

**Universidade de São Paulo  
Escola Superior de Agricultura Luiz de Queiroz**

**Modelo Linear Parcial Generalizado Simétrico**

**Julio Cezar Souza Vasconcelos**

Dissertação apresentada para obtenção do título de Mestre em Ciências, Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba  
2017**

**Julio Cezar Souza Vasconcelos**  
**Licenciado em Matemática**

**Modelo Linear Parcial Generalizado Simétrico**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **CRISTIAN MARCELO VILLEGAS LOBOS**

Dissertação apresentada para obtenção do título de Mestre  
em Ciências, Área de concentração: Estatística e Experi-  
mentação Agronômica

**Piracicaba**  
**2017**

**Dados Internacionais de Catalogação na Publicação  
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Vasconcelos, Julio Cezar Souza

Modelo Linear Parcial Generalizado Simétrico / Julio Cezar Souza Vasconcelos. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2017 .

71 p.

Dissertação (Mestrado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Modelo linear parcial generalizado simétrico 2. Modelos generalizados semiparamétricos 3. Distribuições simétricas 4. Modelos GAMLSS . I. Título.

## DEDICATÓRIA

*Dedico esse trabalho ao meu pai que não está mais entre nós, José Maria Vasconcelos, a minha mãe Maria de Souza Vasconcelos e ao meu irmão, Juliano Souza Vasconcelos. Muito obrigado pela ajuda em todos os momentos.*

## AGRADECIMENTOS

Quero agradecer em primeiro lugar a Deus, por dar-me a saúde, capacidade e oportunidade de continuar a minha formação acadêmica. Ao meu orientador Cristian Marcelo Villegas Lobos pela confiança, parceria, dedicação, apoio e amizade; é uma honra e tenho um imenso prazer de ter sido orientado pelo professor Cristian [”obrigado amigo”]. Gostaria de agradecer a minha família, em especial minha mãe Fátima [”obrigado mãe por ser minha mãe e por sempre me ajudar em tudo e pela suas mensagens de apoio”]; meu querido irmão Juliano [”obrigado meu querido parceiro, admiro muito sua sabedoria ”]; e a uma pessoa que infelizmente não está mais entre nós, meu querido pai Zé Maria [”muito obrigado meu amigo, sempre me lembrarei de tudo que você fez por mim e de ter sido meu grande mestre”]. Sou muito grato aos professores e funcionários do Departamento de Ciências Exatas da ESALQ-USP, aos professores componentes da minha banca: Dr<sup>a</sup> Clarice Garcia Borges Demétrio, Dr. César Gonçalves de Lima, Dr. Gilberto Alvarenga Paula e Dr. Rafael Pimentel Maia, agradeço também meus amigos e colegas de estudo. Agradeço a Universidade de São Paulo por me dar a oportunidade de conseguir conquistar o título de Mestre em Ciências, ao apoio financeiro do CNPq durante o desenvolvimento desse trabalho. Agradeço a todos que contribuíram para a realização dessa empreitada.

## SUMÁRIO

RESUMO . . . . .	7
ABSTRACT . . . . .	8
LISTA DE FIGURAS . . . . .	9
LISTA DE TABELAS . . . . .	10
1 INTRODUÇÃO . . . . .	11
1.1 Objetivos . . . . .	12
2 REVISÃO BIBLIOGRÁFICA . . . . .	13
2.1 Introdução . . . . .	13
2.2 Família exponencial . . . . .	13
2.3 Família de distribuições simétricas . . . . .	14
2.3.1 Distribuição Hiperbólica . . . . .	15
2.3.2 Distribuição <i>Slash</i> . . . . .	16
2.3.3 Distribuição Exponencial potência . . . . .	17
2.3.4 Distribuição t-Student . . . . .	18
2.4 Modelos lineares generalizados . . . . .	19
2.5 Modelos lineares generalizados simétricos . . . . .	20
2.6 Conceitos associados a regressão não-paramétrica . . . . .	21
2.6.1 Graus de liberdade efetivos . . . . .	21
2.6.2 Validação cruzada . . . . .	21
2.6.3 Validação cruzada generalizada . . . . .	22
2.6.4 Spline . . . . .	22
2.6.4.1 Spline cúbico . . . . .	23
2.6.4.2 Spline cúbico natural . . . . .	23
2.6.4.3 B-spline . . . . .	24
2.6.5 Método de penalização . . . . .	24
2.6.6 Algoritmo <i>backfitting</i> . . . . .	25
2.7 Modelo parcialmente linear . . . . .	26
2.8 Modelos parcialmente lineares generalizados . . . . .	27
2.9 Modelos GAMLSS . . . . .	28
2.9.1 Componentes de um modelo GAMLSS . . . . .	28
2.9.2 Algoritmos CG e RS . . . . .	29
2.9.3 Pacote <i>gamlss</i> . . . . .	30
2.10 Análise residual . . . . .	30
2.11 Influência local . . . . .	31
2.12 Considerações . . . . .	33
3 MATERIAL E MÉTODOS . . . . .	35
3.1 Introdução . . . . .	35

3.2	Modelo Linear Parcial Generalizado Simétrico . . . . .	35
3.3	Função Verossimilhança Penalizada . . . . .	36
3.4	Função escore penalizada . . . . .	37
3.5	Matriz hessiana penalizada . . . . .	37
3.6	Processo iterativo para o MLPGS . . . . .	39
3.7	Influência local . . . . .	40
3.7.1	Ponderação de casos . . . . .	41
3.7.2	Perturbação da variável resposta . . . . .	41
3.8	Considerações . . . . .	42
4	RESULTADOS . . . . .	43
4.1	Introdução . . . . .	43
4.2	Modelos propostos . . . . .	44
4.2.1	Modelos Normal, t-Student e Exponencial potência, com função de ligação identidade. . . . .	45
4.2.2	Modelos Normal, t-Student e Exponencial potência, com função de ligação inversa. . . . .	49
4.3	Seleção do modelo . . . . .	53
5	DISCUSSÃO . . . . .	57
5.1	Aspectos computacionais . . . . .	57
5.2	Trabalhos futuros . . . . .	57
	REFERÊNCIAS . . . . .	59
	APÊNDICES . . . . .	63

## RESUMO

### Modelo Linear Parcial Generalizado Simétrico

Neste trabalho foi proposto o modelo linear parcial generalizado simétrico, com base nos modelos lineares parciais generalizados e nos modelos lineares simétricos, em que a variável resposta segue uma distribuição que pertence à família de distribuições simétricas, considerando um preditor linear que possui uma parte paramétrica e uma não paramétrica. Algumas distribuições que pertencem a essa classe são as distribuições: Normal, t-Student, Exponencial potência, *Slash* e Hiperbólica, dentre outras. Uma breve revisão dos conceitos utilizados ao longo do trabalho foram apresentados, a saber: análise residual, influência local, parâmetro de suavização, *spline*, spline cúbico, spline cúbico natural e algoritmo *backfitting*, dentre outros. Além disso, é apresentada uma breve teoria dos modelos GAMLSS (modelos aditivos generalizados para posição, escala e forma). Os modelos foram ajustados utilizando o pacote *gamlss* disponível no software livre R. A seleção de modelos foi baseada no critério de Akaike (AIC). Finalmente, uma aplicação é apresentada com base em um conjunto de dados reais da área financeira do Chile.

**Palavras-chave:** Modelo linear parcial generalizado simétrico; Modelos generalizados semiparamétricos; Distribuições simétricas; Modelos GAMLSS



## ABSTRACT

### Linear Model Partial Generalized Symmetric

In this work we propose the symmetric generalized partial linear model, based on the generalized partial linear models and symmetric linear models, that is, the response variable follows a distribution that belongs to the symmetric distribution family, considering a linear predictor that has a parametric and a non-parametric component. Some distributions that belong to this class are distributions: Normal, t-Student, Power Exponential, Slash and Hyperbolic among others. A brief review of the concepts used throughout the work was presented, namely: residual analysis, local influence, smoothing parameter, spline, cubic spline, natural cubic spline and *backfitting* algorithm, among others. In addition, a brief theory of GAMLSS models is presented (generalized additive models for position, scale and shape). The models were adjusted using the package *gamlss* available in the free R software. The model selection was based on the Akaike criterion (AIC). Finally, an application is presented based on a set of real data from Chile's financial area.

**Keywords:** Symmetric generalized partial linear model; Symmetric distributions; Generalized semi-parametric models; Models GAMLSS

## LISTA DE FIGURAS

2.1	Gráfico das curvas das funções de densidade das distribuições Normal padrão e Hiperbólica com $\nu = 0,85$ e $\nu = 0,9$ . . . . .	16
2.2	Gráfico das curvas das funções de densidade das distribuições Normal padrão e <i>Slash</i> com $\nu = 2$ e $\nu = 5$ . . . . .	17
2.3	Gráfico das curvas das funções de densidade das distribuições Normal padrão e Exponencial potência com $\nu = -0,3$ , $\nu = 0,3$ e $\nu = 1$ . . . . .	18
2.4	Gráfico das curvas das funções de densidade das distribuições Normal padrão e t-Student com $\nu = 2$ e $\nu = 5$ . . . . .	19
4.1	Diagrama de dispersão: (a) Rentabilidade AFP Cuprum versus IPSA e (b) Rentabilidade AFP Cuprum versus tempo. . . . .	43
4.2	Gráfico normal de probabilidade para o resíduo quantílico (esquerda) e função não-paramétrica (direita) sob o modelo Normal. . . . .	44
4.3	Função não-paramétrica dos modelos Normal, t-Student e Exponencial potência ajustada (spline cúbico) à variável explicativa tempo considerando a função de ligação identidade, com a respectiva banda de confiança de 95%. . . . .	47
4.4	Gráficos normais de probabilidade para o resíduo quantílico sob os modelos Normal, t-Student e Exponencial potência ajustados aos dados da AFP com função de ligação identidade. . . . .	48
4.5	Gráfico valores preditos versus observados dos modelos Normal, t-Student e Exponencial potência, considerando a função de ligação identidade. . . . .	49
4.6	Função não-paramétrica dos modelos Normal, t-Student e Exponencial potência ajustada (spline cúbico) à variável explicativa tempo considerando a função de ligação inversa, com a respectiva banda de confiança de 95%. . . . .	51
4.7	Gráficos normais de probabilidade para o resíduo quantílico sob os modelos Normal, t-Student e Exponencial potência ajustados aos dados da AFP com função de ligação inversa. . . . .	52
4.8	Gráfico valores preditos versus observados dos modelos Normal, t-Student e Exponencial potência, considerando a função de ligação inversa. . . . .	53
4.9	Histograma dos resíduos quantílicos do modelo Exponencial potência com função de ligação identidade. . . . .	55

## LISTA DE TABELAS

2.1	Algumas distribuições da família exponencial. . . . .	14
3.1	Expressões $W_h(u)$ , $W'_h(u)$ e $v$ de algumas distribuições simétricas. . . . .	38
4.1	Estimativas de máxima verossimilhança dos modelos Normal, t-Student e Exponencial potência ajustado aos dados da área de finanças do Chile, considerando a função de ligação identidade. . . . .	46
4.2	Estimativas máximas verossimilhança dos modelos Normal, t-Student e Exponencial potência ajustado aos dados da área de finanças do Chile, considerando a função de ligação inversa. . . . .	50
4.3	Resultados do critério AIC para os modelos Normal, t-Student e Exponencial potência, com funções de ligação identidade e inversa. . . . .	54
4.4	Resumo do resíduo quantílico produzido pelo gamlss do modelo Exponencial potência, considerando a função de ligação identidade. . . . .	54

# 1 INTRODUÇÃO

Os modelos lineares generalizados propostos por NELDER e WEDDERBURN (1972) têm sido aplicados em diversas áreas do conhecimento, tais como, agronomia, engenharia florestal, ciências biológicas, toxicologia, entre outras.

Com o avanço computacional outras teorias interessantes têm surgido, tais como os modelos lineares generalizados semiparamétricos propostos por GREEN e YANDELL (1985) em que os autores acrescentaram no preditor linear um termo não paramétrico que geralmente está associado ao tempo. Em particular, eles consideram o modelo de regressão logística para dados de um ensaio biológico do agente de extintores de chama realizado nos Estados Unidos no Programa Nacional de Toxicologia (DINSE e LAGAKOS, 1983). Os dados são de ratos, que foram expostos a várias doses do agente. A variável resposta binária indica a presença ou ausência de hiperplasia do canal biliar no momento da morte, e quatro variáveis explicativas: logaritmo da dose, peso inicial, posição da gaiola e idade no momento da morte. GREEN e YANDELL (1985) ajustaram o modelo de regressão logística, no qual as três primeiras covariáveis foram consideradas de forma paramétrica e a quarta variável (idade no momento da morte) de forma não paramétrica no preditor linear.

As distribuições simétricas estudadas por KELKER (1970); CAMBANIS *et al.* (1981); CHMIELEWSKI (1981); LANGE e SINSHEIMER (1993) vêm sendo utilizadas como uma interessante alternativa à distribuição Normal quando há observações atípicas, já que possuem caudas mais pesadas. Para resolver este problema GALEA *et al.* (2003) propuseram o uso de modelos de regressão simétricos. Uma extensão deste trabalho foi feito por VILLEGAS *et al.* (2013), em que foi considerado o uso de uma função de ligação no mesmo sentido dos modelos lineares generalizados (NELDER e WEDDERBURN, 1972).

O objetivo desse trabalho é desenvolver um novo modelo com base nos modelos lineares generalizados semiparamétricos (GREEN e YANDELL, 1985) e nos modelos lineares simétricos, em que, a variável resposta segue uma distribuição que pertence à família de distribuições simétricas considerando um preditor linear que possui uma parte paramétrica e uma não paramétrica. O componente que é controlado de forma não paramétrica foi estimado utilizando método de suavização por splines. Para encontrar o parâmetro de suavização, utilizou-se o método de validação cruzada generalizada. Ainda, serão desenvolvidas análises de resíduos e de influência local (COOK, 1986). Todas as análises foram feitas utilizando o software livre R (TEAM, 2015), os ajustes dos modelos foram realizados utilizando o pacote *gamlss* (modelos aditivos generalizados para posição, escala e forma), desenvolvido por RIGBY e STASINOPOULOS (1996).

## 1.1 Objetivos

O objetivo principal deste trabalho é propor uma extensão do modelo linear generalizado simétrico (VILLEGAS *et al.*, 2013). Os objetivos específicos são:

1. Estimar os parâmetros do modelo utilizando o método de máxima verossimilhança, com base na função de verossimilhança penalizada;
2. Determinar um critério para escolher o parâmetro de suavização;
3. Propor medidas de diagnóstico, tal como influência local;
4. Aplicar o modelo a um conjunto de dados reais;
5. Deixar disponíveis os códigos do trabalho com base no software livre R.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 Introdução

Embora a suposição de normalidade seja adequada em muitas situações, pode não ser apropriada quando vir de observações de uma distribuição simétrica com caudas mais leves ou mais pesadas do que a distribuição normal. Nestes casos, a família de distribuições simétricas (FANG *et al.*, 1990) surge como uma alternativa interessante, principalmente devido às possíveis variações da curtose e os aspectos de robustez relacionadas às estimativas de máxima verossimilhança versus observações extremas. A família de distribuições simétricas recebeu muita atenção nos últimos anos e tem sido aplicada em diferentes áreas do conhecimento. Para um estudo detalhado das suas propriedades e algumas aplicações veja, por exemplo, KELKER (1970); CAMBANIS *et al.* (1981); CHMIELEWSKI (1981); FANG *et al.* (1990); LANGE e SINSHEIMER (1993); CYSNEIROS *et al.* (2007); PAULA e CYSNEIROS (2009).

Nesse capítulo, será realizada uma revisão sobre alguns conceitos essenciais, como família exponencial, família de distribuições simétricas, modelos lineares generalizados e modelos lineares generalizados simétricos, sendo que nesta breve revisão serão abordados alguns conceitos utilizados na regressão não paramétrica, tais como: formas de como determinar o parâmetro de suavização, cálculo das estimativas do componente não-paramétrico por intermédio de métodos de suavização por splines, critérios de penalização, algoritmo *backfitting*, etc. Posteriormente, os conceitos do modelo parcialmente linear e parcialmente linear generalizado, aditivo generalizado para posição, escala e forma (GAMLSS) serão apresentados. Por fim, será apresentada uma introdução sobre resíduo quantílico e influência local.

### 2.2 Família exponencial

A família exponencial de distribuições foi proposta por Koopman, Pitman e Dar-mois em meados da década de 1930, por meio de aplicações de propriedades de estatística suficiente (CORDEIRO e DEMETRIO, 1986). Tal família é composta por um conjunto de distribuições, como a Normal, Poisson, Binomial, Binomial Negativa, Gama Normal Inversa, entre outras e a Tabela 2.1 exhibe os seus respectivos componentes na família exponencial. Fisher teve uma participação fundamental nesses estudos, estabelecendo na estatística a família exponencial. A família exponencial tem grande ênfase na área de modelos de regressão, principalmente se tratando de modelos lineares generalizados (NELDER e WEDDERBURN, 1972).

Uma variável aleatória  $Y$  pertence à família exponencial se sua função de proba-

**Tabela 2.1.** Algumas distribuições da família exponencial.

Distribuição	$\phi$	$\theta$	$b(\theta)$	$c(y, \phi)$	$\mu(\theta)$	$V(\mu)$
$N(\mu, \sigma^2)$	$\sigma^2$	$\mu$	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left[ \frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right]$	$\theta$	1
$P(\mu)$	1	$\log(\mu)$	$e^\theta$	$-\log(y!)$	$e^\theta$	$\mu$
$B(m, \pi)$	1	$\log\left(\frac{\mu}{m-\mu}\right)$	$m \log(1 + e^\theta)$	$\log\left(\frac{m}{y}\right)$	$\frac{me^\theta}{1+e^\theta}$	$\frac{\mu}{m}(m - \mu)$
$BN(\mu, k)$	1	$\log\left(\frac{\mu}{\mu+k}\right)$	$-k \log(1 - e^\theta)$	$\log\left[\frac{\Gamma(k+y)}{\Gamma(k)y!}\right]$	$k \frac{e^\theta}{1-e^\theta}$	$\mu \left(\frac{\mu}{k} + !\right)$
$G(\mu, \nu)$	$\nu^{-1}$	$-\frac{1}{\mu}$	$-\log(-\theta)$	$\nu \log(\nu y) - \log(y) - \log \Gamma(\nu)$	$-\frac{1}{\theta}$	$\mu^2$
$IG(\mu, \sigma^2)$	$\sigma^2$	$-\frac{1}{2\mu^2}$	$-(-2\theta)^{1/2}$	$-\frac{1}{2} \left[ \log(2\pi\sigma^2 y^3) + \frac{1}{\sigma^2 y} \right]$	$(-2\theta)^{-1/2}$	$\mu^3$

bilidade pode ser escrita como

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \right\}, \quad \phi > 0 \quad (2.1)$$

em que  $b(\theta)$  e  $c(y; \phi)$  são funções conhecidas. A notação  $Y \sim FE(\mu, \phi)$ , com  $\mu = E(Y)$  será utilizada. O valor esperado e a variância das variáveis aleatórias com função densidade de probabilidade (2.1) são, respectivamente

$$E(Y) = \mu = b'(\theta) \text{ e } Var(Y) = \phi b''(\theta).$$

Observa-se, então, que  $\phi$  é um parâmetro de dispersão do modelo e seu inverso  $\phi^{-1}$ , uma medida de precisão (CORDEIRO e DEMETRIO, 1986). Duas distribuições pertencentes a família exponencial são exibidas a seguir.

### Distribuição Normal

$Y \sim N(\mu, \sigma^2)$  tem função densidade de probabilidade

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \frac{(y - \mu)^2}{\sigma^2} \right], \quad y > 0, \quad \mu \in (-\infty, \infty), \quad \sigma^2 > 0, \quad (2.2)$$

em que  $\theta = \mu$ ,  $b(\theta) = \frac{\theta^2}{2}$ ,  $\phi = \sigma^2$  e  $c(y; \phi) = -\frac{1}{2} \left[ \frac{y^2}{\phi} + \log(2\pi\phi) \right]$ .

### Distribuição Binomial

$Y \sim Bin(m, \pi)$  tem função de probabilidade

$$f(y; \pi) = \binom{m}{y} \pi^y (1 - \pi)^{m-y}, \quad y = 0, \dots, m, \quad \pi \in [0, 1], \quad (2.3)$$

em que  $\theta = \log\left(\frac{\pi}{1-\pi}\right)$ ,  $b(\theta) = m \log(1 + \exp\{\theta\})$ ,  $\phi = 1$  e  $c(y; \phi) = \log\left(\frac{m}{y}\right)$ .

## 2.3 Família de distribuições simétricas

Nessa seção, serão apresentadas algumas distribuições que pertencem à família de distribuições simétricas univariadas.

Uma variável aleatória  $Y$  tem uma distribuição simétrica, com parâmetro de posição  $\mu$ , parâmetro de escala  $\phi$  e função geradora de densidade  $h(\cdot)$ , se a sua função de densidade de probabilidade pode ser expressa como

$$f(y, \mu, \phi) = \frac{1}{\sqrt{\phi}} h(u), \quad (2.4)$$

em que  $u = \frac{(y - \mu)^2}{\phi}$ ,  $\mu \in \mathbb{R}$ ,  $\phi > 0$  e  $h(\cdot)$  é uma função definida positiva em  $\mathbb{R}^+$ , de tal modo que  $\int_0^\infty u^{-\frac{1}{2}} h(u) du = 1$  (KELKER, 1970). Denotaremos:  $Y \sim S(\mu, \phi, h)$ . A função característica toma a forma  $\exp(it\mu)\psi(t^2\phi)$  para alguma função  $\psi(\cdot)$ , em que  $i = \sqrt{-1}$ . KELKER (1970) mostrou que se  $u^{\frac{1}{2}(1-k)-1}h(u)$  é integrável, então existem os momentos de ordem  $k$  da variável aleatória  $Y$ . Além disso, se a esperança e a variância de  $Y$  existem, então elas são dadas por  $E(Y) = \mu$  e  $Var(Y) = \xi_h\phi$ , respectivamente, em que  $\xi_h = -2\psi'(0)$  é uma constante positiva.

