma equivalente, resolvendo o sistema de equações obtido para o vetor de parâmetros $\boldsymbol{f}_{\boldsymbol{\mu}}$, temos que:

$$\begin{aligned} (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}}) \boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)} &= \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} - \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \tilde{\boldsymbol{X}} (\boldsymbol{\beta}^{(m+1)} + \boldsymbol{\beta}^{(m)}) \\ &\quad + \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1 (\boldsymbol{y} - \boldsymbol{\mu}) \end{aligned}$$

Portanto, o estimador de $\boldsymbol{f}_{\boldsymbol{\mu}}$ para a $(m+1)$-ésima iteração é definido da seguinte forma:

$$\begin{aligned} \boldsymbol{f}_{\boldsymbol{\mu}}^{(m+1)} &= (\boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 \boldsymbol{N}_{\boldsymbol{\mu}} + \lambda_{\boldsymbol{\mu}} \boldsymbol{K}_{\boldsymbol{\mu}})^{-1} \boldsymbol{N}_{\boldsymbol{\mu}}^\top \boldsymbol{\Psi}^{(2,0)} \boldsymbol{M}_1^2 [\tilde{\boldsymbol{X}} \boldsymbol{\beta}^{(m)} + \boldsymbol{N}_{\boldsymbol{\mu}} \boldsymbol{f}_{\boldsymbol{\mu}}^{(m)} - \tilde{\boldsymbol{X}} \boldsymbol{\beta}^{(m+1)} \\ &\quad + \boldsymbol{M}_1^{-1} (\boldsymbol{y} - \boldsymbol{\mu})]. \end{aligned}$$

Obtenção do processo escore para a dispersão

Considere agora o processo de estimação direcionado para os parâmetros da dispersão, isto é, $\boldsymbol{\gamma}$ e $\boldsymbol{f}_{\boldsymbol{\phi}}$. Logo, o seu respectivo algoritmo de estimação escore de Fisher é obtido resolvendo a seguinte expressão:

$$\begin{pmatrix} \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{\gamma}\boldsymbol{f}_\phi}(\boldsymbol{\gamma}) \\ \mathrm{I}_p^{\boldsymbol{f}_\phi\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathrm{I}_p^{\boldsymbol{f}_\phi\boldsymbol{f}_\phi}(\boldsymbol{\theta}) \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\gamma}^{(m+1)} - \boldsymbol{\gamma}^{(m)} \\ \boldsymbol{f}_\phi^{(m+1)} - \boldsymbol{f}_\phi^{(m)} \end{pmatrix} = \begin{pmatrix} U_p(\boldsymbol{\beta}) \\ U_p(\boldsymbol{f}_\phi) \end{pmatrix}.$$

Substituindo os valores do vetor de funções escores e da matriz de informação de Fisher

esperada, obtemos a seguinte estrutura matrical:

$$\begin{pmatrix} -\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}} & -\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}} \\ -\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}} & -\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}+\lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}} \end{pmatrix}\begin{pmatrix} \boldsymbol{\gamma}^{(m+1)}-\boldsymbol{\gamma}^{(m)} \\ \boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)}-\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)} \end{pmatrix}$$
$$=\begin{pmatrix} \tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(1,1)}\boldsymbol{\Phi}_1(\boldsymbol{y}-\boldsymbol{\mu})+\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Phi}_1 T(\boldsymbol{y})+\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,1)}\boldsymbol{\Phi}_1 \\ \boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(1,1)}\boldsymbol{\Phi}_1(\boldsymbol{y}-\boldsymbol{\mu})+\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Phi}_1 T(\boldsymbol{y})+\boldsymbol{N}^{\top}\boldsymbol{\Psi}^{(0,1)}\boldsymbol{\Phi}_1-\lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}}\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)} \end{pmatrix}.$$

Resolvendo o sistema de equações designado ao vetor de parâmetros $\boldsymbol{\gamma}$ obtemos a igualdade dada por:

$$\begin{aligned} -\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}}\boldsymbol{\gamma}^{(m+1)} &= -\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}}\boldsymbol{\gamma}^{(m)}+\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}(\boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)}-\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)}) \\ &\quad +\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(1,1)}\boldsymbol{\Phi}_1(\boldsymbol{y}-\boldsymbol{\mu})+\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Phi}_1 T(\boldsymbol{y})+\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,1)}\boldsymbol{\Phi}_1. \end{aligned}$$