Na maioria das distribuições simétricas, a função  $h(\cdot)$  depende de um parâmetro adicional, chamado parâmetro de forma, que vai ser indicado como  $\nu$ , controlando a curtose da distribuição. Este parâmetro pode ser fixado ou estimado a partir dos dados. Com base nas distribuições com essas características, neste trabalho, a atenção estará voltada para as seguintes distribuições: Normal, Hiperbólica, *Slash*, Exponencial potência e t-Student. A seguir, serão definidas as funções densidade de probabilidade das distribuições citadas.

### 2.3.1 Distribuição Hiperbólica

Diz-se que uma variável aleatória  $Y$  tem uma distribuição hiperbólica (HOWLANDER, 1989), com parâmetro de posição  $\mu$ , parâmetro de escala  $\phi$ , parâmetro de forma  $\nu$ , se sua função geradora de densidade for expressa por:

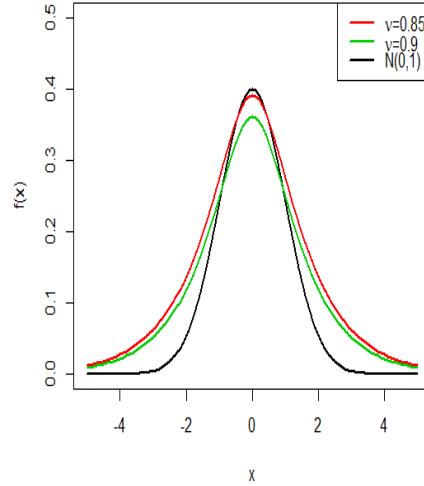
$$h(u) = \frac{1}{2K_1(\nu)} \exp\left\{-\nu\sqrt{1+u}\right\}, \quad u > 0, \quad (2.5)$$

em que  $K_\lambda(\nu) = \frac{1}{2} \int_0^\infty x^{\lambda-1} e^{-\frac{1}{2}\nu(x+x^{-1})} dx$  é a função de Bessel modificada tipo 3 de ordem  $\lambda$ , veja GRADSHTEYN e RYZHIK (2000) para algumas propriedades desta função. Denotemos por  $Y \sim Sh(\mu, \phi, \nu)$ . Temos que  $E(Y) = \mu$  e  $Var(Y) = \xi_h\phi$  com  $\xi_h = \frac{K_2(\nu)}{\nu K_1(\nu)}$ . O coeficiente de curtose é dado por  $k = 3 + \frac{3K_1(\nu)K_3(\nu)}{K_2(\nu)^2}$ . O parâmetro  $\nu$  é o parâmetro de forma e valores pequenos de  $\nu$  indicam distribuições com caudas mais pesadas do que a distribuição normal. Além disso, quando  $\nu$  tende para  $+\infty$  a variável aleatória  $Y$  pode ser representada como uma distribuição normal.

Se definirmos  $Y = \mu + \sqrt{V}Z$ , em que  $V \sim GIG(1, \phi^2, \nu^2)$  é a distribuição Inversa Gaussiana Generalizada (JØRGENSEN, 1982) e  $Z \sim N(0, 1)$  são independentes, então  $Y \sim Sh(\mu, \phi, \nu)$ . Este resultado pode ser usado, por exemplo, para simular variáveis aleatórias



com uma distribuição hiperbólica. Na Figura 2.1 é apresentada uma comparação entre as distribuições Normal padrão e Hiperbólica com  $\nu = 0,85$  e  $\nu = 0,9$ . Observa-se que a distribuição Hiperbólica possui caudas mais pesadas do que a distribuição Normal.



**Figura 2.1.** Gráfico das curvas das funções de densidade das distribuições Normal padrão e Hiperbólica com  $\nu = 0,85$  e  $\nu = 0,9$ .

### 2.3.2 Distribuição *Slash*

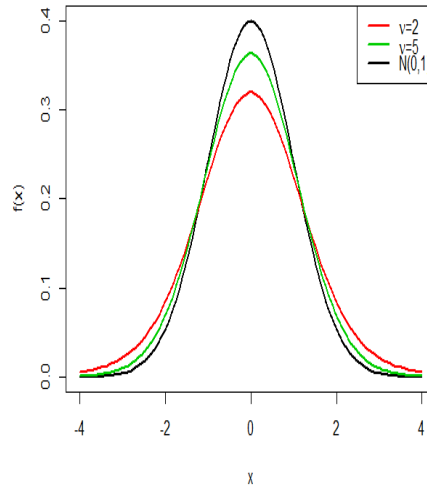
Uma variável aleatória  $Y$  tem uma distribuição *Slash* (LANGE e SINSHEIMER, 1993), com parâmetro de posição  $\mu$ , parâmetro de escala  $\phi$ , parâmetro de forma  $\nu$  e função geradora de densidade expressa por:

$$h(u) = \frac{\nu}{\sqrt{2\pi}} G\left(\nu + \frac{1}{2}, \frac{u}{2}\right), \quad \nu > 0, \quad (2.6)$$

em que  $G(\beta, S) = \int_0^1 x^{\beta-1} e^{-sx} dx$  é a função gama incompleta. Denotaremos:  $Y \sim Sl(\mu, \phi, \nu)$ . Temos que,  $E(Y) = \mu$  e  $Var(Y) = \xi_h \phi$  com  $\xi_h = \frac{\nu}{\nu - 1}$  para  $\nu > 1$ . O coeficiente de curtose é dado por  $k = 3 + \frac{3\nu}{\nu - 2}$  para  $\nu > 2$ .

Se definirmos  $Y = \mu + \sqrt{\phi} V^{-\frac{1}{2}} Z$ , em que  $V$  é uma variável aleatória positiva e  $Z \sim N(0, 1)$  são independentes, então  $Y \sim Sl(\mu, \phi, \nu)$ . Esta propriedade é útil para gerar amostras aleatórias de uma distribuição *Slash*.

Na Figura 2.2 é apresentada uma comparação entre as distribuições Normal padrão e *Slash* com  $\nu = 2$  e  $\nu = 5$ . A distribuição *Slash* possui caudas mais pesadas do que a distribuição Normal.



**Figura 2.2.** Gráfico das curvas das funções de densidade das distribuições Normal padrão e *Slash* com  $\nu = 2$  e  $\nu = 5$ .

### 2.3.3 Distribuição Exponencial potência

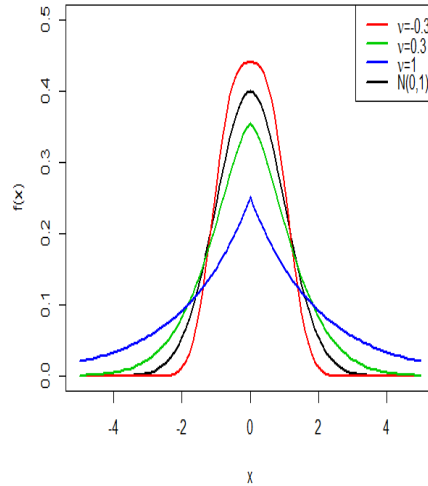
Uma variável aleatória  $Y$  tem uma distribuição Exponencial potência (BOX e TIAO, 1973), com parâmetro de posição  $\mu$ , parâmetro de escala  $\phi$ , parâmetro de forma  $\nu$  e função geradora de densidade expressa por:

$$h(u) = C(\nu) \exp \left\{ -\frac{1}{2} u^{\frac{1}{1+\nu}} \right\}, \quad -1 < \nu \leq 1, \quad (2.7)$$

em que  $C(\nu)^{-1} = \Gamma \left( \frac{3+\nu}{2} \right) 2^{\frac{(3+\nu)}{2}}$ . Representamos por  $Y \sim EP(\mu, \phi, \nu)$ . A distribuição Exponencial potência, descrita por BOX e TIAO (1973) é uma distribuição simétrica com caudas mais curtas ou caudas mais pesadas do que a distribuição normal.

Temos que  $E(Y) = \mu$ ,  $Var(Y) = 2^{1+\nu} \left[ \Gamma \left( \frac{3(1+\nu)}{2} \right) / \Gamma \left( \frac{1+\nu}{2} \right) \right] \phi$ . O coeficiente de curtose é dado por  $k = \left\{ \frac{5}{2}(1+\nu) \right\} \Gamma \left( \frac{1+\nu}{2} \right) / \Gamma^2 \left( \frac{3}{2}(1+\nu) \right)$ . Deve-se observar que, para  $\nu > 0$ , temos  $k > 3$ , isto é, a distribuição é leptocúrtica e para  $\nu < 0$ ,  $k < 3$ , a distribuição é platicúrtica. O parâmetro  $\nu$  pode ser considerado como uma medida de curtose, bem como uma medida da não normalidade, porque quando  $\nu \rightarrow \infty$  a distribuição Exponencial potência converge para a distribuição Normal. Além disso, quando  $\nu = 1$ , temos a distribuição Exponencial dupla (distribuição de Laplace) e se  $\nu$  tende para  $-1$ , a distribuição converge para uma distribuição Uniforme no intervalo  $(\mu - \sqrt{3\phi}, \mu + \sqrt{3\phi})$ . Se definirmos  $Y = (2W)^{\frac{1+\nu}{2}} V$ , em que  $V \sim U(-1, 1)$  e  $W \sim G(\frac{3+\nu}{2}, 1)$  são independentes, então,  $Y \sim EP(0, 1, \nu)$  (DEVROYE, 1986). Utilizando esta propriedade, pode-se gerar amostras de uma distribuição  $EP(0, 1, \nu)$  e, portanto, de uma distribuição  $EP(\mu, \phi, \nu)$ .

Na Figura 2.3 é apresentada uma comparação entre as distribuições Normal padrão e Exponencial potência com  $\nu = -0,3$ ,  $\nu = 0,3$  e  $\nu = 1$ . A distribuição Exponencial potência com  $\nu = -0,3$  possui caudas mais leves do que a distribuição Normal, mas com  $\nu = 0,3$  e  $\nu = 1$ , a função de densidade Hiperbólica possui caudas mais pesadas do que a distribuição normal.



**Figura 2.3.** Gráfico das curvas das funções de densidade das distribuições Normal padrão e Exponencial potência com  $\nu = -0,3$ ,  $\nu = 0,3$  e  $\nu = 1$ .

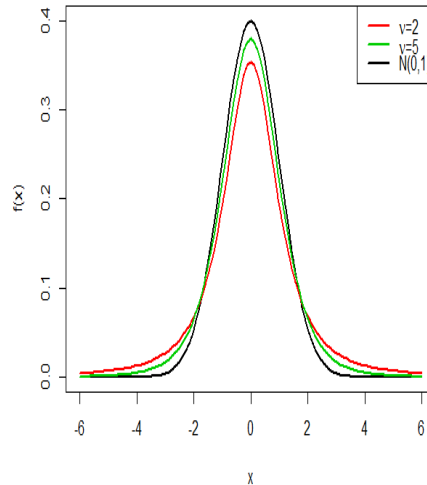
### 2.3.4 Distribuição t-Student

Diz-se que uma variável aleatória  $Y$  tem uma distribuição t-Student, com parâmetro de posição  $\mu$ , parâmetro de escala  $\phi$ , parâmetro de forma  $\nu$  e função geradora de densidade expressa por:

$$h(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{u^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}, \quad \nu > 0. \quad (2.8)$$

Denotemos por,  $Y \sim t(\mu, \phi, \nu)$ . Se definirmos  $Y = \mu + \sqrt{\phi V} Z$ , em que  $V \sim GI\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$  (gama Inversa) e  $Z \sim N(0, 1)$  são independentes, então  $Y \sim t(\mu, \phi, \nu)$ . Temos que,  $E(Y) = \mu$  para  $\nu > 1$  e  $Var(Y) = \xi_h \phi$  com  $\xi_h = \frac{\nu}{(\nu-2)}$  para  $\nu > 2$ . O coeficiente de curtose é dado por  $k = 3 + \frac{6}{\nu-4}$  para  $\nu > 4$ . Podemos observar que para  $k = 3$ , a distribuição t-Student será a distribuição normal. Além disso, podemos ver que a distribuição t-Student com parâmetros  $(\mu, \phi, \nu)$  converge para uma distribuição Normal com média  $\mu$  e variância  $\phi$  quando  $\nu \rightarrow +\infty$ . Se  $\nu = 1$ , temos a distribuição Cauchy com parâmetros  $\mu$  e  $\phi$ , e que não possui momentos de primeira e segunda ordem. Na Figura 2.4 é apresentada uma

comparação entre as distribuições Normal padrão e t-Student com  $\nu = 2$  e  $\nu = 5$ . Em ambos os casos a distribuição t-Student apresenta caudas mais pesadas do que a distribuição Normal.



**Figura 2.4.** Gráfico das curvas das funções de densidade das distribuições Normal padrão e t-Student com  $\nu = 2$  e  $\nu = 5$ .

## 2.4 Modelos lineares generalizados

Os modelos lineares generalizados propostos por NELDER e WEDDERBURN (1972) têm sido aplicados em diversas áreas do conhecimento, tais como, agronomia, engenharia florestal, ciências biológicas, toxicologia, entre outras.

### Componentes de um modelo linear generalizado

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes. Definimos um modelo linear generalizado (NELDER e WEDDERBURN, 1972) quando

1.  $Y_i \sim FE(\mu_i, \phi)$  com  $\mu_i = E(Y_i)$ ;
2.  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ ;
3.  $g(\mu_i) = \eta_i$ ,

em que  $\eta_i$  é o preditor linear e  $g(\mu_i)$  é a função de ligação, ou seja, faz a ligação do componente aleatório com a parte sistemática.

Os casos particulares de um modelo linear generalizado são:

1. Modelo de regressão normal;
2. Regressão logística.

## 2.5 Modelos lineares generalizados simétricos

Os modelos lineares simétricos têm assumido uma função de ligação identidade entre a média resposta e um preditor linear (GALEA *et al.*, 2003). GALEA *et al.* (2005) consideraram uma função de ligação identidade entre a resposta média e um preditor não-linear. Nesses modelos, a interpretação dos parâmetros é crucial. Então, a aplicação de modelos lineares generalizados simétricos, nestes casos, pode levar a uma perda de interpretação biológica dos parâmetros.

A classe de modelos lineares generalizados simétricos é uma generalização de modelos lineares simétricos no mesmo sentido de modelos lineares generalizados, por meio de uma função de ligação que relaciona a resposta média e um preditor linear. Isto nos permite usar diferentes funções de ligação para estudar o mesmo conjunto de dados. Além disso, a notação facilita a derivação de um algoritmo de mínimos quadrados ponderados iterativamente para a estimativa dos parâmetros, bem como alguns resultados inferenciais e gráficos de diagnóstico, semelhantes aos de modelos lineares generalizados.

### Componentes de um modelo linear generalizado simétrico

Aqui, será assumido que as variáveis aleatórias  $Y_1, \dots, Y_n$  são independentes e que cada  $Y_i$  tem uma distribuição simétrica univariada com densidade

$$f(y_i, \mu_i, \phi) = \frac{1}{\sqrt{\phi}} h(u_i), \quad (2.9)$$

em que  $u_i = \frac{(y_i - \mu_i)^2}{\phi}$ , isto é,  $Y_i \sim S(\mu_i, \phi, h)$ , com  $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ . O modelo (2.9) é definido como no modelo linear simétrico. Seja  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$  um preditor linear, em que  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  é um vetor  $p \times 1$  de valores de variáveis explicativas e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  é um vetor de parâmetros de dimensão  $p \times 1$ . Ainda, suponha que o parâmetro  $\mu_i$  satisfaz

$$g(\mu_i) = \eta_i, \quad (2.10)$$

com  $g(\cdot)$ , sendo a função de ligação monótona e diferenciável como em modelos lineares generalizados. O modelo definido em (2.9) e (2.10) é chamado modelo linear generalizado simétrico (VILLEGAS *et al.*, 2013). Considerando que  $\mu_i$  pode assumir qualquer valor real, as possíveis funções de ligação podem ser  $g(\mu_i) = \mu_i$  (identidade) e  $g(\mu_i) = \mu_i^{-1}$  (recíproca). No entanto, se para um caso particular de dados simétricos  $\mu_i$  é positiva, então, outras funções de ligação podem ser consideradas, como  $g(\mu_i) = \log(\mu_i)$  (logarítmica) e  $g(\mu_i) = \sqrt{\mu_i}$  (raiz quadrada).

## 2.6 Conceitos associados a regressão não-paramétrica

### 2.6.1 Graus de liberdade efetivos

Para estimar o componente não-paramétrico é fundamental estimar o número de graus de liberdade total que está sendo utilizado, isto é, estimar o número de graus de liberdade efetivos que é o grau do polinômio da função não-paramétrica. Para os modelos parcialmente em que a variável explicativa segue uma distribuição pertencente a família de distribuições simétricas, a quantidade de parâmetros a serem estimados é  $q = p + 1 + df(\lambda)$ , no qual  $df(\lambda)$  refere-se aos parâmetros do componente não-paramétrico e  $p + 1$  aos parâmetros de  $\beta$  e  $\phi$ . GREEN e SILVERMAN (1993) propuseram uma expressão com a finalidade de obter o valor  $df(\lambda)$  para os MLGs, que pode ser ajustada para os modelos parcialmente em que a variável resposta segue uma distribuição pertencente a família de distribuições simétricas.

### 2.6.2 Validação cruzada

Em AYDIN (2014), é introduzido o conceito da estimativa baseada em *splines* de suavização da seguinte forma. Seja  $\mathbf{f} = (f_\lambda(t_1), \dots, f_\lambda(t_n))^T$  o vetor de valores da função  $f$  nos pontos (nós)  $t_1, \dots, t_n$ . A estimativa do *spline* de suavização  $\hat{f}_\lambda$  do vetor, ou os valores ajustados para os dados  $\mathbf{y} = (y_1, \dots, y_n)^T$  são definidos da seguinte forma:

$$\hat{\mathbf{f}} = (\hat{f}_\lambda(t_1), \dots, \hat{f}_\lambda(t_n))^T = \mathbf{S}_\lambda(\mathbf{y}_1, \dots, \mathbf{y}_n)^T, \text{ assim, } \hat{f}_\lambda = \mathbf{S}_\lambda \mathbf{y}.$$

em que,  $\hat{f}_\lambda$  é um spline cúbico natural com nós  $t_1, \dots, t_n$ , para um parâmetro de suavização fixo  $\lambda > 0$  e  $\mathbf{S}_\lambda$  é uma matriz suavização positiva definida (simétrica) que depende de  $\lambda$  e dos nós  $t_1, \dots, t_n$ , mas não de  $\mathbf{y}$ . Mais especificamente,  $\mathbf{S}_\lambda$  é uma matriz de observações transformadas,  $(y_1, \dots, y_n)^T$  em um modelo  $[y_i = f(t_i) + \epsilon_i]$  com  $\beta = 0$  e os valores ajustados  $\hat{\mathbf{y}} = \mathbf{S}_\lambda \mathbf{f}$ , a função  $\hat{f}_\lambda$  é o estimador da função  $f$ , é obtido por spline cúbico que satisfaz a condição  $\hat{f}(t_i) = (\hat{f})_i$ ,  $i = 1, \dots, n$ , para obter uma melhor estimativa sobre a suavização do spline, veja mais detalhes em EUBANK (1999).

A partir do que foi mencionado, a introdução do conceito fundamental de validação cruzada (VC), nada mais é do que desconsiderar uma das observações  $\{x_i, t_i, y_i\}_{i=1}^n$  sucessivamente, para selecionar o parâmetro de suavização  $\lambda$  que está relacionado com o critério de penalização, pois a sua utilização fará com que minimize a soma de quadrados dos resíduos, e para estimar o soma de quadrados dos resíduos para uma função suavizada em  $t_i$  com base nas restantes  $(n - 1)$  observações. A validação cruzada em função de  $\lambda$  é definida da seguinte forma

$$VC(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{f}_\lambda^{(-i)}(t_i)\}^2 \equiv VC(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{f}_\lambda(t_i)}{1 - (\mathbf{H}_\lambda)_{ii}} \right\}^2, \quad (2.11)$$

em que  $\mathbf{H}_\lambda = \mathbf{S}_\lambda + \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T (\mathbf{I} - \mathbf{S}_\lambda)$  é a matriz suavizada,  $\tilde{\mathbf{X}} = (\mathbf{I} - \mathbf{S}_\lambda) \mathbf{X}$  é a matriz dos resíduos,  $\hat{f} = \mathbf{S}_\lambda (\mathbf{y} - \mathbf{x} \hat{\boldsymbol{\beta}})$  é o ajuste (spline suave) para o parâmetro de suavização  $\lambda$  e  $\hat{f}^{(-i)}$  é o ajuste calculado deixando de fora a  $i$ -ésima observação e  $(\mathbf{H}_\lambda)_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz suavizada  $\mathbf{H}_\lambda$ . Mais detalhes sobre o processo em AYDIN (2014), GREEN e SILVERMAN (1993) e WAHBA (1990).

### 2.6.3 Validação cruzada generalizada

Validação cruzada generalizada (VCG) é uma forma modificada da VC que é um método convencional para a escolha do parâmetro de suavização (AYDIN, 2014). Os autores GREEN e SILVERMAN (1993) afirmam que as ligações com regressões clássicas determinaram outra forma com graus de liberdade equivalente que fornecem uma escolha sobre o número efetivo de parâmetros - em algum sentido - que possui todo o necessário para qualquer valor particular do parâmetro de suavização. Pressupõem, que foi ajustada uma curva ou uma função  $f$  para os dados por regressão paramétrica, assumindo que  $f(t)$  tem a seguinte forma  $\sum_{j=1}^k \theta_j f_j(t)$ . Para algumas funções  $f_j$  fixas e ajustando os parâmetros  $\theta_j$  por mínimos quadrados (por exemplo, regressão linear simples, com  $k = 2$ ,  $f_1(t) = 1$  e  $f_2(t) = t$ ). Assumindo que os parâmetros são determinados com base nas observações disponíveis, a matriz  $\mathbf{H}_\lambda$  é uma projeção sobre um espaço de dimensão  $k$ , o número de parâmetros ajustados, e assim seu traço é igual a  $k$ . Assim, o número de graus de liberdade do modelo é igual ao traço de  $\mathbf{H}_\lambda$ , enquanto o número de graus de liberdade para  $n - k$  ruídos é igual a  $tr(\mathbf{I} - \mathbf{H}_\lambda)$ . A definição de VCG, é construída por analogia ao resultado VC; pode ser obtida a partir dos resíduos comuns dividindo pelos fatores  $1 - \mathbf{H}_{\lambda ii}$ . O modelo subjacente da VCG é substituir os fatores  $1 - \mathbf{H}_{\lambda ii}$  na equação (2.11) com o fator  $1 - n^{-1} tr(\mathbf{H}_\lambda)$ . Assim, pela soma de quadrados de resíduos corrigidos e o fator de  $\{1 - n^{-1} tr(\mathbf{H}_\lambda)\}^2$  (mais detalhes em CRAVEN e WAHBA (1978) e WAHBA (1990)), de forma análoga à validação cruzada comum, a função que representa VCG pode ser definida da seguinte forma:

$$VCG(\lambda) = \frac{1}{n} \frac{\sum_{i=1}^n \{y_i - \hat{f}_\lambda(t_i)\}^2}{\{1 - n^{-1} tr(\mathbf{H}_\lambda)\}^2} = \frac{n^{-1} \|(\mathbf{I} - \mathbf{H}_\lambda) \mathbf{y}\|^2}{[n^{-1} tr(\mathbf{I} - \mathbf{H}_\lambda)]^2}. \quad (2.12)$$

### 2.6.4 Spline

Um spline é uma curva polinomial formada por partições, podendo ser separada em algumas categorias, duas delas são: spline de aproximação, na qual essa curva passa por apenas alguns pontos de controle que são chamados de nós (ressalta-se que todos os nós são pontos de controle, mas nem todos pontos de controle são nós) e spline de interpolação, na qual passa por todos os pontos de controle, como consequência são todos nós.

Nesse trabalho serão utilizados métodos de suavização por splines que se encaixam na categoria de splines de aproximação. Assim, o conceito de suavização por spline, para  $f \in W_2^m[a, b] = \{f : f' \text{ é absolutamente contínua e } \int_a^b (f^m)^2 < \infty\}$  o espaço de funções Sobolev, em que desejamos uma estimativa de  $f$  que se ajuste bem aos dados e que ao mesmo tempo tenha um certo grau de suavidade (DIAS, 2001), para um conjunto de pontos que chamado de nós  $a = f_0 < f_1 < \dots < f_m = b$  de modo que  $f(t)$  é uma função polinomial de grau  $n - 1$  em cada subintervalo  $[t_i, t_{i+1}]$ ,  $i = 0, \dots, m - 1$ . Note que um spline de ordem 2 é apenas a interpolação linear comum por partes da forma

$$f(x) = y_i + \frac{x - x_i}{x_{i+1} - x_i}(y_{i+1} - y_i), x \in [x_i, x_{i+1}]. \quad (2.13)$$

#### 2.6.4.1 Spline cúbico

Sejam  $t_1, \dots, t_n$  pertencentes ao intervalo  $[a, b]$ , com  $a < t_1 < t_2 < \dots < t_n < b$ . A função definida em  $[a; b]$ , é um spline cúbico se, e somente se:

- (i) A função  $f$  é um polinômio cúbico nos intervalos  $(a, t_1), (t_1, t_2), (t_2, t_3), \dots, (t_n, b)$ ;
- (ii) As partições polinomiais se encaixam nos pontos de  $t_i$  de tal forma que  $f$  tenha primeira e segunda derivadas contínuas em cada  $t_i$ , denominados de nós e, conseqüentemente sobre a totalidade de  $[a, b]$  (GREEN e SILVERMAN, 1993).

Várias formas equivalentes de descrever um spline cúbico são apresentadas. Uma delas é a condição de atribuir os quatro coeficientes polinomiais de cada partição cúbica, por exemplo, sob a forma

$$f(t) = d_i(t - t_i)^3 + c_i(t - t_i)^2 + b_i(t - t_i) + a_i, \quad (2.14)$$

em que  $t_i \leq t \leq t_{i+1}$ .

Sejam as constantes  $a_i, b_i, c_i, d_i$ , para  $i = 1, \dots, n$ ; definimos  $t_1 = a$  e  $t_n = b$ . As condições de continuidade de  $f$  em suas derivadas de 1ª e 2ª ordem implicam várias relações entre os coeficientes. Por exemplo, a continuidade de  $f$  crescente em  $t_{i+1}$ , para  $i = 1, \dots, n$ , tem a seguinte forma

$$d_i(t_{i+1} - t_i)^3 + c_i(t_{i+1} - t_i)^2 + b_i(t_{i+1} - t_i) + a_i = a_{i+1} \quad (2.15)$$

uma vez que (2.14) e (2.15) são iguais a a curva suave  $f(t_{i+1})$ .

#### 2.6.4.2 Spline cúbico natural

Um spline cúbico, em um intervalo  $[a, b]$ , será um spline cúbico natural (SCN), se as derivadas de 2ª e 3ª ordem são zero nos pontos  $a$  e  $b$ . Estas condições são chamadas as condições de contorno natural (GREEN e SILVERMAN, 1993). Elas implicam que  $d_0 = c_0 = d_n = c_n = 0$ , de modo que a derivada segunda de  $f$  é zero, ou seja,  $f$  é linear nos dois



intervalos extremos  $] -\infty, t_1]$  e  $[t_n, +\infty[$ , mais detalhes veja BOOR e MATHÉMATICIEN (1978).

### 2.6.4.3 B-spline

Todas as técnicas de suavização por splines estão contidas numa classe chamada de B-splines proposto por BOOR (1972), que é formada por bases polinomiais ortogonais. Essa classe é bastante utilizada, principalmente quando a matriz  $\mathbf{X}^T\mathbf{X}$  tem problemas de singularidade e caso seja escolhido um número de nós elevado. Essa classe também é favorável pois requer pouco esforço computacional. Em OKI (2015), representa B-splines da seguinte forma

$$S(x) = \sum_{i=1}^{q+4} \gamma_i B_i(x)$$

$$B_i(x) = \sum_{j=i-4}^i \left[ \frac{(x - t_j)_+^3}{\prod_{m=i-1, m \neq j}^i (t_j - t_m)} \right] \quad (2.16)$$

em que  $i = 1, \dots, q + 4$ ,  $\gamma_i$  são parâmetros a serem estimados, mais detalhes podem ser vistos em OKI (2015), BOOR (1977) e DIERCKX (1993).

### 2.6.5 Método de penalização

Cuidados no ajuste de uma curva devem ser tomados, ou seja, na escolha do número de nós, se for muito grande pode ocorrer o ajuste próximo à interpolação, se forem escolhidos poucos nós a curva estimada talvez não seja a melhor ajustada. Para se obter melhores resultados, foram desenvolvidos critérios envolvendo, por exemplo, a penalização da curva.

Segundo GREEN e SILVERMAN (1993), a ideia básica da abordagem de penalização é estimar a informação de uma curva que sofre oscilações repentinas, em seguida, representar o problema de estimação de forma que torna explícita a necessidade do ajuste entre dois propósitos diferentes nas estimativas da curva: estimar a rugosidade da curva e a regressão por mínimos quadrados penalizados. Nesse trabalho, a forma de estimar a rugosidade de uma curva é dada por  $\int_a^b [f''(t)]^2 dt$ , em que  $f$  pertence ao intervalo  $[a, b]$  e contínua nas derivadas de 1ª e 2ª ordem avaliada no domínio. Outra forma de penalização foi proposta por EILERS e MARX (1996), chamada de P-splines, que é uma combinação de B-splines e penalidades de diferença (sobre os coeficientes estimados). A penalização é realizada por meio de contrastes, é discreta e penaliza os coeficientes diretamente, em vez de penalizar a curva, reduzindo a dimensionalidade do problema. A vantagem dos P-splines é ser muito mais rápido para calcular o traço de  $\mathbf{H}$  (matriz utilizada para estimar os graus de liberdade efetivos), do que outro tipo de suavização (DURBÁN *et al.*, 2008).

### 2.6.6 Algoritmo *backfitting*

O algoritmo *backfitting*, proposto por BREIMAN e FRIEDMAN (1985), é um método iterativo simples utilizado para ajustar os Modelos Aditivos Generalizados (GAM). Na maioria dos casos, é equivalente ao algoritmo baseado no método de Gauss-Seidel. O algoritmo é utilizado para calcular todos os  $f_j$  o que são dados em termos de um spline cúbico suavizado ( $S_j$ ) arbitrário e um gráfico de dispersão. Modelos aditivos é uma classe de modelos de regressão não-paramétrica da seguinte forma:

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \epsilon, \quad (2.17)$$

em que  $X_j$  é a variável explicativa que é controlada de forma não-paramétrica,  $Y$  é a variável resposta,  $\epsilon$  representa o erro que supostamente tem média zero e  $f_j$  representa as funções suavizadas desconhecidas de um único  $X_j$ . O parâmetro constante  $\alpha$  não é identificável, uma vez que pode adicionar ou subtrair quaisquer constantes a cada uma das funções  $f_j$  e ajustar  $\alpha$  conseqüentemente. Geralmente  $\alpha$  é ajustado supondo que  $\sum_{j=1}^p f(X_j) = 0$  para todo  $j$ , ou seja, as funções tem média zero em relação aos dados.

O algoritmo *backfitting* é definido da seguinte forma:

Inicialize  $\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N y_i$ ,  $\hat{f}_j \equiv 0$ ,  $\forall i, j$ .

Intervalo  $j = 1, \dots, p$ .

(i)  $\hat{f}_j \leftarrow S_j[(Y - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k(x_{ik}))_1^N]$  (passo *backfitting*)

(ii)  $\hat{f}_j \leftarrow \hat{f}_j - \frac{1}{N} \sum_{i=1}^N \hat{f}_j(x_{ij})$  (centralização da estimativa média da função)

Em teoria, o passo (ii) no algoritmo não é necessário, pois é necessário que a soma das estimativas da função seja zero. No entanto, devido à problemas numéricos isso pode se tornar um problema na prática. Existe também um procedimento iterativo simples para encontrar a solução, em que  $\hat{\alpha}$  deve ser fixado. Aplica-se um spline cúbico suavizado ( $S_j$ ) em  $\{y_i - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k(x_{ik})\}_1^N$ , como a função de  $x_{ij}$ , para se obter uma nova estimativa  $\hat{f}_j$ . Isto é feito para cada preditor, utilizando as estimativas atuais das outras funções  $\hat{f}_k$  ao calcular  $y_i - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k(x_{ik})$ . O processo continua até que as estimativas  $\hat{f}_j$  estabilizem. Vários autores utilizam a suavização spline cubico, mas pode ser considerados outros métodos de ajuste como: regressão polinomial local, métodos do kernel suavização, mais operadores complexos, tais como superfícies suavizadas para segunda e interações de ordem superior, mais detalhes em HASTIE e TIBSHIRANI (1990).

## 2.7 Modelo parcialmente linear

O modelo parcialmente linear (GREEN e SILVERMAN, 1993) é definido da seguinte forma

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (2.18)$$

em que  $Y_i$  é uma variável aleatória que representa a variável resposta,  $\mathbf{x}_i$  é o vetor que contém os valores de variáveis explicativas,  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos,  $f(t_i)$  é uma função não-paramétrica ( $t_i$  é o valor da variável explicativa que deve ser controlada de maneira não-paramétrica e  $\varepsilon_i$  é uma variável aleatória que representa o erro experimental. O modelo é definido como parcialmente linear, porque inclui apenas uma função não-paramétrica. A equação (2.18) também pode ser denotada na forma matricial da seguinte forma

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \boldsymbol{\varepsilon}, \quad (2.19)$$

em que  $\mathbf{Y}$  é um vetor  $n \times 1$ ,  $\mathbf{X}$  é a matriz planejamento  $n \times p$ ,  $\boldsymbol{\beta}$  é o vetor de  $p$  parâmetros desconhecidos de dimensão  $p \times 1$ ,  $\mathbf{N}$  é a matriz de incidência ( $n \times q$ ), com elementos  $n_{ij} = I(t_i = t_j^0)$ , em que  $I(t_i = t_j^0)$  é a função indicadora e  $t_j^0$ , em que  $j = 1, \dots, q$  são os valores de  $\mathbf{t}$  ordenados e distintos e  $q$  é a quantidade de valores distintos da variável explicativa que é controlada de forma não-paramétrica. A matriz  $\mathbf{N} = \mathbf{I}_n$  quando não há empates, ou seja, quando todos os valores de  $\mathbf{t}$  forem diferentes,  $\mathbf{f} = (f_1, \dots, f_q)^T$  de ordem  $q \times 1$  e  $\boldsymbol{\varepsilon}$  de ordem  $n \times 1$ .

Admitindo que  $f(\cdot)$  pertence ao espaço funcional, respeitando algumas condições estabelecidas, a pretensão é encontrar uma combinação linear de funções desse espaço que seja capaz de explicar o comportamento e assim ajustar a variável explicativa que é controlada de forma não-paramétrica. Utilizando métodos de suavização por *splines*, um critério de penalização e com a subvenção do parâmetro de suavização, serão obtidas as estimativas do componente não-paramétrico  $\mathbf{t}$ , ou seja, para cada ponto  $t_i$  denominado nó, será obtida uma estimativa  $f(t_i)$ . É fundamental escolher um número de nós de forma correta, mas, para isso é necessário tomar cuidado em relação à escolha da quantidade de nós; se for um número muito grande pode causar um ajuste próximo à interpolação dos dados, o que não é conveniente; se o número de nós for muito baixo também pode acarretar um mau ajuste. Geralmente o número de nós é grande, tendendo à interpolação. Em seguida, uma alternativa interessante para suavizar a função é aplicar um critério de penalização (O'SULLIVAN, 1986) sobre a curva  $f(\cdot)$ , critério que será especificado mais adiante na Seção 2.6.5.

### Exemplo da matriz $\mathbf{N}$

A seguir é apresentado um exemplo de como determinar a matriz  $\mathbf{N}$ :

- Seja  $\mathbf{t} = (t_1, \dots, t_9)^T = (10, 5, 3, 5, 10, 6, 3, 5, 2)^T$  o vetor com os valores da variável explicativa que deve ser controlada de forma não-paramétrica.
- Seja  $\mathbf{t}^0 = (t_1^0, t_2^0, t_3^0, t_4^0, t_5^0)^T$  o vetor com os valores de  $\mathbf{t}$  ordenados e distintos, isto é,  $\mathbf{t}^0 = (2, 3, 5, 6, 10)^T$ . Calculando a primeira linha de  $\mathbf{N}$ , com base no vetor  $t^0$ , temos que  $n_{1j} = I(t_1 = t_j^0)$  com  $j = 1, \dots, 5$ , isto é,

$$t_1 = 10, \begin{cases} t_1^0 = 2, \text{então } n_{11} = 0 \\ t_2^0 = 3, \text{então } n_{12} = 0 \\ t_3^0 = 5, \text{então } n_{13} = 0 \\ t_4^0 = 6, \text{então } n_{14} = 0 \\ t_5^0 = 10, \text{então } n_{15} = 1 \end{cases}$$

Após encontrados os elementos de cada linha da matriz  $\mathbf{N}$ , esta pode ser escrita da seguinte forma

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Esta matriz pode ser calculada utilizando o código no software R, que está disponível no Apêndice B.

## 2.8 Modelos parcialmente lineares generalizados

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes. Os modelos parcialmente lineares generalizados (GREEN e YANDELL, 1985), aqui considerados, tem componentes expressos da seguinte forma:

1.  $Y_i \sim FE(\mu_i, \phi)$ , com  $\mu_i = E(Y_i)$ ;
2.  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i)$ ;
3.  $g(\mu_i) = \eta_i$ ,

em que  $\eta_i$  é o preditor linear e  $g(\mu_i)$  é a função de ligação para  $i = 1, \dots, n$ . Um caso particular é  $Y_i \sim N(\mu_i, \phi)$ , desde que o parâmetro  $\mu_i$  satisfaça

$$g(\mu_i) = \mu_i = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i).$$

## Regressão logística parcialmente linear

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes, em que cada  $Y_i \sim \text{Bin}(m_i, \pi_i)$ , com  $\mu_i = E(Y_i) = m_i \pi_i$ . Assumindo a função logito como função de ligação, tem-se,

$$g(\mu_i) = \log \left( \frac{\pi_i}{1 - \pi_i} \right) = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i). \quad (2.20)$$

O modelo (2.20) é chamado modelo de regressão logística parcialmente linear proposto por GREEN e YANDELL (1985). Esse modelo foi aplicado aos dados de um ensaio biológico do agente de extintores de chama realizada nos Estados Unidos no Programa Nacional de Toxicologia (DINSE e LAGAKOS, 1983). Os dados são de ratos, que foram expostos a várias doses do agente. A variável resposta binária  $Y_i$  indica a presença ou ausência de hiperplasia do canal biliar no momento da morte, e as quatro variáveis explicativas foram: logaritmo da dose, peso inicial, posição da gaiola e idade no momento da morte.

## 2.9 Modelos GAMLSS

Modelos Aditivos Generalizados para Posição, Escala e Forma (GAMLSS) foram introduzidos por RIGBY e STASINOPOULOS (2001) e AKANTZILIOTOU *et al.* (2002) como uma forma de superar algumas das limitações associadas aos Modelos Lineares Generalizados (GLM) e aos Modelos Aditivos Generalizados (GAM) (NELDER e WEDDERBURN (1972) e HOWLADER (1989), respectivamente).

No GAMLSS, as suposições da família exponencial de distribuições para a variável resposta ( $Y$ ) são expandidas e substituídas por uma família de distribuições gerais, incluindo distribuições altamente assimétricas ou curtóticas.

A parte sistemática do modelo é expandida permitindo modelar não apenas a média (ou posição), mas outros parâmetros da distribuição de  $y$  como funções paramétricas e também aditivas não-paramétricas de variáveis explicativas ou efeitos aleatórios. A estimativa de máxima verossimilhança (penalizada) é utilizada para ajustar os modelos. Existem dois algoritmos para ajustar os modelos: os algoritmos CG e RS, que são discutidos em detalhes em RIGBY e STASINOPOULOS (2005).

### 2.9.1 Componentes de um modelo GAMLSS

O GAMLSS assume observações independentes  $y_i$  para  $i = 1, 2, \dots, n$ , com função densidade de probabilidade  $f(y_i | \boldsymbol{\theta}^i)$  condicional em  $\boldsymbol{\theta}^i$ , em que  $\boldsymbol{\theta} = (\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_{i2}, \dots, \boldsymbol{\theta}_{ip})$  é um vetor de  $p$  parâmetros, cada um dos quais está relacionado com as variáveis explicativas. Em várias situações práticas o número de parâmetros é no máximo  $p = 4$ . A implementação no software *R* denota estes parâmetros como  $(\mu_i, \phi_i, \nu_i, \tau_i)$ . Os dois primeiros parâmetros  $\mu_i$  e  $\phi_i$  são geralmente caracterizados como parâmetros de posição e escala respectivamente, enquanto os demais parâmetro(s), se houver, são caracterizados como

parâmetros de forma, embora o modelo possa ser aplicado de maneira mais geral aos parâmetros de qualquer distribuição. Seja  $y^T = (y_1, y_2, \dots, y_n)$  de dimensão  $n$  do vetor da variável resposta. Também para  $k = 1, 2, 3, 4$ , sejam  $g_k(\cdot)$  funções de ligação monótonas conhecidas, com o  $k$ -ésimo parâmetro  $\theta_k$  relacionando com as variáveis explicativas dos modelos aditivos semiparamétricos, definidas como:

$$\begin{aligned}
 g_1(\mu) &= \eta_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}(\mathbf{x}_{j1}) \\
 g_2(\phi) &= \eta_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} h_{j2}(\mathbf{x}_{j2}) \\
 g_3(\nu) &= \eta_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} h_{j3}(\mathbf{x}_{j3}) \\
 g_4(\tau) &= \eta_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}(\mathbf{x}_{j4}),
 \end{aligned} \tag{2.21}$$

em que  $\boldsymbol{\mu}, \phi, \boldsymbol{\nu}, \boldsymbol{\tau}$  e  $\boldsymbol{\eta}_k$  e  $\mathbf{x}_{jk}$ , para  $j = 1, 2, \dots, J_k$  e  $k = 1, 2, 3, 4$ , são vetores de comprimento  $n$ . A função  $h_{jk}$  é uma função aditiva não-paramétrica da variável explicativa  $\mathbf{X}_{jk}$ , avaliada em  $\mathbf{x}_{jk}$ . Os vetores explicativos  $\mathbf{x}_{jk}$  são assumidos fixos e conhecidos. Também  $\mathbf{X}_k$ , para  $k = 1, 2, 3, 4$ , são fixos na matriz de planejamento enquanto  $\boldsymbol{\beta}_k$  são os vetores de parâmetros da regressão. Note que, em aplicações típicas, um modelo constante ou outro modelo simples é frequentemente adequado para cada um dos dois parâmetros de forma  $\nu$  e  $\tau$ ). O modelo (2.21) é chamado de modelo GAMLSS.

### 2.9.2 Algoritmos CG e RS

O método de máxima verossimilhança (penalizada) é utilizada para ajustar os modelos parcialmente lineares, semiparamétricos e GAMLSS. Há dois algoritmos básicos que serão aplicados para ajustar o modelo, tais como o algoritmo *backfitting*. O primeiro, o algoritmo CG, proposto por COLE e GREEN (1992) é um algoritmo em que é utilizado as derivadas primeiras e os valores esperados da derivada segunda e derivadas cruzadas da função de verossimilhança em relação  $\boldsymbol{\theta} = (\mu, \phi, \nu)^T$ . No entanto, para muitas funções densidade de probabilidade  $f(y|\boldsymbol{\theta})$  os parâmetros  $\boldsymbol{\theta}$  são informações ortogonais (uma vez que os valores esperados das derivadas cruzadas da função de verossimilhança são nulas), isto é, para modelos com parâmetros de posição e escala ou modelos com parâmetro de dispersão, ou ainda utilizando modelos que envolvem os três parâmetros. Neste caso, o segundo algoritmo (RS) não utiliza os valores das derivadas cruzadas, é uma generalização do algoritmo proposto por RIGBY e STASINOPOULOS (1996) para ajuste da média e dispersão de modelos.