Dessa forma, o estimador de $\boldsymbol{\gamma}$ para a $(m+1)$-ésimo iteração é expresso pela seguinte forma:

$$\begin{aligned} \boldsymbol{\gamma}^{(m+1)} &= (\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}})^{-1}\tilde{\boldsymbol{S}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2(\tilde{\boldsymbol{S}}\boldsymbol{\gamma}^{(m)}+\boldsymbol{N}_{\boldsymbol{\phi}}\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)}-\boldsymbol{N}_{\boldsymbol{\phi}}\boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} \\ &\quad +(-\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}[\boldsymbol{\Psi}^{(1,1)}(\boldsymbol{y}-\boldsymbol{\mu})+T(\boldsymbol{y})+\boldsymbol{\Psi}^{(0,1)}]). \end{aligned}$$

Por outro lado, resolvendo o sistema de equações obtido para o vetor de parâmetros $\boldsymbol{f}_{\boldsymbol{\phi}}$, temos que:

$$\begin{aligned} -(\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}+\lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}})\boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} &= -\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)}+\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\tilde{\boldsymbol{S}}(\boldsymbol{\gamma}^{(m+1)}-\boldsymbol{\gamma}^{(m)}) \\ &\quad +\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(1,1)}\boldsymbol{\Phi}_1(\boldsymbol{y}-\boldsymbol{\mu})+\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Phi}_1 T(\boldsymbol{y})+\boldsymbol{N}^{\top}\boldsymbol{\Psi}^{(0,1)}\boldsymbol{\Phi}_1. \end{aligned}$$

Finalmente, o estimador de $\boldsymbol{f}_{\boldsymbol{\mu}}$ para a $(m+1)$-ésimo iteração é definido da seguinte forma:

$$\begin{aligned} \boldsymbol{f}_{\boldsymbol{\phi}}^{(m+1)} &= (\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\boldsymbol{N}_{\boldsymbol{\phi}}-\lambda_{\boldsymbol{\phi}}\boldsymbol{K}_{\boldsymbol{\phi}})^{-1}\boldsymbol{N}_{\boldsymbol{\phi}}^{\top}\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1^2\{\tilde{\boldsymbol{S}}\boldsymbol{\gamma}^{(m)}+\boldsymbol{N}_{\boldsymbol{\phi}}\boldsymbol{f}_{\boldsymbol{\phi}}^{(m)}-\tilde{\boldsymbol{S}}\boldsymbol{\gamma}^{(m+1)} \\ &\quad +(-\boldsymbol{\Psi}^{(0,2)}\boldsymbol{\Phi}_1)^{-1}[\boldsymbol{\Psi}^{1,1}(\boldsymbol{y}-\boldsymbol{\mu})+T(\boldsymbol{y})+\boldsymbol{\Psi}^{(0,1)}]\}. \end{aligned}$$

APÊNDICE B

Banco de Dados

Tabela B.1: Dados referente as medidas dos frutos das goiabeiras "Pedro Sato".

daa	coleta	rep	long	trans	peso	volume
13	1	1	29.15	20.54	6.66	6
13	1	2	32.94	22.76	8.51	9
13	1	3	36.34	22.93	9.93	11
13	1	4	34.68	22.72	9.14	10
13	1	5	33.31	21.36	7.60	7
13	1	6	33.35	20.86	7.54	8
13	1	7	30.28	20.45	6.27	6
13	1	8	35.53	23.74	9.94	9
13	1	9	34.50	21.69	7.95	8
13	1	10	32.99	20.97	7.79	8
13	1	11	33.63	23.14	9.05	10
13	1	12	33.47	21.60	8.06	9
13	1	13	33.43	22.61	8.73	9
13	1	14	35.26	23.28	9.74	9
13	1	15	33.60	21.67	8.09	8
28	2	1	38.80	26.13	14.77	15
28	2	2	34.32	24.80	11.03	11
28	2	3	34.67	26.10	12.98	13
28	2	4	38.12	28.26	15.51	15