### 2.9.3 Pacote gamlss

O pacote `gamlss`, proposto por STASINOPOULOS *et al.* (2006), permite ao usuário modelar a distribuição da variável resposta utilizando uma variedade de famílias de distribuições com um, dois, três e quatro parâmetros. As distribuições implementadas atualmente podem ser encontradas em `gamlss.family`. Outras distribuições podem ser facilmente adicionadas. Na implementação atual do GAMLSS foram adotados vários termos aditivos, incluindo splines de regressão, splines de suavização, splines penalizados, dentre outros. Outros termos aditivos podem ser facilmente adicionados. Uma opção para ajustar os modelos parcialmente lineares é utilizar o pacote `gamlss` que está no software R (TEAM, 2015). O argumento `method` recebe o algoritmo CG propostos por COLE e GREEN (1992), ou RS que foi apresentado por RIGBY e STASINOPOULOS (1996) ou ainda a mistura de ambos os algoritmos. Cada um deles podem ser especificados da seguinte forma:

- `RS ()`: O método padrão é o algoritmo RS, que não requer valores iniciais precisos nos parâmetros  $\mu, \phi, \nu, \tau$  para garantir a convergência (os valores iniciais padrão, frequentemente constantes, são geralmente adequados). Esse método é mais rápido para grandes conjuntos de dados.
- `CG ()`: O algoritmo CG, que pode ser melhor para distribuições com estimativas de parâmetros potencialmente altamente correlacionados.
- `mixed()`: Esse é uma mistura dos algoritmos RS e CG e começa com o algoritmo RS e termina com o CG. `method = mixed(n1, n2)` é utilizado para a distribuição que está sendo considerada, fazendo com que acelere a convergência. Esse argumento instrui o `gamlss()` a usar o algoritmo `RS()` para as  $n_1$  primeiras iterações (para estabilizar o processo de ajuste) e então (se ainda não convergiu) mudar para o algoritmo `CG()` e continuar com até  $n_2$  iterações. `CG()` converge mais rápido perto do máximo quando os parâmetros da distribuição são altamente não ortogonais no máximo. Mais detalhes dos algoritmos RS e CG em RIGBY e STASINOPOULOS (2005).

### 2.10 Análise residual

A análise de resíduos é uma técnica importante, pois serve para identificar possíveis observações discrepantes. Além disso, detecta alguns desvios das suposições do ajuste que podem comprometer os resultados inferenciais. Dentre os resíduos considerados na literatura, enfatizam-se os resíduos quantílicos. DUNN e SMYTH (1996) definem os resíduos quantílicos da seguinte forma:

Seja a função densidade de probabilidade contínua  $f(y_i|\boldsymbol{\theta})$ , com  $i = 1, \dots, n$ , então os resíduos quantílicos normalizados ajustados (DUNN e SMYTH, 1996) são definidos como:

$$\hat{r}_i = \Phi^{-1}(\hat{u}_i), \quad (2.22)$$

em que  $\Phi^{-1}(\cdot)$  é a inversa da função densidade de probabilidade acumulada e  $\hat{u}_i$  são os resíduos quantílicos. Esses resíduos podem ser utilizado em modelos, tais como, modelos lineares normais, modelos lineares generalizados, modelos semiparamétricos, modelos lineares generalizados simétricos, entre outros. Se  $y$  é uma observação de uma variável resposta contínua, então,  $u = F(y|\theta)$  e  $\hat{u} = F(y|\hat{\theta})$  é o modelo ajustado de funções densidade de probabilidade acumulada. Se o modelo é especificado corretamente,  $u$  tem a distribuição Uniforme entre zero e um. Isso é chamado de transformação integral de probabilidade. Atualmente o pacote `gamlss` (STASINOPOULOS *et al.*, 2006), disponível no software livre R (TEAM, 2015), realiza esse tipo de análise. A função `plot()` além de gerar sumário dos resíduos quantílicos também gera quatro gráficos, em que um desses é o *QQplot* que está no manual do pacote `gamlss` (STASINOPOULOS *et al.*, 2006).

## 2.11 Influência local

A análise de diagnóstico é utilizada para detectar possíveis observações influentes. Este procedimento é conhecido como análise de sensibilidade. O método de influência local é utilizado com o objetivo de identificar possíveis observações extremas. Isto ocorre devido os modelos de regressão serem sensíveis aos pressupostos exigidos para o ajuste. Este método foi proposto por COOK (1986) e tem como objetivo avaliar a qualidade do ajuste, utilizando uma medida de influência apropriada, a robustez das estimativas fornecidas pelo modelo ajustado.

Dentre as medidas de influência mais aplicadas, destaca-se a que avalia o efeito das perturbações nas estimativas fornecidas pelo modelo através do afastamento da verossimilhança dada por

$$LD(\omega) = 2 | L(\hat{\theta}) - L(\hat{\theta}_\omega) |, \quad (2.23)$$

em que  $LD(\omega) \geq 0$ ,  $\omega$  é o vetor de perturbações,  $\hat{\theta}$  e  $\hat{\theta}_\omega$  são as estimativas de máxima verossimilhança do modelo não perturbado  $L(\theta)$  e perturbado  $L(\theta|\omega)$  por  $\omega$ , respectivamente. Temos que  $LD(\omega_0) = 0$ . A equação (2.23) é utilizada para avaliar a influência de variáveis,  $\omega \in \Omega$ . A curvatura normal de uma superfície  $\alpha(\omega)$ , resulta da geometria diferencial (considerando-se, por exemplo KREYSZIG (1991), p.35) que  $C_i = |\ddot{L}D\{\omega(a)\}|_{a=0} = |\partial^2 LD\{\omega(a)\}/\partial a^2|_{a=0}$ , mais informações estão disponíveis em BATES e WATTS (1980). Por conveniência (q+1)-dimensional, considera-se a superfície geométrica  $\alpha(\omega)$  definido por

$$\alpha(\omega) = \begin{pmatrix} \omega \\ LD(\omega) \end{pmatrix} \quad (2.24)$$



quando  $\omega$  varia em  $\Omega$ . Essa superfície é chamada gráfico de influência, para reproduzir curvaturas normais. Curvaturas em pontos diferentes de  $\omega_0$  podem ser importantes na avaliação do comportamento global de um gráfico de influência. O interesse principal é especificar uma descrição de como a superfície  $\alpha(\omega)$  desloca de seu plano tangente em  $\omega_0$  que é denominado linha projetada. Tal descrição pode ser obtida mediante análise da curvatura de algumas curvas na superfície que passam por  $\alpha(\omega_0)$ . Visualizadas em  $R^3$ , as determinadas curvas são as secções normais formadas pela intersecção da superfície com planos que contêm o vetor ortogonal ao plano tangente em  $\omega_0$  (STOKER, 1969). As curvaturas dessas secções normais são chamadas curvaturas normais e são utilizadas para caracterizar o comportamento de um gráfico de influência em torno de  $\omega_0$ . Como  $LD(\omega)$  atinge um mínimo local em  $\omega_0$ , uma simples representação de secções normais é possível. Para construir uma secção normal, considere uma reta em  $\Omega$  passando por  $\omega_0$ , que pode ser representada por

$$\omega(a) = \omega_0 + al \quad (2.25)$$

em que  $a \in R$  e  $l$  é um vetor unitário fixo não nulo de comprimento em  $R^q$ . Esta linha gera uma linha de elevação no gráfico de influência  $\alpha(\omega_0)$  passando por  $\omega_0$ . Cada direção  $l$  especifica a linha de elevação e cada linha de elevação corresponde a uma secção normal.

A curvatura  $C_l$  pode ser definida como

$$C_l = 2|l^T \ddot{\mathbf{F}} l| \quad (2.26)$$

em que  $\|l\| = 1$  e  $\ddot{\mathbf{F}}$  é a matriz  $q \times q$  com elementos  $\frac{\partial^2 L(\hat{\theta}_\omega)}{\partial \omega_k \partial \omega_j}$ ,  $j, k = 1, \dots, q$ . Esta é a forma básica para curvaturas normais que serão utilizadas. Para (2.26) ser útil, devemos ter uma maneira direta de avaliar  $\ddot{\mathbf{F}}$ . Utilizando a regra de cadeia para a diferenciação,  $\ddot{\mathbf{F}}$  pode ser expressa como

$$\ddot{\mathbf{F}} = \mathbf{J}^T \ddot{\mathbf{L}} \mathbf{J} \quad (2.27)$$

em que  $-\ddot{\mathbf{L}}$  é a informação observada para o modelo proposto ( $\omega = \omega_0$ ) e  $\mathbf{J}$  é a matriz  $p \times q$  com elementos  $\frac{\partial \hat{\theta}_{i\omega}}{\partial \omega_j}$ ,  $i = 1, 2, \dots, p$ ,  $j = 1, 2, \dots, q$ , em que  $\hat{\theta}_{i\omega}$  é o  $i$ -ésimo componente de  $\theta_\omega$ . Em seguida, para avaliar  $\mathbf{J}$ , é utilizado o fato de que

$$\left. \frac{\partial L(\theta|\omega)}{\partial \theta_j} \right|_{\theta=\hat{\theta}} = 0 \quad (2.28)$$

para  $j = 1, 2, \dots, p$  e para todo  $\omega$  em  $\Omega$ . Diferenciando ambos os lados de (2.28) em relação a  $\omega$  e avaliando em  $\omega_0$ , segue que

$$\mathbf{J} = -(\ddot{\mathbf{L}})^{-1} \mathbf{\Delta} \quad (2.29)$$

em que  $\mathbf{\Delta}$  é a matriz  $p \times q$  com elementos

$$\Delta_{ij} = \frac{\partial^2 L(\theta|\omega)}{\partial \theta_i \partial \omega_j}$$

avaliado em  $\theta = \hat{\theta}$  e  $\omega = \omega_0$ ,  $i = 1, 2, \dots, p$ ,  $j = 1, 2, \dots, q$ . Substituindo (2.29) em (2.27) obtemos

$$\ddot{\mathbf{F}} = \mathbf{\Delta}^T (\ddot{\mathbf{L}})^{-1} \mathbf{\Delta} \quad (2.30)$$

e portanto de (2.26)

$$C_l = 2|\mathbf{l}^T \mathbf{\Delta}^T (\ddot{\mathbf{L}})^{-1} \mathbf{\Delta} \mathbf{l}|. \quad (2.31)$$

em que  $\|\mathbf{l}\| = 1$ , é um vetor de 1's, considerando a partição

$$\ddot{\mathbf{L}} = \begin{pmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{pmatrix},$$

o interesse está em determinar um subvetor  $\boldsymbol{\theta}_1$  de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ . Assim a curvatura normal na direção do vetor unitário  $l$  é definido da forma

$$C_l(\boldsymbol{\theta}_1) = 2|\mathbf{l}^T \mathbf{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_1) \mathbf{\Delta} \mathbf{l}| \quad (2.32)$$

em que

$$\mathbf{B}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{L}_{22}^{-1} \end{pmatrix}$$

com  $-\ddot{\mathbf{L}}_{\hat{\boldsymbol{\theta}}_2 \hat{\boldsymbol{\theta}}_2}^{-1}$  denotando a informação de Fisher observada para  $\boldsymbol{\theta}_2$ . O gráfico do maior autovetor de  $\mathbf{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_1) \mathbf{\Delta}$  versus a ordem das observações pode evidenciar os pontos com maior influência local em  $\hat{\boldsymbol{\theta}}_1$ .

## 2.12 Considerações

Nesse capítulo foi realizada uma síntese sobre alguns dos principais modelos no âmbito de família exponencial, família de distribuições simétricas, modelos lineares generalizados, modelos lineares generalizados simétricos, alguns conceitos associados a regressão não-paramétrica, modelo parcialmente linear, modelo parcialmente linear generalizado, foi especificado o GAMLSS, resíduo quantílico e por fim foi feita uma revisão sobre influência local.



## 3 MATERIAL E MÉTODOS

### 3.1 Introdução

Levando em consideração os modelos que foram exibidos no Capítulo 2, dar-se-á início ao capítulo fundamental, pois trata-se de um dos propósitos deste trabalho definir o modelo linear parcial generalizado simétrico (MLPGS). Este modelo é flexível para a variável resposta com distribuição simétrica, associada a um preditor linear, em que pelo menos uma covariável se relaciona com  $Y$  de forma paramétrica e somente uma covariável que se relaciona com  $Y$  de forma não-paramétrica, ou seja, não há uma função não linear conhecida que explique a relação entre a variável resposta e a covariável. Esta covariável, denotada por  $t$ , (geralmente é considerada como tempo, umidade ou temperatura), é acrescentada no modelo para diminuir o resíduo, funciona de modo similar ao bloco num delineamento casualizado em blocos. Também no decorrer do Capítulo será exposta a função de verossimilhança penalizada, vetor escore penalizado, matriz hessiana penalizada e terminando o capítulo com influência local.

### 3.2 Modelo Linear Parcial Generalizado Simétrico

Sejam  $n$  variáveis aleatórias independentes  $Y_1, \dots, Y_n$ , cada uma com função de densidade de probabilidade expressa por

$$f(y_i; \mu_i, \phi) = \frac{1}{\sqrt{\phi}} h(u_i),$$

em que  $u = \frac{(y - \mu)^2}{\phi}$ ,  $\mu \in \mathbb{R}$ ,  $\phi > 0$  e  $h(\cdot)$  é uma função definida positiva em  $\mathbb{R}^+$ .

#### Componentes de um Modelo Linear Parcial Generalizado Simétrico

1.  $Y_i \sim S(\mu_i, \phi, h)$  com  $\mu_i = E(Y_i)$ ;
2.  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(t_i)$ ;
3.  $g(\mu_i) = \eta_i$ .

O modelo definido acima é chamado Modelo Linear Parcial Generalizado Simétrico, em que  $\eta_i$  é o preditor linear,  $f(t_i)$  é uma função arbitrária que depende de da covariável  $t_i$  que é controlada de forma não paramétrica e  $g(\mu_i)$  é a função de ligação.

### 3.3 Função Verossimilhança Penalizada

O logaritmo da função de verossimilhança penalizada do modelo linear parcial generalizado simétrico é dado por:

$$L_p(\boldsymbol{\theta}, \lambda) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log h(u_i) - \frac{\lambda}{2} \int_a^b [f''(t_i)]^2 dt, \quad (3.1)$$

em que  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T$ ,  $\lambda > 0$  é o parâmetro de suavização que controla a qualidade do ajuste. O valor deste parâmetro, não pode ser muito alto nem muito baixo, pois se  $\lambda \rightarrow 0$  a terceira parcela de (3.1) tem pouca intervenção e antepõe o ajuste por interpolação da curva, se  $\lambda \rightarrow \infty$  se antepõe a suavidade da curva  $f$ , com a ressalva de  $f'' = 0$  com a intenção da curva se linear, devido a isso é importante a escolha adequada de  $\lambda$ , se  $\lambda$  minimizar o escore  $VCG(\lambda)$ , será esse o melhor parâmetro de suavização. Para facilitar os cálculos, a integral que esta na equação (3.1) pode ser da forma

$$\int_a^b [f''(t_i)]^2 dt = \mathbf{f}^T \mathbf{K} \mathbf{f}, \text{ para } a \leq t \leq b, \quad (3.2)$$

em que a matriz positiva definida  $\mathbf{K}$  ( $q \times q$ ) é definida como

$$\mathbf{K} = \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T, \quad (3.3)$$

em que  $\mathbf{Q}$  é uma matriz tridiagonal de ordem  $q \times (q - 2)$  e seja  $d_i$  a distância entre dois nós subsequentes e distintos  $i$  e  $i + 1$ , isto é,  $d_i = t_{i+1}^0 - t_i^0$ ,  $i = 1, \dots, q - 1$ . Os elementos de  $\mathbf{Q}$  são definidos como:

$$\begin{aligned} \delta_{j-1,j} &= d_{j-1}^{-1}, & \delta_{jj} &= -d_{j-1}^{-1} - d_j^{-1}, & \delta_{j+1,j} &= d_j^{-1} & j = 2, \dots, q - 1, \\ \delta_{ij} &= 0 & \text{para } |i - j| &\geq 2, & i = 1, \dots, q & \text{ e } j = 2, \dots, q - 1. \end{aligned}$$

A matriz  $\mathbf{R}$  de ordem  $(q-2) \times (q-2)$ , em que seus elementos  $r_{ij}$  com  $i = 2, \dots, q-1$  e  $j = 2, \dots, q-1$ , são determinados por:

$$\begin{aligned} r_{ii} &= \frac{1}{3}(d_{i-1} + d_i) & i = 2, \dots, q - 1, \\ r_{i,i+1} &= r_{i+1,i} = \frac{1}{6}d_i & \text{com } i = 2, \dots, q - 2 \text{ e} \\ r_{ij} &= 0 & \text{para } |i - j| \geq 2, \end{aligned}$$

mais detalhes em GREEN e SILVERMAN (1993).

É importante ressaltar que está sendo utilizados splines cúbicos em (3.1) e a solução de (3.2) existe se,  $f(t)$  é um spline cúbico em que os nós são os valores distintos

de  $t_1, \dots, t_n$ , ordenados e podem ser denotados por  $t_1^0 < t_2^0 < \dots < t_q^0$ . Finalmente, com base em (3.1) e (3.2) o logaritmo da função de verossimilhança penalizada pode ser expresso da forma:

$$L_p(\boldsymbol{\theta}, \lambda) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log h(u_i) - \frac{\lambda}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}. \quad (3.4)$$

### 3.4 Função escore penalizada

Derivando o logaritmo da função verossimilhança dada pela equação (3.4) em relação  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{f}^T, \phi)^T$ , obtém-se a função escore penalizada denotada por  $\mathbf{U}^p(\boldsymbol{\theta})$ , isto é,

$$\mathbf{U}^p(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{U}_{\beta}^p(\boldsymbol{\theta}) \\ \mathbf{U}_f^p(\boldsymbol{\theta}) \\ U_{\phi}^p(\boldsymbol{\theta}) \end{pmatrix},$$

cujos elementos são definidos a seguir (detalhes no Apêndice A)

$$\begin{aligned} \mathbf{U}_{\beta}^p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}} = \frac{1}{\phi} \sum_{i=1}^n v_i \dot{a}_i (y_i - \mu_i) \mathbf{x}_i, \\ \mathbf{U}_f^p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \mathbf{f}} = \frac{1}{\phi} \sum_{i=1}^n v_i \dot{a}_i (y_i - \mu_i) \mathbf{n}_i - \lambda \mathbf{K} \mathbf{f}, \\ U_{\phi}^p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \phi} = -\frac{n}{2\phi} + \frac{1}{2\phi^2} \sum_{i=1}^n v_i (y_i - \mu_i)^2, \end{aligned}$$

em que  $v_i = -2W_h(u_i)$ ,  $W_h(u_i) = \frac{d \log(h(u_i))}{du_i} = \frac{h'(u_i)}{h(u_i)}$  e  $\dot{a}_i = \frac{d\mu_i}{d\eta_i}$ . Para mais detalhes, ver a Tabela 3.1. Matricialmente, a função escore penalizada pode ser expressa por

$$\begin{aligned} \mathbf{U}_{\beta}^p(\boldsymbol{\theta}) &= \frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) (\mathbf{y} - \boldsymbol{\mu}), \\ \mathbf{U}_f^p(\boldsymbol{\theta}) &= \frac{1}{\phi} \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) (\mathbf{y} - \boldsymbol{\mu}) - \lambda \mathbf{K} \mathbf{f}, \\ U_{\phi}^p(\boldsymbol{\theta}) &= -\frac{n}{2\phi} + \frac{1}{2\phi^2} (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{v}) (\mathbf{y} - \boldsymbol{\mu}), \end{aligned}$$

em que  $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1, \dots, v_n\}$ ,  $\mathbf{D}(\boldsymbol{\xi}) = \text{diag}\{\dot{a}_1, \dots, \dot{a}_n\}$ ,  $\mathbf{y} = (y_1, \dots, y_n)^T$  e  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ .

### 3.5 Matriz Hessiana penalizada

A seguir, serão apresentados os elementos da matriz Hessiana penalizada denotada por  $\mathbf{L}^p(\boldsymbol{\theta})$ , isto é,

$$\mathbf{L}^p(\boldsymbol{\theta}) = \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{pmatrix} \mathbf{L}_{\beta\beta}^p(\boldsymbol{\theta}) & \mathbf{L}_{\beta f}^p(\boldsymbol{\theta}) & \mathbf{L}_{\beta\phi}^p(\boldsymbol{\theta}) \\ \mathbf{L}_{f\beta}^p(\boldsymbol{\theta}) & \mathbf{L}_{ff}^p(\boldsymbol{\theta}) & \mathbf{L}_{f\phi}^p(\boldsymbol{\theta}) \\ \mathbf{L}_{\phi\beta}^p(\boldsymbol{\theta}) & \mathbf{L}_{\phi f}^p(\boldsymbol{\theta}) & \ddot{L}_{\phi\phi}^p(\boldsymbol{\theta}) \end{pmatrix}$$

**Tabela 3.1.** Expressões  $W_h(u)$ ,  $W'_h(u)$  e  $v$  de algumas distribuições simétricas.

Distribuição	$W_h(u)$	$W'_h(u)$	$v$
Normal	$-\frac{1}{2}$	0	1
Hiperbólica	$-\frac{\nu}{2\sqrt{1+u}}$	$\frac{\nu\sqrt{1+u}}{4(1+u)^2}$	$\frac{\nu}{\sqrt{1+u}}$
Exponencial Potência	$-\frac{u^{-\frac{\nu}{(1+\nu)}}}{2(1+\nu)}$	$\frac{\nu u^{-\frac{(2\nu+1)}{(1+\nu)}}}{2(1+\nu)^2}$	$\frac{u^{-\frac{\nu}{(1+\nu)}}}{(1+\nu)}$
t-Student	$-\frac{(1+\nu)}{2(\nu+u)}$	$\frac{(1+\nu)}{2(\nu+u)^2}$	$\frac{(1+\nu)}{(\nu+u)}$

cujos elementos são (detalhes no Apêndice A)

$$\begin{aligned} \mathbf{L}_{\beta\beta}^p(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\frac{1}{\phi} \sum_{i=1}^n [a_i \dot{a}_i^2 - v_i \ddot{a}_i (y_i - \mu_i)] \mathbf{x}_i \mathbf{x}_i^T, \\ \mathbf{L}_{ff}^p(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \mathbf{f} \partial \mathbf{f}^T} = -\frac{1}{\phi} \sum_{i=1}^n [a_i \dot{a}_i^2 - v_i \ddot{a}_i (y_i - \mu_i)] \mathbf{n}_i \mathbf{n}_i^T - \lambda \mathbf{K}, \\ \ddot{L}_{\phi\phi}^p(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \phi \partial \phi} = \frac{n}{2\phi^2} + \frac{1}{\phi^2} \sum_{i=1}^n [c_i u_i^2 - v_i u_i], \\ \mathbf{L}_{\beta\phi}^p(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \phi} = \frac{2}{\phi^2} \sum_{i=1}^n b_i (y_i - \mu_i) \dot{a}_i \mathbf{x}_i, \\ \mathbf{L}_{f\phi}^p(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \mathbf{f} \partial \phi} = \frac{2}{\phi^2} \sum_{i=1}^n b_i (y_i - \mu_i) \dot{a}_i \mathbf{n}_i, \\ \mathbf{L}_{\beta f}^p(\boldsymbol{\theta}) &= \frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta} \partial \mathbf{f}^T} = -\frac{1}{\phi} \sum_{i=1}^n [a_i \dot{a}_i^2 - v_i \ddot{a}_i (y_i - \mu_i)] \mathbf{x}_i \mathbf{n}_i^T, \end{aligned}$$

em que  $a_i = [v_i - 4W'_h(u_i)u_i]$ ,  $u_i = \frac{(y_i - \mu_i)^2}{\phi}$ ,  $c_i = W'_h(u_i) = \frac{dW_h(u_i)}{du_i}$ ,  $\ddot{a}_i = \frac{d^2 \mu_i}{d\eta_i^2}$  e  $b_i = [W_h(u_i) + W'_h(u_i)u_i]$ . Para mais detalhes veja a Tabela 3.1. Matricialmente, os elementos de  $\mathbf{L}^P(\boldsymbol{\theta})$ , podem ser escritos como

$$\begin{aligned} \mathbf{L}_{\beta\beta}^p(\boldsymbol{\theta}) &= -\frac{1}{\phi} \mathbf{X}^T [\mathbf{D}(\mathbf{a})\mathbf{D}(\boldsymbol{\xi})^2 - \mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\rho})\mathbf{D}(\mathbf{y} - \boldsymbol{\mu})] \mathbf{X}, \\ \mathbf{L}_{ff}^p(\boldsymbol{\theta}) &= -\frac{1}{\phi} \mathbf{N}^T [\mathbf{D}(\mathbf{a})\mathbf{D}(\boldsymbol{\xi})^2 - \mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\rho})\mathbf{D}(\mathbf{y} - \boldsymbol{\mu})] \mathbf{N} - \lambda \mathbf{K}, \\ \ddot{L}_{\phi\phi}^p(\boldsymbol{\theta}) &= \frac{n}{2\phi^2} + \frac{1}{\phi^2} \left[ \mathbf{u}^T \mathbf{D}(\mathbf{c}) \mathbf{u} - \frac{1}{\phi} (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{v}) (\mathbf{y} - \boldsymbol{\mu}) \right], \\ \mathbf{L}_{\beta\phi}^p(\boldsymbol{\theta}) &= \frac{2}{\phi^2} \mathbf{X}^T \mathbf{b} (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\xi}, \\ \mathbf{L}_{f\phi}^p(\boldsymbol{\theta}) &= \frac{2}{\phi^2} \mathbf{N}^T \mathbf{b} (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\xi}, \\ \mathbf{L}_{\beta f}^p(\boldsymbol{\theta}) &= -\frac{1}{\phi} \mathbf{X}^T [\mathbf{D}(\mathbf{a})\mathbf{D}(\boldsymbol{\xi})^2 - \mathbf{D}(\mathbf{v})\mathbf{D}(\boldsymbol{\rho})\mathbf{D}(\mathbf{y} - \boldsymbol{\mu})] \mathbf{N}, \end{aligned}$$

em que  $\mathbf{u} = (u_1, \dots, u_n)^T$ ,  $\mathbf{b} = (b_1, \dots, b_n)^T$ ,  $\boldsymbol{\xi} = (\dot{a}_1, \dots, \dot{a}_n)^T$ ,  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ ,  $\mathbf{D}(\mathbf{a}) = \text{diag}\{a_1, \dots, a_n\}$ ,  $\mathbf{D}(\mathbf{c}) = \text{diag}\{c_1, \dots, c_n\}$ ,  $\mathbf{D}(\boldsymbol{\rho}) = \text{diag}\{\ddot{a}_1, \dots, \ddot{a}_n\}$ , Para  $g(\mu_i) = \mu_i = \eta_i$ , então  $\boldsymbol{\xi} = (1, \dots, 1)^T$ , assim  $\mathbf{D}(\boldsymbol{\xi}) = \mathbf{I}_n$ ,  $\boldsymbol{\rho} = (0, \dots, 0)^T$  e portanto  $\mathbf{D}(\boldsymbol{\rho}) = \mathbf{0}_n$ .

### 3.6 Processo iterativo para o MLPGS

Existem várias sugestões de obter os estimativas dos parâmetros no contexto de modelos parcialmente lineares, uma das quais envolve a utilização do pacote `gamlss` (RIGBY e STASINOPOULOS, 2005) disponível no software R (TEAM, 2015), em que as estimativas que maximizam logaritmo da função verossimilhança penalizada são obtidas com a utilização do escore de Fisher. Assim, o objetivo é determinar as estimativas dos parâmetros que maximizam sua função de verossimilhança penalizada, isto é, os valores que maximizam o logaritmo da função verossimilhança penalizada. Sobretudo, a estimativa ( $\hat{\boldsymbol{\theta}}$ ) obtida mediante à verossimilhança penalizada deve satisfazer a desigualdade

$$L_p(\hat{\boldsymbol{\theta}}, \lambda) \geq \sup_{\boldsymbol{\theta} \in \Theta} L_p(\boldsymbol{\theta}, \lambda).$$

Considerando o vetor escore penalizado  $\mathbf{U}_{\boldsymbol{\theta}}^p = \begin{pmatrix} \mathbf{U}_{\beta}^p(\boldsymbol{\theta}) \\ \mathbf{U}_f^p(\boldsymbol{\theta}) \\ U_{\phi}^p(\boldsymbol{\theta}) \end{pmatrix}$ , adaptando o algoritmo de *backfitting* e o processo iterativo proposto por IBACACHE-PULGAR e PAULA (2011), é definido um procedimento prático para determinar as estimativas de  $\boldsymbol{\theta}$  para MPLGS em três etapas, o que equivale a resolução das seguintes equações

$$\mathbf{U}_{\beta}^p(\boldsymbol{\theta}) = 0, \mathbf{U}_f^p(\boldsymbol{\theta}) = 0 \text{ e } U_{\phi}^p(\boldsymbol{\theta}) = 0.$$

Tem-se que  $g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f}$  e a partir de  $\mathbf{U}_{\boldsymbol{\theta}}^p$ , e obtêm as seguintes equações:

$$\begin{aligned} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) \mathbf{X} \boldsymbol{\beta} &= \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) (\mathbf{y} - \mathbf{N}\mathbf{f}), \\ (\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) \mathbf{N} + \lambda \phi \mathbf{K}) \mathbf{f} &= \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \\ (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{v}) (\mathbf{y} - \boldsymbol{\mu}) &= n\phi. \end{aligned}$$

Portanto, a partir do seguinte algoritmo *backfitting* com matriz de peso dado por  $\mathbf{D}(\mathbf{v})$  tem-se

$$\boldsymbol{\beta}^{(r+1)} = (\mathbf{X}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)}) (\mathbf{y} - \mathbf{N}\mathbf{f}^{(r)}), \quad (3.5)$$

$$\mathbf{f}^{(r+1)} = (\mathbf{N}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)}) \mathbf{N} + \lambda \phi^{(r)} \mathbf{K})^{-1} \mathbf{N}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(r+1)}), \quad (3.6)$$

$$\phi^{(r+1)} = \frac{1}{\phi} (\mathbf{y} - \boldsymbol{\mu}^{(r)})^T \mathbf{D}(\mathbf{v}^{(r)}) (\mathbf{y} - \boldsymbol{\mu}^{(r)}), \quad (3.7)$$



para  $r = 1, \dots$  em que  $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1, \dots, v_n\}$ ,  $\mathbf{D}(\boldsymbol{\xi}) = \text{diag}\{\dot{a}_1, \dots, \dot{a}_n\}$ ,  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$  e  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ , é necessário iniciar o processo iterativo (3.5) - (3.7) com valores iniciais  $\boldsymbol{\beta}^{(0)}$ ,  $\mathbf{f}^{(0)}$  e  $\phi^{(0)}$ , por exemplo, as estimativas a partir do modelo normal com função de ligação identidade. Se as condições do Teorema 4.1 (GREEN e SILVERMAN, 1993) forem satisfeitas, garantir-se-á solução única e correspondente limite do processo iterativo (3.5) - (3.7). Estas condições significam que a matriz de pesos  $\mathbf{D}(\mathbf{v})$  deve ser positiva definida, que é verificada a partir de  $v_i > 0$ , para todo  $i = 1, \dots, n$  e a matriz  $(\mathbf{N}, \mathbf{NT})$  tem posto completo, em que  $\mathbf{T} = (\mathbf{1}, \mathbf{t}^0)^T$ ,  $\mathbf{t}^0 = (t_1^0, \dots, t_1^q)^T$  e  $\mathbf{1}$  é um vetor de 1's de ordem  $q \times 1$ .

Utilizando o algoritmo *backfitting* e o número de graus de liberdade efetivos, RELVAS (2013) define a equação em função do parâmetro de suavização da seguinte forma:

$$df(\lambda) = \text{traço}(\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{N} + \phi \lambda \mathbf{K})^{-1} (\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{N}).$$

Adaptando para MLPGS a equação é da forma:

$$df(\lambda) = \text{traço}(\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) \mathbf{N} + \phi \lambda \mathbf{K})^{-1} (\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) \mathbf{N})$$

### 3.7 Influência local

Após ajustar o modelo de regressão definido na Seção (3.2), recomenda-se utilizar a análise de diagnóstico para detectar observações influentes com base no método de influência local proposto por COOK (1986).

Considerando o afastamento da verossimilhança dada por  $LD(\boldsymbol{\omega}) = 2\{L_p(\hat{\boldsymbol{\theta}}) - L_p(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ , em que  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  são as estimativas de máxima verossimilhança do modelo proposto,  $L_p(\hat{\boldsymbol{\theta}})$  que é o logaritmo da função verossimilhança penalizada na estimativa máxima verossimilhança quando o modelo proposto é  $L_p(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})$ , em que  $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  denota as EMV do modelo perturbado.

A curvatura normal para  $\boldsymbol{\theta}$  na direção unitária  $\mathbf{l}$ , é dada por:

$$C_1(\boldsymbol{\theta}) = 2|\mathbf{l}^T \boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta} \mathbf{l}|,$$

em que  $\boldsymbol{\Delta}$  é a matriz  $(p + q + 1) \times n$  que depende do tipo de perturbação em que seus elementos são dados por  $\Delta_{zi} = \partial^2 L_p(\boldsymbol{\theta}|\boldsymbol{\omega}) / \partial \theta_z \partial \omega_i$ ,  $z = 1, 2, \dots, p + q + 1$  e  $i = 1, 2, \dots, n$  avaliada em  $\hat{\boldsymbol{\theta}}$  e  $\boldsymbol{\omega}_0$ , em que  $\boldsymbol{\omega}_0$  é o vetor de não perturbação.

As curvaturas normais  $C_1(\boldsymbol{\beta})$ ,  $C_1(\mathbf{f})$ ,  $C_1(\phi)$  podem ser calculadas para fazer os gráficos de índices, por exemplo, o gráfico de índices  $\mathbf{l}_{max}$ , do autovetor correspondente ao  $C_{\mathbf{l}_{max}}$ , maior autovalor da matriz  $\mathbf{B} = -\boldsymbol{\Delta}^T [\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta}$  e o gráficos de índices  $C_{\mathbf{l}_i}(\boldsymbol{\beta})$ ,  $C_{\mathbf{l}_i}(\mathbf{f})$  e  $C_{\mathbf{l}_i}(\phi)$ , que juntos denotam a influência local total (ver, por exemplo, LESAFFRE e VERBEKE (1998)), em que  $\mathbf{l}_i$  denota um vetor  $n \times 1$  de zeros com 1 na  $i$ -ésima posição. Assim, a curvatura na direção  $\mathbf{l}_i$  assume a forma  $C_i = 2|\mathbf{l}_i^T \boldsymbol{\Delta}^T [\ddot{\mathbf{L}}_p(\boldsymbol{\theta})]^{-1} \boldsymbol{\Delta} \mathbf{l}_i|$ . Um critério

descritivo é destacar as observações tais que  $C_i > 2\bar{C}$  em que  $\bar{C} = \sum_{i=1}^n \frac{C_i}{n}$ . Essa análise foi realizada para dois casos de perturbação, a matriz:

$$\Delta_{zi} = \left( \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_z \partial \omega_i} \right)_{(p+q+1) \times n},$$

em que  $z = 1, 2, \dots, p + q + 1$  e  $i = 1, \dots, n$ . Para as definições a seguir foi considerado o modelo definido na Seção (3.2), a função verossimilhança penalizada (3.1) e o vetor de pesos  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ , ou seja, o vetor de perturbações.

### 3.7.1 Ponderação de casos

Considere o vetor de pesos  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ . Neste caso, a perturbação da função log-verossimilhança tem a forma:

$$L_p(\boldsymbol{\theta}; \lambda|\boldsymbol{\omega}) = -\frac{1}{2} \sum_{i=1}^n \omega_i \log \phi + \sum_{i=1}^n \omega_i \log h(u_i) - \frac{\lambda}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

em que  $u_i = \frac{(y_i - \mu_i)^2}{\phi}$ ,  $0 \leq \omega_i \leq 1$ ,  $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$  e  $i = 1, \dots, n$ . Os elementos da matriz  $\Delta = (\Delta_{\boldsymbol{\beta}}^T, \Delta_{\mathbf{f}}^T, \Delta_{\phi}^T)^T$ , com as respectivas ordens  $p \times n$ ,  $q \times n$  e  $1 \times n$ , podem ser obtidas da forma

$$\Delta_{\boldsymbol{\beta}} = \frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) (\mathbf{y} - \boldsymbol{\mu}) \mathbf{e}^T$$

$$\Delta_{\mathbf{f}} = \frac{1}{\phi} \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi}) (\mathbf{y} - \boldsymbol{\mu}) \mathbf{e}^T$$

$$\Delta_{\phi} = \frac{1}{2\phi} \left[ \mathbf{e}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\mathbf{u}) - \mathbf{e}^T \right],$$

em que  $\mathbf{e} = [e_1 \dots e_n]^T$ ,  $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1, \dots, v_n\}$ ,  $\mathbf{D}(\boldsymbol{\xi}) = \text{diag}\{\dot{a}_1, \dots, \dot{a}_n\}$ ,  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$  e  $\mathbf{u} = \frac{(\mathbf{y} - \boldsymbol{\mu})^T (\mathbf{y} - \boldsymbol{\mu})}{\phi}$ .

### 3.7.2 Perturbação da variável resposta

Considere que cada  $y_i$  seja perturbada por  $y_{i\omega} = y_i + \omega_i$ . Assim, o logaritmo da função de verossimilhança perturbada é definida da seguinte forma:

$$L_p(\boldsymbol{\theta}; \lambda|\boldsymbol{\omega}) = \sum_{i=1}^n \left\{ -\frac{1}{2} \log \phi + \log h(u_{i\omega}) \right\} - \frac{\lambda}{2} \mathbf{f}^T \mathbf{K} \mathbf{f},$$

em que  $h(u_{i\omega})$  é uma distribuição simétrica univariada com perturbação na variável resposta,  $u_{i\omega} = \frac{(y_{i\omega} - \mu_i)^2}{\phi}$ ,  $0 \leq \omega_i \leq 1$ ,  $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$  e  $i = 1, \dots, n$ . Os elementos da matriz

$\Delta = (\Delta_{\beta}^T, \Delta_{\mathbf{f}}^T, \Delta_{\phi}^T)^T$ , com as respectivas ordens  $p \times n$ ,  $q \times n$  e  $1 \times n$ , podem ser obtidas da forma

$$\Delta_{\beta} = \frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{e}) \mathbf{D}(\mathbf{a}) \mathbf{D}(\xi)$$

$$\Delta_{\mathbf{f}} = \frac{1}{\phi} \mathbf{N}^T \mathbf{D}(\mathbf{e}) \mathbf{D}(\mathbf{a}) \mathbf{D}(\xi)$$

$$\Delta_{\phi} = \frac{1}{\phi^2} \left[ (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{m}) \mathbf{D}(\mathbf{e}) \right]$$

em que  $\mathbf{D}(\mathbf{e}) = \text{diag}\{e_1, \dots, e_n\}$ ,  $a_i = [v_{i\omega} - 4W'_h(u_{i\omega})u_{i\omega}]$ ,  $\mathbf{D}(\mathbf{a}) = \text{diag}\{a_1, \dots, a_n\}$ ,  $m_i = [v_{i\omega} - 2W'_h(u_{i\omega})u_{i\omega}]$ ,  $\mathbf{m} = (m_1, \dots, m_n)^T$ ,  $\mathbf{D}(\mathbf{m}) = \text{diag}\{m_1, \dots, m_n\}$ ,  $\mathbf{D}(\xi) = \text{diag}\{\dot{a}_1, \dots, \dot{a}_n\}$ ,  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ .

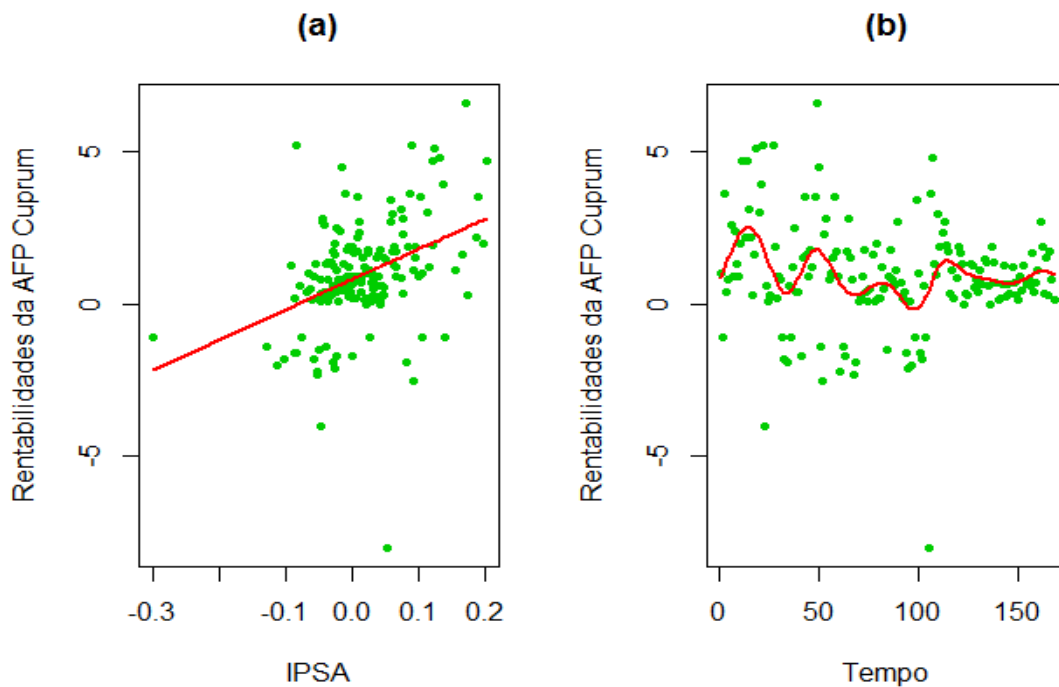
### 3.8 Considerações

Nesse capítulo definiu-se o Modelo Linear Parcial Generalizado Simétrico, efetuou-se os cálculos das derivadas da função escore penalizada a partir da função verossimilhança penalizada, bem como matriz hessiana penalizada, fez-se uma adaptação do algoritmo *backfitting* para o MLPGS e por fim as derivadas de alguns casos de perturbação para as curvaturas normais de influência local foram obtidas.

## 4 RESULTADOS

### 4.1 Introdução

Nesse Capítulo será realizada uma aplicação do modelo linear parcial generalizado simétrico, baseado no trabalho de IBACACHE-PULGAR e PAULA (2011). O conjunto de dados que será utilizado é referente a área de finanças, mais especificamente das instituições financeiras privadas encarregadas de administrar os fundos de contas individuais de poupança para pensões (AFP) do Chile. As rentabilidades mensais da administradora de fundos de pensões *Cuprum* representa a variável resposta, a rentabilidade do indicador macroeconômico IPSA é a variável explicativa e o tempo (em meses) corresponde ao período de janeiro de 1990 a dezembro de 2003 em que foram registradas as rentabilidades, ou seja, num total de 168 meses.



**Figura 4.1.** Diagrama de dispersão: (a) Rentabilidade AFP Cuprum versus IPSA e (b) Rentabilidade AFP Cuprum versus tempo.