daa	coleta	rep	long	trans	peso	volume
28	2	5	39.01	27.67	15.29	14
28	2	6	31.36	25.68	10.23	11
28	2	7	31.10	23.23	8.34	8
28	2	8	40.02	27.17	15.26	15
28	2	9	38.84	27.97	15.38	15
28	2	10	36.81	22.89	9.79	10
28	2	11	36.20	22.66	9.57	10
28	2	12	35.65	26.74	13.09	12
28	2	13	35.42	24.45	10.49	10
28	2	14	38.88	25.90	12.55	13
28	2	15	38.69	25.65	12.52	13
42	3	1	46.94	31.21	22.70	24
42	3	2	39.72	29.07	16.87	17
42	3	3	42.19	29.82	17.98	19
42	3	4	47.55	32.13	24.19	23
42	3	5	39.11	26.78	13.20	12
42	3	6	38.31	27.30	13.76	15
42	3	7	47.54	33.05	25.30	26
42	3	8	43.88	28.44	19.58	20
42	3	9	37.67	27.49	13.86	14
42	3	10	43.88	30.34	19.73	20
42	3	11	42.15	29.60	17.95	19
42	3	12	44.44	30.76	19.91	21
42	3	13	40.46	28.46	17.18	17
42	3	14	38.88	28.90	17.57	16
42	3	15	41.63	29.84	19.31	19
59	4	1	46.68	34.00	28.69	60
59	4	2	46.15	32.43	24.56	50
59	4	3	43.35	31.93	20.81	40
59	4	4	46.05	33.36	26.29	50
59	4	5	47.87	37.00	33.98	65
59	4	6	43.18	31.46	21.15	40
59	4	7	42.32	31.96	22.52	45
59	4	8	44.29	33.74	25.79	55
59	4	9	45.86	33.75	26.21	55
59	4	10	44.99	32.19	23.39	50
59	4	11	44.30	30.74	20.50	45
59	4	12	48.68	35.14	31.13	65

daa	coleta	rep	long	trans	peso	volume
59	4	13	42.49	31.64	22.30	45
59	4	14	43.06	32.15	23.82	50
59	4	15	44.99	33.64	24.52	50
73	5	1	60.58	44.39	59.32	55
73	5	2	48.21	35.83	26.80	25
73	5	3	51.86	36.94	36.05	35
73	5	4	51.47	36.21	33.62	40
73	5	5	49.57	36.22	31.55	30
73	5	6	56.62	37.92	39.57	45
73	5	7	57.55	42.60	53.21	50
73	5	8	50.81	36.74	33.72	30
73	5	9	51.10	36.15	31.74	30
73	5	10	46.08	32.51	23.34	20
73	5	11	51.47	39.91	39.05	30
73	5	12	51.11	39.67	38.66	40
73	5	13	53.83	36.79	35.54	30
73	5	14	51.45	38.15	37.90	40
73	5	15	53.31	34.66	32.11	35
88	6	1	54.23	41.83	49.76	–
88	6	2	52.92	40.42	43.21	–
88	6	3	56.84	42.86	53.24	–
88	6	4	61.64	45.18	63.79	–
88	6	5	57.48	39.57	45.47	–
88	6	6	55.90	39.56	43.03	–
88	6	7	60.45	44.93	60.38	–
88	6	8	53.18	38.19	38.83	–
88	6	9	61.56	44.44	62.31	–
88	6	10	71.31	50.16	91.28	–
88	6	11	50.29	37.38	55.67	–
88	6	12	58.30	43.38	38.07	–
88	6	13	63.05	47.72	69.38	–
88	6	14	53.12	38.99	40.58	–
88	6	15	63.18	44.53	67.38	–
103	7	1	67.33	52.70	89.56	95
103	7	2	77.38	60.34	133.45	140
103	7	3	73.81	56.27	106.49	115
103	7	4	68.12	54.35	98.90	105
103	7	5	67.55	55.64	97.39	105