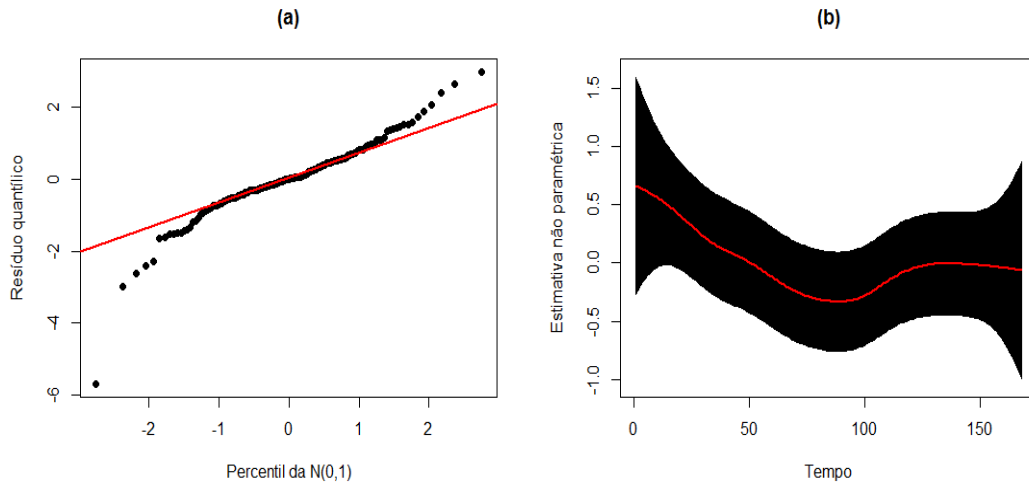
A Figura 4.1 (a) apesar de algumas observações com comportamento atípico, apresenta fortes indícios de uma tendência linear entre as variáveis  $y_i$ , que representa rentabilidade da AFP CUPRUM e  $x_i$  que representa IPSA, o mesmo não acontece na Figura 4.1 (b) que mostra uma relação não linear entre rentabilidade da AFP CUPRUM e o tempo (meses), que pode ser considerada de forma não-paramétrica.

## 4.2 Modelos propostos

Nota-se na Figura 4.2(a) possui varias observações atípicas nos extremos, indicando que o modelo

$$Rentabilidade_i \sim N(\mu_i, \phi), \text{ em que } g(\mu_i) = \mu_i = \beta_1 + \beta_2 IPSA_i + f(tempo_i),$$

não é ideal, já na Figura 4.2(b), referente a função não-paramétrica, observa-se uma variabilidade maior que 1,5, contudo ao realizar ajustes de outros modelos, permitirá dizer se há modelos que têm menor variabilidade da função não-paramétrica, ou seja, uma curva mais suave, garantindo assim um melhor ajuste.



**Figura 4.2.** Gráfico normal de probabilidade para o resíduo quantílico (esquerda) e função não-paramétrica (direita) sob o modelo Normal.

Os demais modelos a serem estudados são apresentados a seguir, no qual foram consideradas apenas as funções de ligação identidade ( $\mu_i$ ) e inversa ( $\frac{1}{\mu_i}$ ), uma vez que no conjunto de dados haviam valores negativos, impossibilitando a utilização das funções de ligação raiz quadrada ( $\sqrt{\mu_i}$ ) e logarítmica ( $\log(\mu_i)$ ).

1.  $Rentabilidade_i \sim t(\mu_i, \phi, \nu)$ , em que  $g(\mu_i) = \mu_i = \beta_1 + \beta_2 * IPSA_i + f(tempo_i)$ ;
2.  $Rentabilidade_i \sim EP(\mu_i, \phi, \nu)$ , em que  $g(\mu_i) = \mu_i = \beta_1 + \beta_2 * IPSA_i + f(tempo_i)$ ;
3.  $Rentabilidade_i \sim N(\mu_i, \phi)$ , em que  $g(\mu_i) = \frac{1}{\mu_i} = \beta_1 + \beta_2 * IPSA_i + f(tempo_i)$ ;
4.  $Rentabilidade_i \sim t(\mu_i, \phi, \nu)$ , em que  $g(\mu_i) = \frac{1}{\mu_i} = \beta_1 + \beta_2 * IPSA_i + f(tempo_i)$ ;
5.  $Rentabilidade_i \sim EP(\mu_i, \phi, \nu)$ , em que  $g(\mu_i) = \frac{1}{\mu_i} = \beta_1 + \beta_2 * IPSA_i + f(tempo_i)$ .

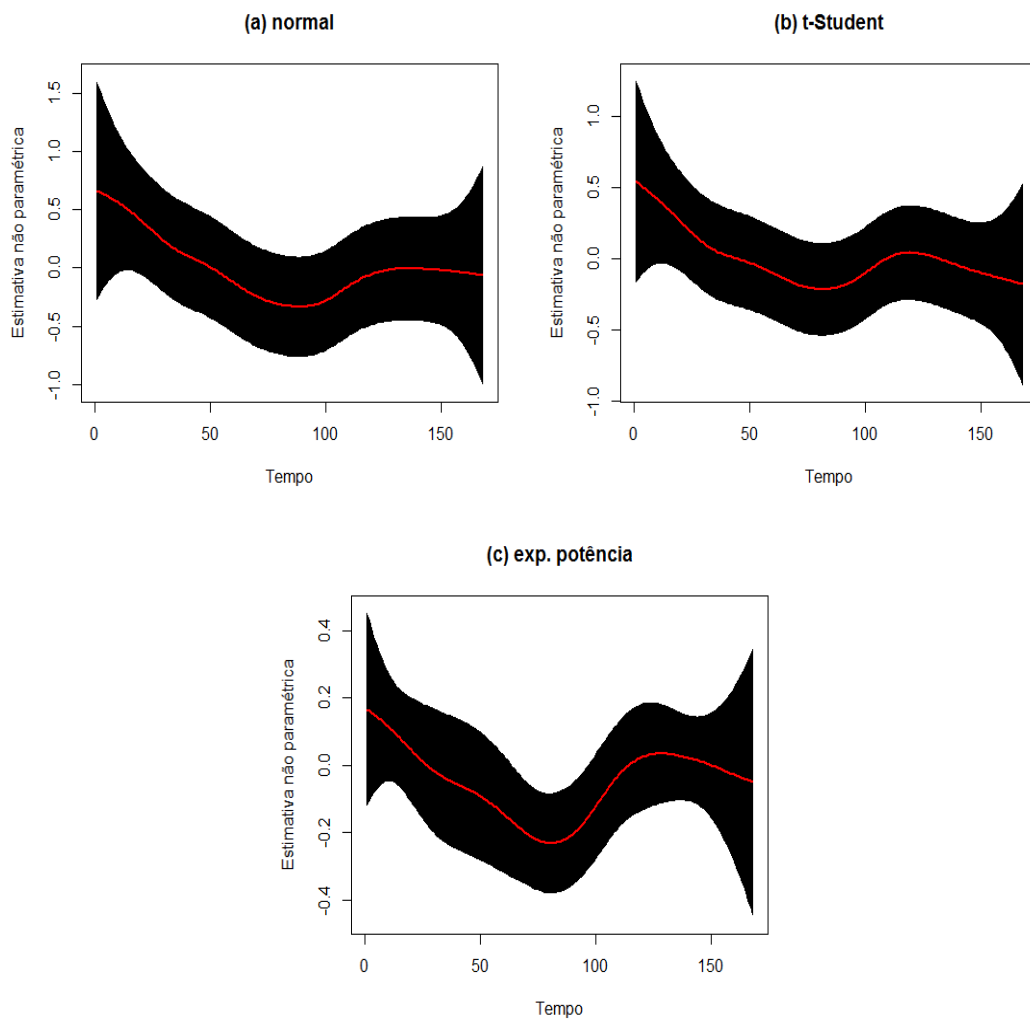
#### 4.2.1 Modelos Normal, t-Student e Exponencial potência, com função de ligação identidade.

Constata-se por intermédio da Tabela 4.1, que as estimativas dos coeficientes de regressão apresentam pouca diferença, no qual o modelo Normal e Exponencial potência apresentam uma diferença um pouco maior em relação ao modelo t-Student. Porém é possível notar que quando mudam as EMV na parte paramétrica dos 3 modelos o erro padrão de  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$  do modelo Exponencial potência possui o menor valor, ou seja, indica menor variabilidade em relação aos demais modelos, mostrando que o modelo que apresenta caudas mais pesadas presume proporcionar estimativas mais precisas para o coeficiente de regressão. O erro padrão de  $\hat{\phi}$  do modelo Normal apresentou menor valor em relação aos demais modelos. Foi estimado o valor de  $\nu$  do modelo t-Student e do modelo Exponencial potência e para o modelo Exponencial potência a estimativa foi próximo de 1, indicando que o modelo está convergindo para a distribuição Exponencial dupla (distribuição de Laplace), a estimativa de  $\lambda$  dos três modelos foi próximo de zero, indicando que tem pouca intervenção e antepõe o ajuste por interpolação da curva. Também, o número de graus de liberdade efetivos de todos os modelos foram próximos de três, ou seja, as partições da curva interligadas por cada nó tem grau três.

**Tabela 4.1.** Estimativas de máxima verossimilhança dos modelos Normal, t-Student e Exponencial potência ajustado aos dados da área de finanças do Chile, considerando a função de ligação identidade.

Modelo	Parâmetro	Estimativa	Erro padrão	valor t	valor-p
Normal	$\beta_1$	1,0845	0,2530	4,2860	<0,0001
	$\beta_2$	9,0802	1,7738	5,1190	<0,0001
	$\phi$	1,5919	0,0869	18,3300	<0,0001
	$\lambda$	0,0104			
	$df$	3,0006			
t-Student	$\beta_1$	1,0693	0,2123	5,0370	<0,0001
	$\beta_2$	8,4995	1,7050	4,9850	<0,0001
	$\phi$	0,9685	0,1169	8,2850	<0,0001
	$\nu$	2,7624			
	$\lambda$	0,0104			
Exp. potência	$\beta_1$	0,7966	0,0201	39,5830	<0,0001
	$\beta_2$	7,6178	0,0776	98,1880	<0,0001
	$\phi$	1,6940	0,1989	8,5150	<0,0001
	$\nu$	0,7051			
	$\lambda$	0,0083			
	$df$	2,9994			

A Figura 4.3 descreve o ajuste da variável explicativa tempo. O modelo que apresentou maior variabilidade foi o modelo Normal e o que apresenta menor variabilidade o modelo Exponencial potência. Sendo assim, há indícios que o modelo que melhor ajustou é o Exponencial potência com a função de ligação identidade.

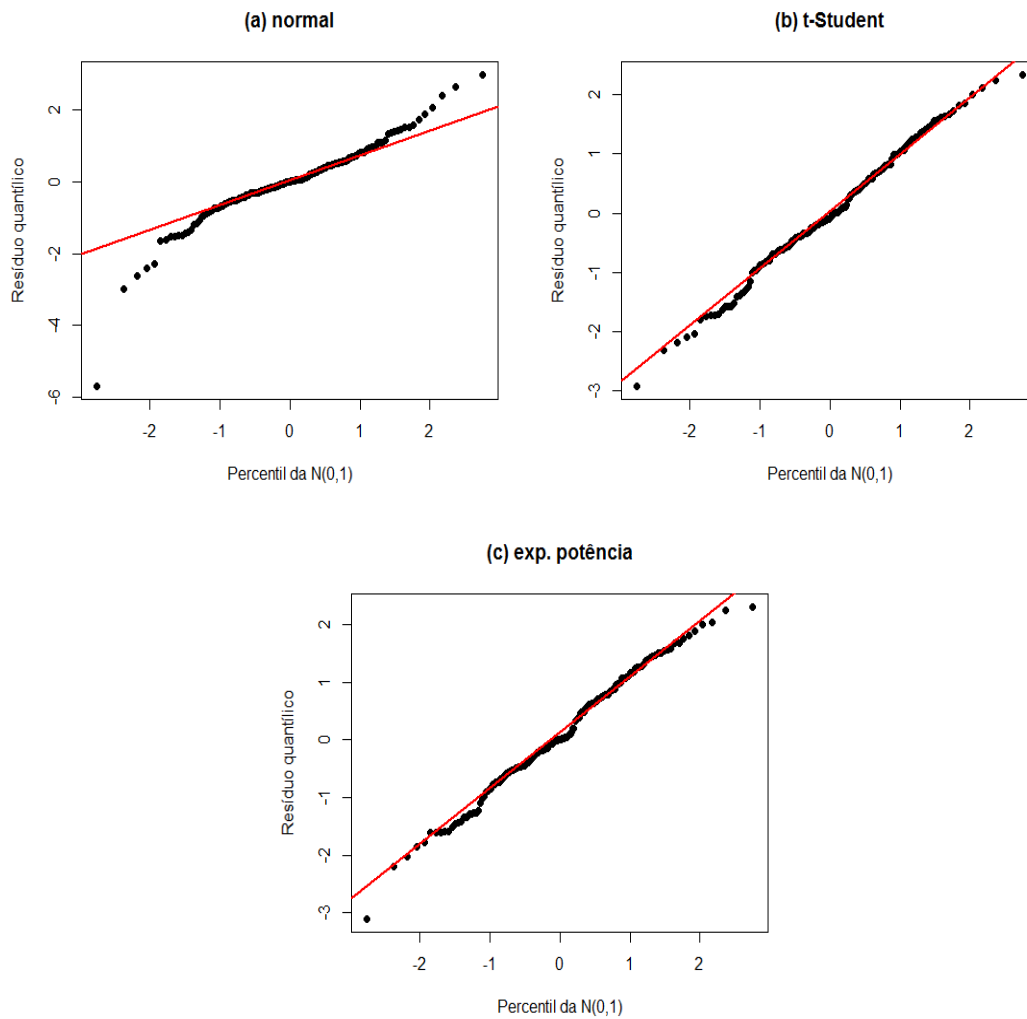


**Figura 4.3.** Função não-paramétrica dos modelos Normal, t-Student e Exponencial potência ajustada (spline cúbico) à variável explicativa tempo considerando a função de ligação identidade, com a respectiva banda de confiança de 95%.

Há várias maneiras de avaliar a qualidade do ajuste de um modelo. Uma delas foi proposta por LANGE *et al.* (1989) para casos multivariados e que pode ser modificada para o contexto univariado. No presente trabalho, será utilizada para os modelos Normal, t-Student e Exponencial potência. Com base nos gráficos normais de probabilidade para o resíduo quantílico sob os modelos apresentados e nos valores de  $L_p(\hat{\theta}, \hat{\lambda})$  na Figura 4.4, apesar de algumas observações atípicas nas extremidades, há indícios que o modelo Exponencial potência com  $\nu = 0,7051$  obteve um ajuste mais propício em comparação ao modelo Normal que apresenta muitas observações atípicas nas extremidades indicando a necessidade de ajustar um modelo com distribuição que possui caudas mais pesadas do que a mesma. Já o gráfico do modelo t-Student, ficou semelhante ao do modelo Exponencial potência, embora seja possível notar que há mais observações que não estão sobre a reta

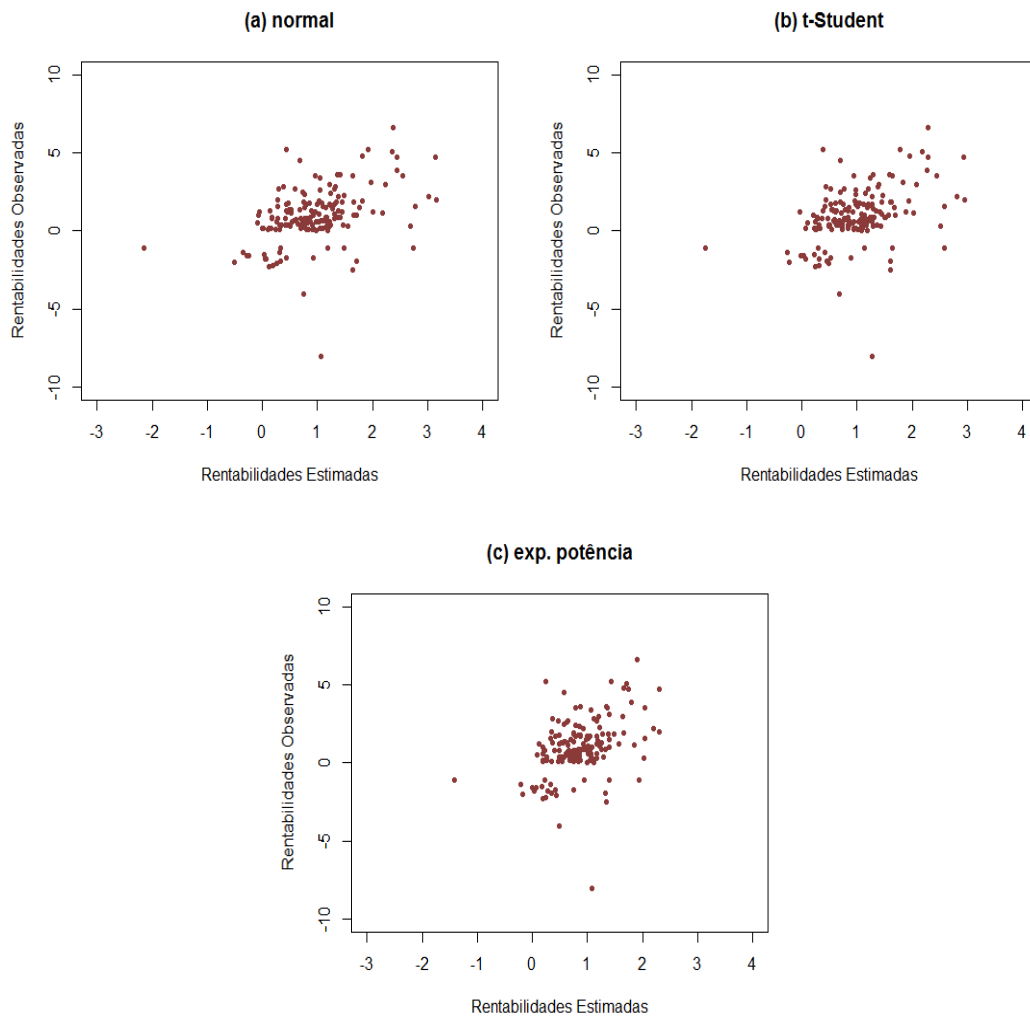


vermelha.



**Figura 4.4.** Gráficos normais de probabilidade para o resíduo quantílico sob os modelos Normal, t-Student e Exponencial potência ajustados aos dados da AFP com função de ligação identidade.

Note que, na Figura 4.5, o modelo que apresenta uma tendência linear com menor variabilidade é o modelo Exponencial potência.



**Figura 4.5.** Gráfico valores preditos versus observados dos modelos Normal, t-Student e Exponencial potência, considerando a função de ligação identidade.

#### 4.2.2 Modelos Normal, t-Student e Exponencial potência, com função de ligação inversa.

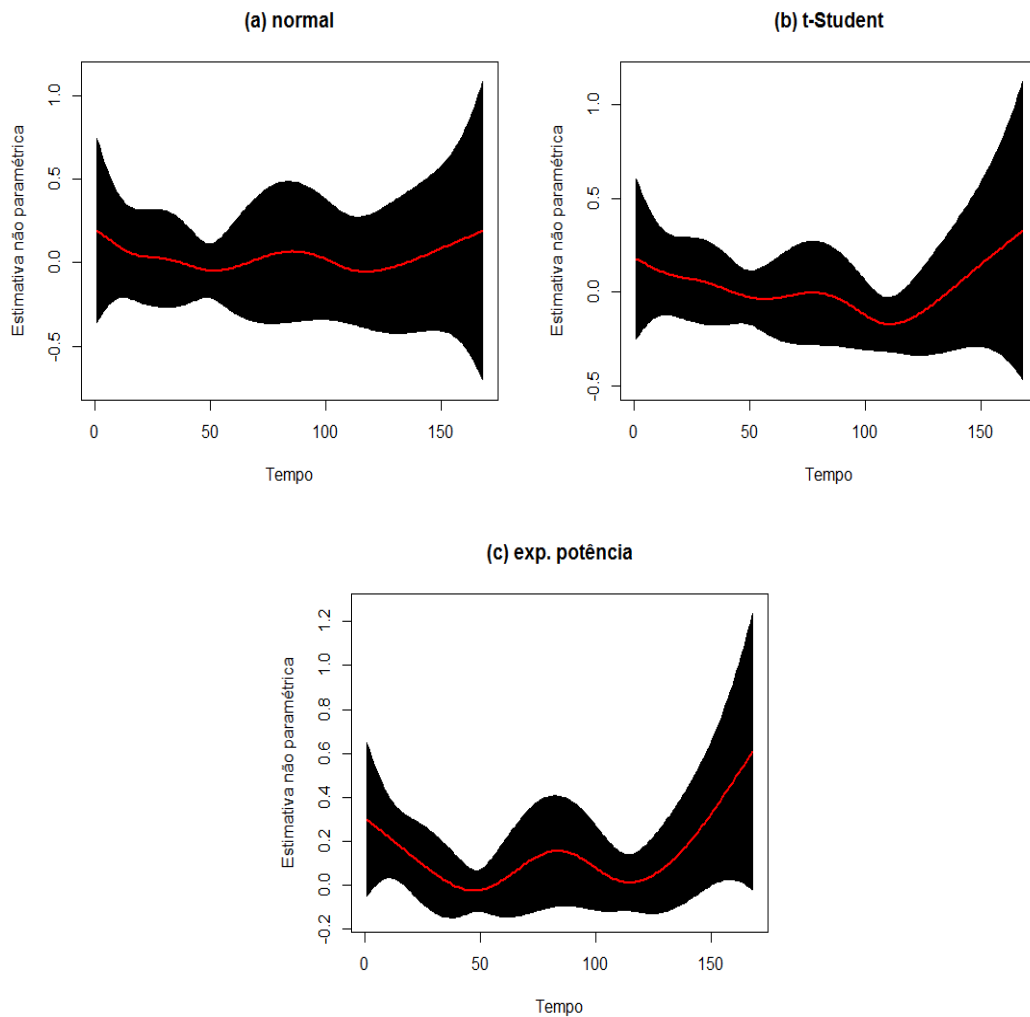
Da Tabela 4.2, verifica-se que as estimativas dos coeficientes de regressão dos três modelos apresentam pouca diferença. Contudo, o erro padrão de  $\hat{\beta}$  do modelo Exponencial potência é menor que àqueles para as distribuições Normal e t-Student, mostrando que o modelo que apresenta caudas mais pesadas presume fornecer estimativas mais precisas para o coeficiente de regressão. O erro-padrão de  $\hat{\phi}$  do modelo Normal apresentou menor valor em relação aos demais modelos. Utilizando a função exponencial foi calculado a estimativa de  $\nu$  do modelo t-Student e do modelo Exponencial potência. Da mesma forma como do modelo Exponencial potência, a função de ligação identidade também foi próximo de 1 e a estimativa do parâmetro de suavização foi muito próximo de zero, indicando que

**Tabela 4.2.** Estimativas máximas verossimilhança dos modelos Normal, t-Student e Exponencial potência ajustado aos dados da área de finanças do Chile, considerando a função de ligação inversa.

Modelo	Parâmetro	Estimativa	Erro padrão	valor t	valor-p
Normal	$\beta_1$	1,2153	0,2219	5,4790	<0,0001
	$\beta_2$	-4,9047	1,0419	-4,7070	<0,0001
	$\phi$	1,6324	0,0891	18,3200	<0,0001
	$\lambda$	0,0026			
	$df$	2,9995			
t-Student	$\beta_1$	1,2457	0,1784	6,9840	<0,0001
	$\beta_2$	-4,7766	0,8459	-5,6460	<0,0001
	$\phi$	0,8814	0,1269	6,9460	<0,0001
	$\nu$	2,1271			
	$\lambda$	0,0042			
Exp. potência	$\beta_1$	1,2995	0,1404	9,2530	<0,0001
	$\beta_2$	-4,8185	0,6380	-7,5520	<0,0001
	$\phi$	1,6850	0,1451	11,6200	<0,0001
	$\nu$	0,7451			
	$\lambda$	0,0028			
	$df$	3,0006			

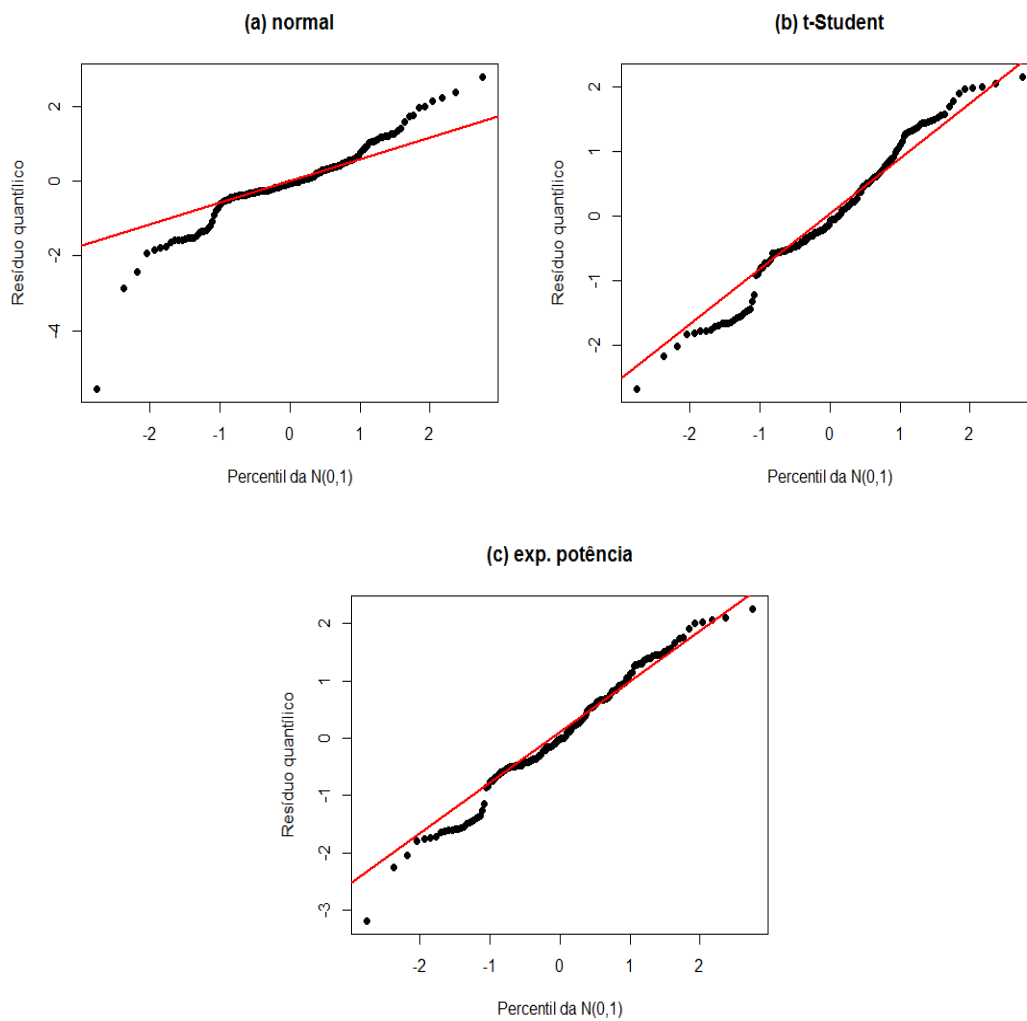
tem pouca influência e prevalece o ajuste por interpolação da curva. Ademais, o número de graus de liberdade efetivos foram próximo de três, ou seja, o grau do polinômio de cada partição da curva que estão interligadas por cada nó é três.

A Figura 4.6 aponta o modelo Exponencial potência com menor variabilidade, descrevendo o ajuste da variável tempo. Dessa maneira, há indícios de que o modelo que melhor ajustou foi o Exponencial potência com função de ligação inversa.



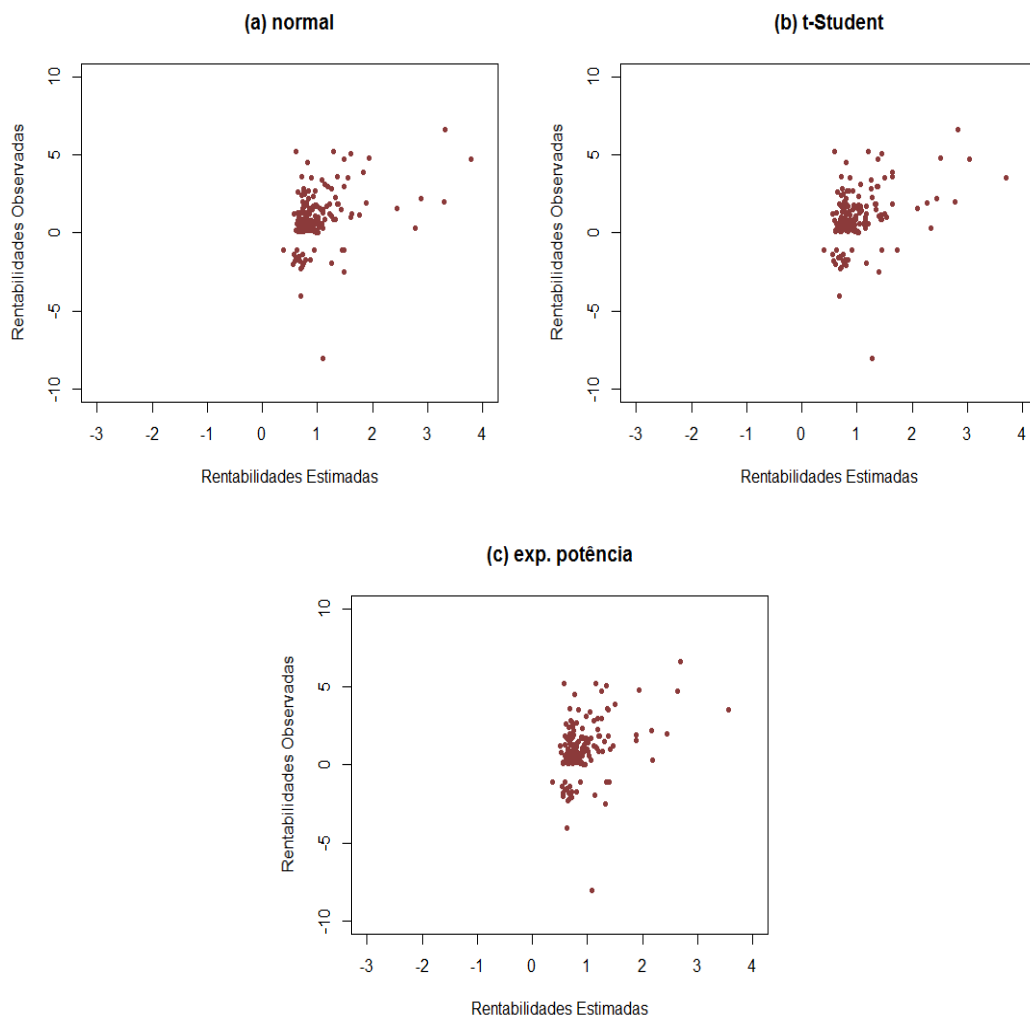
**Figura 4.6.** Função não-paramétrica dos modelos Normal, t-Student e Exponencial potência ajustada (spline cúbico) à variável explicativa tempo considerando a função de ligação inversa, com a respectiva banda de confiança de 95%.

Na Figura 4.7 considerou-se a função de ligação inversa nos modelos Normal, t-Student e Exponencial potência. Com base nos gráficos normais de probabilidade para o resíduo quantílico sob os modelos apresentados e nos valores de  $L_p(\hat{\theta}, \hat{\lambda})$ , os três gráficos apresentaram um mal ajuste. No entanto, o que apresentou o número maior de observações atípicas nas extremidades foi o modelo Normal, indicando a necessidade da escolha de distribuições que possuem caudas mais pesadas do que a mesma.



**Figura 4.7.** Gráficos normais de probabilidade para o resíduo quantílico sob os modelos Normal, t-Student e Exponencial potência ajustados aos dados da AFP com função de ligação inversa.

A partir da Figura 4.8, o ajuste não ficou bom com nenhum modelo, considerando a função de ligação inversa.



**Figura 4.8.** Gráfico valores preditos versus observados dos modelos Normal, t-Student e Exponencial potência, considerando a função de ligação inversa.

### 4.3 Seleção do modelo

Por meio da Tabela 4.3, observa-se que o modelo Exponencial potência com a função de ligação identidade foi o que apresentou menor AIC, ou seja, o melhor modelo selecionado. Este pode ser representado da seguinte forma

$$Rentabilidade_i \sim EP(\mu_i, \phi, \nu), \text{ em que } g(\mu_i) = \mu_i = \beta_1 + \beta_2 * IP_{SA_i} + f(tempo_i).$$

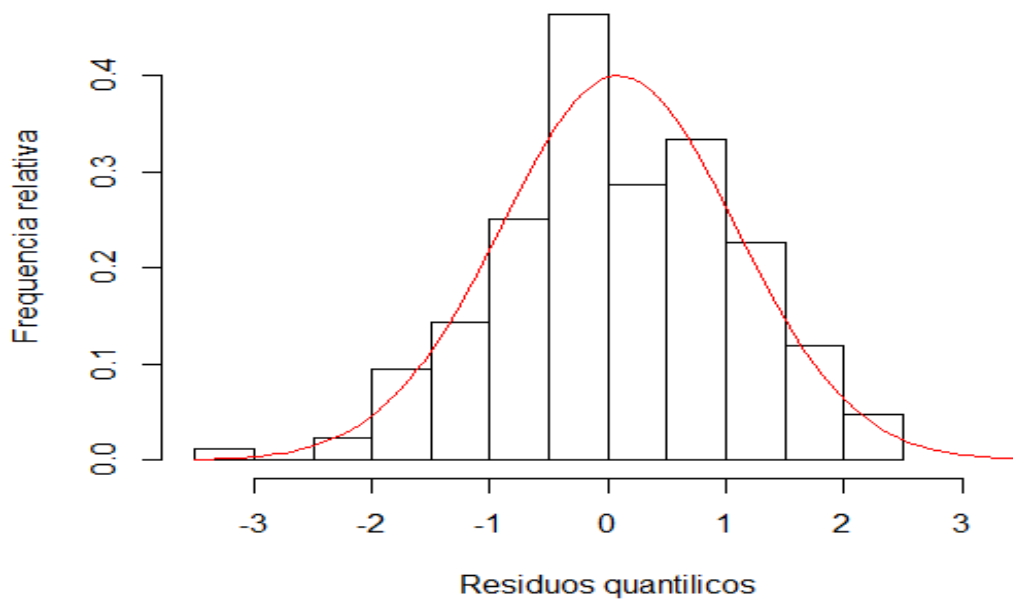
**Tabela 4.3.** Resultados do critério AIC para os modelos Normal, t-Student e Exponencial potência, com funções de ligação identidade e inversa.

Distribuição	Função de ligação	AIC
Exp. potência	identidade	602,9726
Exp. potência	inversa	610,1567
t-Student	identidade	611,2789
t-Student	inversa	620,5971
Normal	identidade	646,9863
Normal	inversa	655,4291

Por meio da Tabela 4.4 e da Figura 4.9, observa-se, para ambos dos resíduos quantílicos, o modelo Exponencial potência com função de ligação identidade segue uma distribuição simétrica tendendo a distribuição Normal padrão. Os resíduos se comportam bem, pois: a média = 0,0852, ou seja, é próximo de zero, sua variância = 0,9969, isto é, próximo de um, seu coeficiente de assimetria =  $-0,1573$ , ou seja, próximo de zero, seu coeficiente de curtose = 2,7726, próximo de três.

**Tabela 4.4.** Resumo do resíduo quantílico produzido pelo gamlss do modelo Exponencial potência, considerando a função de ligação identidade.

média	0,0852
variância	0,9969
coef. de assimetria	-0,1573
coef. de curtose	2,7726



**Figura 4.9.** Histograma dos resíduos quantílicos do modelo Exponencial potência com função de ligação identidade.





## 5 DISCUSSÃO

Nessa dissertação foi proposta uma extensão do modelo linear generalizado simétrico, foram apresentados resultados de estimação para o modelo linear parcial generalizado simétrico. Para a função não paramétrica, foram utilizados splines cúbicos e aplicadas funções de verossimilhança penalizada para a obtenção das estimativas de máxima verossimilhança penalizadas. Além de uma revisão bibliográfica para ser adaptados procedimentos iterativos para a estimação dos parâmetros, a estimação do parâmetro de suavização e do número de graus de liberdade efetivos, foi constatado que distribuições que possuem caudas mais pesadas do que a distribuição Normal, podem se ajustar melhor ao conjunto de dados reais referente a área de finanças, mais especificamente das instituições financeiras privadas encarregadas de administrar os fundos de contas individuais de poupança para pensões (AFP) do Chile, mediante as estimativas dos parâmetros, análise de diagnósticos e do critério de AIC.

### 5.1 Aspectos computacionais

Para estimar o coeficiente de regressão nessa dissertação, foi utilizado o processo iterativo baseado no algoritmo de score de Fisher e algoritmo de *backfitting*, as estimativas da parte paramétrica, as funções não paramétricas, as estimativas do parâmetro de suavização, os graus de liberdade efetivos e os gráficos, foram obtidos com uso do pacote *gamlss* (RIGBY e STASINOPOULOS, 2005) disponível no software R (TEAM, 2015).

### 5.2 Trabalhos futuros

1. Uma extensão do modelo proposto é considerar outros tipos de penalização, tais como por exemplo P-splines, que apresenta a vantagem de diminuir o custo computacional na estimação do componente não paramétrico.
2. Considerar mais de uma função não paramétrica no preditor linear, ou seja, considerar mais covariáveis que se relacionam de forma não paramétrica (não linear) com a variável resposta.



## REFERÊNCIAS

- AKANTZILIOTOU, C., R. RIGBY, e D. STASINOPOULOS, 2002 The r implementation of generalized additive models for location, scale and shape. In *Statistical modelling in Society: Proceedings of the 17th International Workshop on statistical modelling*, pp. 75–83, Statistical Modelling Society.
- AYDIN, D., 2014 Estimations of the partially linear models with smoothing spline based on different selection methods: A comparative study. *Journal of Approximation Theory* **30**: 35–56.
- BATES, D. M. e D. G. WATTS, 1980 Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–25.
- BOOR, C., 1972 On calculating with b-splines. *Journal of Approximation Theory* **6**: 50–62.
- BOOR, C., 1977 Package for calculating with b-splines. *SIAM Journal on Numerical Analysis* **14**: 441–472.
- BOOR, C. e E.-U. MATHÉMATICIEN, 1978 *A practical guide to splines*, volume 27. Springer-Verlag New York.
- BOX, G. e G. TIAO, 1973 *Bayesian Inference in Statistical Analysis*. Addison-Wesley, California.
- BREIMAN, L. e J. H. FRIEDMAN, 1985 Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association* **80**: 580–598.
- CAMBANIS, S., S. HUANG, e G. SIMONS, 1981 On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* **11**: 368–385.
- CHMIELEWSKI, M., 1981 Elliptically symmetric distributions: a review and bibliography. *International Statistical Review* **49**: 67–74.
- COLE, T. e P. GREEN, 1992 Smoothing reference centile curves: the lms method and penalized likelihood. *Statistics in medicine* **11**: 1305–1319.
- COOK, R., 1986 Assessment of local influence (with discussion). *Journal Royal Statistics Society B* **48**: 133–169.
- CORDEIRO, G. M. e C. DEMETRIO, 1986 *Modelos lineares generalizados*. Univ. Estadual de Campinas. Dep. de Estatística Campinas.
- CRAVEN, P. e G. WAHBA, 1978 Smoothing noisy data with spline functions. *Numerische Mathematik* **31**: 377–403.

- CYSNEIROS, F., G. PAULA, e M. GALEA, 2007 Heteroscedastic symmetrical linear models. *Statistics and Probability Letters* **77**: 1084–1090.
- DEVROYE, L., 1986 *Non-Uniform Random Variable Generation*. Springer-Verlag, Canada.
- DIAS, R., 2001 O uso de splines em regressão não paramétrica. Coloquio do depto. de Estatística, IME-USP .
- DIERCKX, P., 1993 *Curve and Surface Fitting with Splines*. Clarendon Press, New York.
- DINSE, G. e S. LAGAKOS, 1983 Regression analysis of tumour prevalence data. *Journal of the Royal Statistical Society C* **32**: 236–248.
- DUNN, P. e G. SMYTH, 1996 Randomized quantile residuals. *Journal of Computational and Graphical Statistics* pp. 236–244.
- DURBÁN, M., D. LEE, e M. D. UGARTE, 2008 *Splines con penalizaciones (P-splines): teoría y aplicaciones*. Universidad Pública de Navarra Nafarroako Unibertsitate Publikoa.
- EILERS, P. H. C. e B. D. MARX, 1996 Flexible smoothing with b-splines and penalties. *Statistical Science* **11**: 89–102.
- EUBANK, R. L., 1999 *Nonparametric regression and spline smoothing*. CRC press.
- FANG, K., S. KOTZ, e K. NG, 1990 *Symmetric Multivariate and Related Distribution*. Chapman & Hall, London, 36th edition.
- GALEA, M., G. PAULA, e F. CYSNEIROS, 2005 On diagnostics in symmetrical nonlinear models. *Statistics and Probability Letters* **73**: 459–467.
- GALEA, M., G. PAULA, e M. URIBE-OPAZO, 2003 On influence diagnostics in univariate elliptical linear regression models. *Statistical Papers* **44**: 23–45.
- GRADSHTEYN, I. e I. RYZHIK, 2000 *Table of Integrals, Series, and Products*. Academic Press, New York, 6nd ed edition.
- GREEN, P. e B. SILVERMAN, 1993 *Nonparametric regression and generalized linear models: A roughness penalty approach*. Chapman and Hall/CRC, New York.
- GREEN, P. e B. YANDELL, 1985 Semi-parametric generalized linear models. *Lecture Notes in Statistics* **32**: 44–55.
- HASTIE, T. e R. TIBSHIRANI, 1990 *Generalized Additive Models*. Chapman & Hall/CRC.

- HOWLADER, H., 1989 Estimation of the location and scale of a symmetric hyperbolic distribution. *Statistical Papers* pp. 271–278.
- IBACACHE-PULGAR, G. e G. PAULA, 2011 Local influence for student partially linear models pp. 1462–1478.
- JØRGENSEN, B., 1982 *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer, New York, 2nd ed edition.
- KELKER, D., 1970 Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya A* pp. 419–430.
- KREYSZIG, E., 1991 Principal normal, curvature, osculating circle. *Differential Geometry*. New York: Dover pp. 34–36.
- LANGE, K. e J. SINSHEIMER, 1993 Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* pp. 175–198.
- LANGE, K. L., R. J. LITTLE, e J. M. TAYLOR, 1989 Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**: 881–896.
- LESAFFRE, E. e G. VERBEKE, 1998 Local influence in linear mixed models. *International Biometric Society* **54**: 570–582.
- NELDER, J. e R. WEDDERBURN, 1972 Generalized linear models. *Journal of the Royal Statistical Society A* pp. 370–384.
- OKI, F., 2015 Modelos semiparametricos com resposta binomial negativa .
- O’SULLIVAN, F., 1986 A statistical perspective on ill-posed inverse problems. *Statistical science* pp. 502–518.
- PAULA, G. e F. CYSNEIROS, 2009 Systematic risk estimation in symmetric models. *Applied Economics Letters* pp. 217–221.
- RELVAS, C. E. M., 2013 *Modelos parcialmente lineares com erros simétricos autoregressivos de primeira ordem*. Ph.D. thesis, Universidade de São Paulo.
- RIGBY, R. e D. STASINOPOULOS, 1996 A semi-parametric additive model for variance heterogeneity. *Statistics and Computing* **6**: 57–65.
- RIGBY, R. e D. STASINOPOULOS, 2001 The gamlss project: a flexible approach to statistical modelling. In *New trends in statistical modelling: Proceedings of the 16th international workshop on statistical modelling*, pp. 337–345.

- RIGBY, R. A. e D. M. STASINOPOULOS, 2005 Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**: 507–554.
- STASINOPOULOS, D., R. RIGBY, e C. AKANTZILIOTOU, 2006 Instructions on how to use the gamlss package in r. Accompanying documentation in the current GAMLSS help files,(see also <http://www.gamlss.org/>) .
- STOKER, J. J., 1969 Differential geometry. *Pure and Applied Mathematics* **20**.
- TEAM, R. C., 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- VILLEGAS, C., G. PAULA, F. CYSNEIROS, e M. GALEA, 2013 Influence diagnostics in generalized symmetric linear models. *Computational Statistics and Data Analysis* pp. 161–170.
- WAHBA, G., 1990 *Spline models for observational data*, volume 59. Siam.

## APÊNDICES

### Apêndice A

O Apêndice A é referente aos cálculos de forma detalhada dos seguintes itens: função escore, algoritmo *backfitting*, influência local e da função  $W_h$ .