daa	coleta	rep	long	trans	peso	volume
103	7	6	72.38	55.21	105.87	115
103	7	7	63.88	48.58	72.28	70
103	7	8	65.59	50.42	82.42	85
103	7	9	76.54	60.11	132.32	140
103	7	10	80.18	60.50	129.66	140
103	7	11	78.20	55.11	112.11	120
103	7	12	75.72	53.02	96.80	100
103	7	13	62.63	49.46	74.71	80
103	7	14	65.38	50.65	82.37	90
103	7	15	78.66	60.17	139.60	150
110	8	1	70.92	63.32	138.52	160
110	8	2	75.00	62.64	141.10	150
110	8	3	72.31	60.28	124.10	120
110	8	4	85.45	65.77	172.81	180
110	8	5	84.61	59.59	136.45	150
110	8	6	77.38	63.67	141.59	160
110	8	7	73.54	63.12	145.38	160
110	8	8	80.77	67.32	165.66	180
110	8	9	77.06	61.80	128.61	140
110	8	10	73.68	66.40	143.71	160
110	8	11	76.09	61.32	125.95	140
110	8	12	72.92	62.47	133.40	150
110	8	13	75.11	66.08	155.88	180
110	8	14	75.45	65.05	156.87	180
110	8	15	77.62	64.76	164.59	180
117	9	1	84.51	71.44	223.25	230
117	9	2	88.99	71.99	224.93	225
117	9	3	77.80	65.21	170.05	195
117	9	4	75.92	64.52	153.65	185
117	9	5	85.38	70.42	203.16	230
117	9	6	81.86	70.55	193.77	220
117	9	7	78.75	66.54	160.00	200
117	9	8	85.59	72.30	227.52	270
117	9	9	91.25	78.00	252.58	295
117	9	10	91.81	73.95	249.30	290
117	9	11	82.52	68.46	196.40	230
117	9	12	83.81	75.51	232.32	250
117	9	13	84.68	69.27	173.55	200

daa	coleta	rep	long	trans	peso	volume
117	9	14	90.59	73.71	254.93	280
117	9	15	81.08	66.31	178.42	200
124	10	1	87.47	69.71	201.54	250
124	10	2	80.91	66.14	185.34	200
124	10	3	83.42	70.22	195.57	210
124	10	4	94.65	78.71	260.63	310
124	10	5	85.62	68.83	181.42	210
124	10	6	88.79	72.11	211.72	225
124	10	7	85.47	70.51	203.87	220
124	10	8	83.60	75.25	204.53	265
124	10	9	75.76	71.74	179.15	235
124	10	10	82.39	69.23	201.29	220
124	10	11	90.04	79.07	269.34	300
124	10	12	101.07	76.83	291.30	310
124	10	13	98.82	69.78	202.22	220
124	10	14	83.75	70.83	217.94	230
124	10	15	70.87	67.34	144.27	170
131	11	1	82.78	69.46	192.31	200
131	11	2	84.79	67.89	174.63	175
131	11	3	83.74	72.98	206.74	215
131	11	4	81.54	72.70	195.70	210
131	11	5	86.42	72.31	214.36	225
131	11	6	71.32	64.88	142.95	145
131	11	7	69.31	62.47	126.62	130
131	11	8	82.93	69.44	185.82	195
131	11	9	84.62	73.55	224.33	230
131	11	10	68.53	76.14	159.82	180
131	11	11	86.19	73.24	222.40	230
131	11	12	96.73	78.14	304.11	310
131	11	13	85.95	72.35	217.35	220
131	11	14	78.46	71.76	190.89	200
131	11	15	72.87	66.52	169.41	170
138	12	1	85.33	73.55	226.16	205
138	12	2	82.19	68.45	171.82	190
138	12	3	90.23	74.95	260.60	215
138	12	4	70.21	61.09	126.90	110
138	12	5	77.45	71.55	194.75	200
138	12	6	89.31	76.24	252.47	210

daa	coleta	rep	long	trans	peso	volume
138	12	7	68.80	64.34	160.60	140
138	12	8	84.57	74.71	242.65	240
138	12	9	77.09	66.28	178.00	140