#### Função escore

O logaritmo da função de verossimilhança penalizada é da forma:

$$L_p(\boldsymbol{\theta}, \lambda) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log h(u_i) - \frac{\lambda}{2} \mathbf{f}^T \mathbf{K} \mathbf{f}.$$

$$\begin{aligned} \mathbf{U}_{\beta}^p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left\{ \frac{h'(u_i)}{h(u_i)} \left[ -\frac{2(y - \mu_i)}{\phi} \left( \frac{dg^{-1}(\eta_i)}{d\eta_i} \right) \mathbf{x}_i \right] \right\} \\ &= \sum_{i=1}^n \left\{ W_h(u_i) \left[ -\frac{2(y - \mu_i)}{\phi} \left( \frac{dg^{-1}(\eta_i)}{d\eta_i} \right) \mathbf{x}_i \right] \right\} \\ &= \frac{1}{\phi} \sum_{i=1}^n v_i \dot{a}_i(y_i - \mu_i) \mathbf{x}_i. \end{aligned}$$

$$\begin{aligned} \mathbf{U}_f^p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \mathbf{f}} = \sum_{i=1}^n \left\{ \frac{h'(u_i)}{h(u_i)} \left[ -\frac{2(y - \mu_i)}{\phi} \left( \frac{dg^{-1}(\eta_i)}{d\eta_i} \right) \mathbf{n}_i \right] \right\} - \lambda \mathbf{K} \mathbf{f} \\ &= \sum_{i=1}^n \left\{ W_h(u_i) \left[ -\frac{2(y - \mu_i)}{\phi} \left( \frac{dg^{-1}(\eta_i)}{d\eta_i} \right) \mathbf{n}_i \right] \right\} - \lambda \mathbf{K} \mathbf{f} \\ &= \frac{1}{\phi} \sum_{i=1}^n v_i \dot{a}_i(y_i - \mu_i) \mathbf{n}_i - \lambda \mathbf{K} \mathbf{f}. \end{aligned}$$

$$\begin{aligned} U_{\phi}^p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \lambda)}{\partial \phi} = -\frac{n}{2\phi} - \frac{1}{\phi^2} \sum_{i=1}^n \left\{ \frac{h'(u_i)}{h(u_i)} (y - \mu_i)^2 \right\} \\ &= -\frac{n}{2\phi} - \frac{1}{\phi^2} \sum_{i=1}^n [W_h(y - \mu_i)^2] \\ &= -\frac{n}{2\phi} + \frac{1}{2\phi^2} \sum_{i=1}^n v_i (y_i - \mu_i)^2. \end{aligned}$$

#### Adaptação do algoritmo backfitting

Temos que,

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f}$$



A a matriz da função escore em relação a  $\beta$  é

$$\mathbf{U}_{\beta}^p(\boldsymbol{\theta}) = \frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \boldsymbol{\mu}). \text{ Seja } \mathbf{U}_{\beta}^p(\boldsymbol{\theta}) = \mathbf{0}, \text{ então}$$

$$\begin{aligned} \frac{1}{\phi} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \boldsymbol{\mu}) &= 0 \\ \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})[\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f})] &= 0 \\ \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{y} - \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{N}\mathbf{f} &= 0 \\ \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{N}\mathbf{f}) \\ (\mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X})^{-1}(\mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X}^T)\boldsymbol{\beta} &= (\mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{N}\mathbf{f}) \\ \mathbf{I}\boldsymbol{\beta} &= (\mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{N}\mathbf{f}) \end{aligned}$$

e conseqüentemente ao seguinte algoritmo de *backfitting* com a matriz de peso dado por  $\mathbf{D}(\mathbf{v})$  é

$$\boldsymbol{\beta}^{(r+1)} = (\mathbf{X}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)})\mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)})(\mathbf{y} - \mathbf{N}\mathbf{f}^{(r)})$$

A a matriz da função escore em relação a  $\mathbf{f}$  é

$$\mathbf{U}_{\mathbf{f}}^p(\boldsymbol{\theta}) = \frac{1}{\phi} \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \boldsymbol{\mu}) - \lambda \mathbf{K}\mathbf{f}. \text{ Seja } \mathbf{U}_{\mathbf{f}}^p(\boldsymbol{\theta}) = \mathbf{0}, \text{ então}$$

$$\begin{aligned} \frac{1}{\phi} \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \boldsymbol{\mu}) - \lambda \mathbf{K}\mathbf{f} &= 0 \\ \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})[\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f})] &= \phi \lambda \mathbf{K}\mathbf{f} \\ \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{y} - \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{X}\boldsymbol{\beta} - \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{N}\mathbf{f} &= \phi \lambda \mathbf{K}\mathbf{f} \\ (\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{N} + \phi \lambda \mathbf{K})\mathbf{f} &= \mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \mathbf{f} &= (\mathbf{N}^T \mathbf{D}(\mathbf{v}) \mathbf{D}(\boldsymbol{\xi})\mathbf{N} + \phi \lambda \mathbf{K})^{-1} \mathbf{N}^T \mathbf{D}(\mathbf{v}) \\ &\quad \mathbf{D}(\boldsymbol{\xi})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

e conseqüentemente ao seguinte algoritmo de *backfitting* com a matriz de peso dado por  $\mathbf{D}(\mathbf{v})$  é

$$\mathbf{f}^{(r+1)} = (\mathbf{N}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)})\mathbf{N} + \lambda \phi^{(r)} \mathbf{K})^{-1} \mathbf{N}^T \mathbf{D}(\mathbf{v}^{(r)}) \mathbf{D}(\boldsymbol{\xi}^{(r)})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(r+1)}).$$

E finalmente a matriz da função escore em relação a  $\phi$

$$\text{Temos que } U_{\phi}^p(\boldsymbol{\theta}) = -\frac{n}{2\phi} + \frac{(\mathbf{y} - \boldsymbol{\mu})^T}{2\phi^2} \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu}). \text{ Seja } U_{\phi}^p(\boldsymbol{\theta}) = 0, \text{ então}$$

$$\begin{aligned}
-\frac{n}{2\phi} + \frac{(\mathbf{y} - \boldsymbol{\mu})^T}{2\phi^2} \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu}) &= 0 \\
\frac{n}{2\phi} &= \frac{(\mathbf{y} - \boldsymbol{\mu})^T}{2\phi^2} \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu}) \\
\phi n &= (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu})
\end{aligned}$$

assim o seguinte algoritmo de *backfitting* com a matriz de peso dado por  $\mathbf{D}(\mathbf{v})$  é

$$\phi^{(r+1)} = \frac{1}{n} (\mathbf{y} - \boldsymbol{\mu}^{(r)})^T \mathbf{D}(\mathbf{v}^{(r)}) (\mathbf{y} - \boldsymbol{\mu}^{(r)}).$$

### Influência local

Ponderação de casos

$$\begin{aligned}
\frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T} &= \frac{\partial}{\partial \boldsymbol{\omega}} \sum_{i=1}^n \omega_i \left\{ W_h(u_i) \left[ -\frac{2(y - \mu_i)}{\phi} \left( \frac{dg^{-1}(\eta_i)}{d\eta_i} \right) \mathbf{x}_i \right] \right\} \\
&= \sum_{i=1}^n e_{in} \left\{ W_h(u_i) \left[ -\frac{2(y - \mu_i)}{\phi} \dot{a} \mathbf{x}_i \right] \right\} \\
&= \frac{1}{\phi} \sum_{i=1}^n v_i (y - \mu_i) \dot{a}_i \mathbf{x}_i e_{in}^T.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \boldsymbol{\omega}^T} &= \frac{\partial}{\partial \boldsymbol{\omega}} \sum_{i=1}^n \omega_i \left\{ W_h(u_i) \left[ -\frac{2(y - \mu_i)}{\phi} \left( \frac{dg^{-1}(\eta_i)}{d\eta_i} \right) \mathbf{n}_i \right] \right\} - \lambda \mathbf{K} \mathbf{f} \\
&= \sum_{i=1}^n e_{in} \left\{ W_h(u_i) \left[ -\frac{2(y - \mu_i)}{\phi} \dot{a} \mathbf{n}_i \right] \right\} \\
&= \frac{1}{\phi} \sum_{i=1}^n v_i (y - \mu_i) \dot{a}_i \mathbf{n}_i e_{in}^T.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^T} &= \frac{\partial}{\partial \boldsymbol{\omega}} \left[ -\frac{1}{2\phi} \sum_{i=1}^n \omega_i + \frac{1}{2\phi^2} \sum_{i=1}^n \omega_i [-2W_h(u_i)(y - \mu_i)^2] \right] \\
&= -\frac{1}{2\phi} \sum_{i=1}^n e_{in}^T + \frac{1}{2\phi^2} \sum_{i=1}^n [-2W_h(u_i)(y - \mu_i)^2] e_{in}^T \\
&= -\frac{1}{2\phi} \sum_{i=1}^n e_{in}^T + \frac{1}{2\phi^2} \sum_{i=1}^n v_i (y - \mu_i)^2 e_{in}^T.
\end{aligned}$$

Perturbação na variável resposta

$$\begin{aligned}
\frac{\partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[ \frac{2}{\phi} \sum_{i=1}^n W_h(u_{i\boldsymbol{\omega}}) (y_{i\boldsymbol{\omega}} - \mu_i) \mathbf{e}_{in}^T \right] \\
&= \frac{2}{\phi} \sum_{i=1}^n [-2W'_h(u_{i\boldsymbol{\omega}}) u_{i\boldsymbol{\omega}} \dot{a}_i \mathbf{x}_i - W_h(u_{i\boldsymbol{\omega}}) \dot{a}_i \mathbf{x}_i] \mathbf{e}_{in}^T \\
&= \frac{1}{\phi} \sum_{i=1}^n [v_{i\boldsymbol{\omega}} - 4W'_h(u_{i\boldsymbol{\omega}}) u_{i\boldsymbol{\omega}}] \dot{a}_i \mathbf{x}_i \mathbf{e}_{in}^T.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \mathbf{f} \partial \boldsymbol{\omega}^T} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[ \frac{2}{\phi} \sum_{i=1}^n W_h(u_{i\boldsymbol{\omega}}) (y_{i\boldsymbol{\omega}} - \mu_i) \mathbf{e}_{in}^T \right] \\
&= \frac{2}{\phi} \sum_{i=1}^n [-2W'_h(u_{i\boldsymbol{\omega}}) u_{i\boldsymbol{\omega}} \dot{a}_i \mathbf{n}_i - W_h(u_{i\boldsymbol{\omega}}) \dot{a}_i \mathbf{n}_i] \mathbf{e}_{in}^T \\
&= \frac{1}{\phi} \sum_{i=1}^n [v_{i\boldsymbol{\omega}} - 4W'_h(u_{i\boldsymbol{\omega}}) u_{i\boldsymbol{\omega}}] \dot{a}_i \mathbf{n}_i \mathbf{e}_{in}^T.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L_p(\boldsymbol{\theta}, \lambda | \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^T} &= \frac{\partial}{\partial \phi} \left[ \frac{2}{\phi} \sum_{i=1}^n W_h(u_{i\boldsymbol{\omega}}) (y_{i\boldsymbol{\omega}} - \mu_i) \mathbf{e}_{in}^T \right] \\
&= \frac{1}{\phi^2} \sum_{i=1}^n \left\{ -2W'_h(u_{i\boldsymbol{\omega}}) u_{i\boldsymbol{\omega}} - 2W_h(u_{i\boldsymbol{\omega}}) \right\} (y_{i\boldsymbol{\omega}} - \mu_i) \mathbf{e}_{in}^T \\
&= \frac{1}{\phi^2} \sum_{i=1}^n [v_{i\boldsymbol{\omega}} - 2W'_h(u_{i\boldsymbol{\omega}}) u_{i\boldsymbol{\omega}}] (y_{i\boldsymbol{\omega}} - \mu_i) \mathbf{e}_{in}^T.
\end{aligned}$$

**Função  $W_h$  para as funções geradoras de densidade Hiperbólica, Exponencial potência e t-Student.**

Calculando  $W_h = \frac{h'(u)}{h(u)}$  para a função geradora de densidade Hiperbólica.

$$h(u) = \frac{1}{2} \frac{e^{-\nu(\sqrt{1+u})}}{K_1(\nu)}$$

Derivando

$$\begin{aligned}
h'(u) &= \frac{1}{2K_1(\nu)} e^{-\nu(\sqrt{1+u})} \left( -\nu(1+u)^{\frac{1}{2}} \right)' \\
h'(u) &= \frac{1}{2K_1(\nu)} e^{-\nu(\sqrt{1+u})} \left( -\frac{1}{2}\nu(1+u)^{\frac{1}{2}-1} \right) \\
h'(u) &= \frac{1}{2K_1(\nu)} e^{-\nu(\sqrt{1+u})} \left( -\frac{1}{2}\nu(1+u)^{-\frac{1}{2}} \right) \\
h'(u) &= \frac{1}{2K_1(\nu)} e^{-\nu(\sqrt{1+u})} \left( -\frac{\nu}{2\sqrt{1+u}} \right)
\end{aligned}$$

Logo temos

$$\frac{h'(u)}{h(u)} = \frac{\left( \frac{1}{2K_1(\nu)} e^{-\nu(\sqrt{1+u})} \left( -\frac{\nu}{2\sqrt{1+u}} \right) \right)}{\frac{1}{2K_1(\nu)} e^{-\nu(\sqrt{1+u})}}$$

Portanto

$$\frac{h'(u)}{h(u)} = -\frac{\nu}{2\sqrt{1+u}} = W_h(u).$$

Calculando  $W_h = \frac{h'(u)}{h(u)}$  para a função geradora de densidade Exponencial potência.

$$h(u) = \frac{2^{\frac{3}{2} + \frac{\nu}{2}} e^{-\frac{u}{2} \frac{1}{1+\nu}}}{\Gamma\left(3 + \frac{\nu}{2}\right)}$$

Derivando

$$h'(u) = \Gamma\left(3 + \frac{\nu}{2}\right)^{-1} 2^{\frac{3}{2} + \frac{\nu}{2}} e^{-\frac{u}{2} \frac{1}{1+\nu}} \left[ -\frac{u}{2} \frac{1}{1+\nu} \right]'$$

$$h'(u) = \Gamma\left(3 + \frac{\nu}{2}\right)^{-1} 2^{\frac{3}{2} + \frac{\nu}{2}} e^{-\frac{u}{2} \frac{1}{1+\nu}} \left[ \left( \frac{1}{1+\nu} \right) \left( -\frac{u}{2} \right)^{-\left(\frac{1}{1+\nu}\right)} \right]$$

$$h'(u) = \Gamma\left(3 + \frac{\nu}{2}\right)^{-1} 2^{\frac{3}{2} + \frac{\nu}{2}} e^{-\frac{u}{2} \frac{1}{1+\nu}} \left[ -\frac{u^{-\left(\frac{\nu}{1+\nu}\right)}}{2(1+\nu)} \right]$$

Logo temos

$$\frac{h'(u)}{h(u)} = \frac{\Gamma\left(3 + \frac{\nu}{2}\right)^{-1} 2^{\frac{3}{2} + \frac{\nu}{2}} e^{-\frac{u}{2} \frac{1}{1+\nu}} \left[ -\frac{u^{-\left(\frac{\nu}{1+\nu}\right)}}{2(1+\nu)} \right]}{\Gamma\left(3 + \frac{\nu}{2}\right)^{-1} 2^{\frac{3}{2} + \frac{\nu}{2}} e^{-\frac{u}{2} \frac{1}{1+\nu}}}$$

Portanto

$$\frac{h'(u)}{h(u)} = -\frac{u^{-\left(\frac{\nu}{1+\nu}\right)}}{2(1+\nu)} = W_h(u)$$

Calculando  $W_h = \frac{h'(u)}{h(u)}$  para a função geradora de densidade t-Student

$$h(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+1)}{2}}$$

Derivando

$$h'(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right) \left[ -\frac{1}{2}(\nu+1) \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+1)}{2}-1} \left(1 + \frac{u}{\nu}\right)' \right]}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}$$

$$h'(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right) \left[ -\frac{1}{2}(\nu+1) \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+3)}{2}} \left(\frac{1}{\nu}\right) \right]}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}$$

$$h'(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right) \left[ -\frac{\nu+1}{2\nu} \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+3)}{2}} \right]}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}$$

Logo temos

$$\frac{h'(u)}{h(u)} = \frac{\frac{\Gamma\left(\frac{\nu+1}{2}\right) \left[ -\frac{\nu+1}{2\nu} \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+3)}{2}} \right]}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}}{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+1)}{2}}}$$

Portanto

$$\frac{h'(u)}{h(u)} = \left[ -\frac{(\nu+1)}{2\nu} \left(1 + \frac{u}{\nu}\right)^{\left(\left(-\frac{\nu}{2} + \frac{\nu}{2}\right) + \left(-\frac{3}{2} + \frac{1}{2}\right)\right)} \right]$$

$$\frac{h'(u)}{h(u)} = -\frac{(1+\nu)}{2\nu} \left(\frac{\nu+u}{\nu}\right)^{-1}$$

$$\frac{h'(u)}{h(u)} = -\frac{(1+\nu)}{2\nu} \left(\frac{\nu}{\nu+u}\right)$$

$$\frac{h'(u)}{h(u)} = -\frac{(1+\nu)}{2(\nu+u)} = W_h(u)$$

## Apêndice B

### Códigos realizados no R:

Matriz de incidência.

```
matrizN<- function(t){ n<- length(t);ti<- sort(unique(t));r<- length(ti)
N<- matrix(0,n,r)
for (i in 1:n){ for(j in 1:r){ N[i, j] <- ifelse(t[i]==ti[j], 1,0) }}
N}
```

A seguir apresentamos os códigos R para ajustar os modelos propostos no Capítulo 4 com base no conjunto de dados da área financeira do Chile.

```

rm(list=ls(all=TRUE))
require(xlsx)
library(gamlss)
dados1<-read.xlsx("dados.xlsx",1)
#grafico dispersao dos dados
par(mfrow=c(1,2))
#y vs x
plot(x,y,pch=20,main="(1)",col="3",xlab="IPSA",
ylab="Rentabilidades da AFP Cuprum")
#y vs t
plot(t,y,pch=20,main="(2)",col="3",xlab="tempo",
ylab="Rentabilidades da AFP Cuprum")
#####
#_Função de ligação identidade _____
#####
#_____Normal_____
Normal_identidade<- gamlss(y~x+cs(t),
family=NO(mu.link = identity,sigma.link = identity),
data=dados1, trace=FALSE)

summary(Normal_identidade)
#_____t_Student_____
tStudent_identidade<- gamlss(y~x+cs(t),
family=TF(mu.link = identity,sigma.link = identity),
data=dados1, trace=FALSE)

summary(tStudent_identidade)
#_____Exponencial potência_____
Exponencial_potencia_identidade<- gamlss(y~x+cs(t),
family=PE(mu.link = identity,sigma.link = identity),
nu.start = 1, data=dados1, trace=TRUE,method=mixed(50,50))

summary(Exponencial_potencia_identidade)#resumo

#_Estimativas de $\nu$_____
exp(tStudent_identidade$nu.coefficients)
exp(Exponencial_potencia_identidade$nu.coefficients)

#_Estimativas do parâmetro de suavização e
#as estimativasos graus de liberdade efetivos.

paste("lambda=",round(getSmo(Normal_identidade)$lambda1,5))
paste("df=",round(sum(getSmo(Normal_identidade)$lev),5))
paste("lambda=",round(getSmo(tStudent_identidade)$lambda1,5))
paste("df=",round(sum(getSmo(tStudent_identidade)$lev),5))
paste("lambda=",round(getSmo(Exponencial_potencia_identidade)$lambda1,5))
paste("df=",round(sum(getSmo(Exponencial_potencia_identidade)$lev),5))

```

```

#Gráfico.
par(mfrow=c(3,1))

term.plot(Normal_identidade, main="Normal",what='mu',terms=2)

term.plot(tStudent_identidade, main="t-Student",what='mu',terms=2)

term.plot(Exponencial_potencia_identidade, main="Exponencial potência",what='mu',terms=2)

#Gráfico do resíduo quantilico
#Gráfico do gamlss
plot(Normal_identidade)

plot(tStudent_identidade)

plot(Exponencial_potencia_identidade)

#####
#__Função de ligação inversa na média_____
#####

#_____Normal_____
Normal_inversa<- gamlss(y~x+cs(t),
family=NO(mu.link = inverse,sigma.link = identity),
data=dados1, trace=FALSE)

summary(Normal_inversa)

#_____t-Student_____
tStudent_inversa<- gamlss(y~x+cs(t),
family=TF(mu.link = inverse,sigma.link = identity),
data=dados1, trace=FALSE)

summary(tStudent_inversa)

#_____Exponencial potência_____
Exponencial_potencia_inversa<- gamlss(y~x+cs(t),
family=PE(mu.link = inverse,sigma.link = identity),
nu.start = 0.7,data=dados1, trace=TRUE,method=mixed(50,50))

summary(Exponencial_potencia_inversa)

```

```

#Estimativas de $\nu$
exp(tStudent_inversa$nu.coefficients)

exp(Exponencial_potencia_inversa$nu.coefficients)

#_Estimativas do parâmetro de suavização e
#as estimativasos graus de liberdade efetivos

paste("lambda=",round(getSmo(Normal_inversa)$lambda1,5))
paste("df=",round(sum(getSmo(Normal_inversa)$lev),5))
paste("lambda=",round(getSmo(tStudent_inversa)$lambda1,5))
paste("df=",round(sum(getSmo(tStudent_inversa)$lev),5))
paste("lambda=",round(getSmo(Exponencial_potencia_inversa)$lambda1,5))
paste("df=",round(sum(getSmo(Exponencial_potencia_inversa)$lev),5))

#Gráfico.
par(mfrow=c(3,1))

term.plot(Normal_inversa, main="Normal",what='mu',terms=2)

term.plot(tStudent_inversa, main="t-Student",what='mu',terms=2)

term.plot(Exponencial_potencia_inversa, main="Exponencial potência",what='mu',terms=2)

#Gráfico do resíduo quantilico
wp(Normal_inversa)

wp(tStudent_inversa)

wp(Exponencial_potencia_inversa)

AIC(Normal_identidade,
tStudent_identidade,
Exponencial_potencia_identidade,
Normal_inversa,
tStudent_inversa,
Exponencial_potencia_inversa)

```