UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA

Camila Galhardo Toledo

Modelos Aditivos Parcialmente Lineares Multivariados

Juiz de Fora
2025

Camila Galhardo Toledo

Modelos Aditivos Parcialmente Lineares Multivariados

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de bacharel em Estatística.

Orientador: Prof. Dr. Clécio da Silva Ferreira

Juiz de Fora

2025

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Toledo, Camila Galhardo.
Modelos Aditivos Parcialmente Lineares Multivariados / Camila Galhardo Toledo. – 2025.
38 f. : il.

Orientador: Clécio da Silva Ferreira
Trabalho de Conclusão de Curso (graduação) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Departamento de Estatística, 2025.

1. Palavra-chave. 2. Palavra-chave. 3. Palavra-chave. I. Sobrenome, Nome do orientador, orient. II. Título.

Camila Galhardo Toledo

Modelos Aditivos Parcialmente Lineares Multivariados

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de bacharel em Estatística.

Aprovada em (dia) de (mês) de (ano)

BANCA EXAMINADORA

Prof. Dr. Clécio da Silva Ferreira - Orientador
Universidade Federal de Juiz de Fora

Profa. Dra. Camila Borelli Zeller
Universidade Federal de Juiz de Fora

Prof. Dr. Tiago Maia Magalhães
Universidade Federal de Juiz de Fora

Dedico este trabalho aos meus pais.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que esteve comigo em todos os momentos, sustentando-me e guiando o meu caminho. Agradeço aos meus pais, Maria e Almir, que nunca colocaram empecilhos para que eu alcançasse meus sonhos. Pelo contrário, sempre foram incentivadores e apoiadores de tudo que acreditavam ser o melhor para os seus filhos.

Sou igualmente grato aos meus irmãos, que sempre me ajudaram no que foi necessário (e, às vezes, me perturbaram quando não era, rs), e que me inspiraram a estudar e lutar pelos meus objetivos.

Aos meus amigos Arthur, Gustavo, Joysce, Natasha e Pedro, minha gratidão por terem caminhado comigo pelas disciplinas, trabalhos, idas ao RU, dividindo os desafios, os surtos e muitas risadas. Às meninas da república, especialmente Sara e Analice, que compartilharam o dia a dia comigo sob o mesmo teto, tornando-se, assim, como uma família para mim.

Agradeço também ao meu grande amor, Patryck, por todo o apoio e por fazer da minha vida mais leve e feliz.

Por fim, mas não menos importante, agradeço ao meu professor e orientador, Clécio, que foi um mestre extremamente paciente e competente ao me guiar, com dedicação, na construção deste trabalho. Sem ele, certamente nada disso teria sido possível.

"Àquele que é capaz de fazer infinitamente mais do que tudo o que pedimos ou pensamos, de acordo com o seu poder que atua em nós, a ele seja a glória na igreja e em Cristo Jesus, por todas as gerações, para todo o sempre! Amém"(Efésios 3:20-21, NVI (2001))

RESUMO

No presente trabalho são apresentados os Modelos Aditivos Parcialmente Lineares Multivariados, isto é, modelos que abrangem relações tanto lineares, quanto não lineares entre a variável resposta e as variáveis explicativas. Como é um modelo multivariado, temos que a variável resposta é apresentada como um vetor de variáveis. O objetivo do trabalho é estudar tal modelo e implementar um algoritmo que possibilite a estimação dos parâmetros. Para descrever as componentes não paramétricas são utilizados os *B-Splines*. Afim de obter os estimadores dos parâmetros do modelo, utilizamos o método de máxima verossimilhança penalizada, e os parâmetros de suavização, dado por $\boldsymbol{\lambda}$, são escolhidos através do método de Validação Cruzada. Por meio da função de verossimilhança penalizada também obtemos a matriz de informação, a qual nos dará as estimativas dos desvios-padrão dos estimadores dos parâmetros. Quando a mesma não é inversível, aplicamos o método de bootstrap. Por fim, para avaliar o modelo proposto, foram realizados estudos de simulação e uma aplicação em um conjunto de dados da edição do PISA de 2003, uma avaliação realizada a cada três anos e organizada pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) que mede o desempenho dos alunos de 15 anos em leitura, matemática e ciências.
Palavras-chave: modelos aditivos parcialmente lineares multivariados; *b-splines*; máxima verossimilhança penalizada.

ABSTRACT

In the present work, Multivariate Partially Linear Additive Models are presented, that is, models that encompass both linear and nonlinear relationships between the response variable and the explanatory variables. As it is a multivariate model, the response variable is presented as a vector of variables. The objective of the work is to study such models and implement an algorithm that allows for the estimation of the parameters. To describe the non-parametric components, B-Splines are used. In order to obtain the estimators of the model parameters, the penalized maximum likelihood method is used, and the smoothing parameters, given by $\boldsymbol{\lambda}$, are chosen through the Cross Validation method. Through the penalized likelihood function, we also obtain the information matrix, which will provide the standard deviation estimates of the parameter estimators. When the information matrix is not invertible, we apply the bootstrap method. Finally, to evaluate the proposed model, simulation studies were carried out, and an application was made to a dataset from the 2003 edition of PISA, an assessment conducted every three years by the Organisation for Economic Co-operation and Development (OECD) that measures the performance of 15-year-old students in reading, mathematics, and science.

Keywords: multivariate partially linear additive models; b-splines; penalized maximum likelihood.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Um modelo estatístico é uma representação matemática baseada em dados observados ou teoria subjacente. O mesmo é usado para descrever, entender e fazer previsões sobre o comportamento de variáveis em um sistema. Modelos estatísticos são fundamentais em várias áreas de conhecimento, e podem variar em sua complexidade.

Modelos de regressão são uma classe de modelos estatísticos que têm como objetivo estudar a relação entre uma ou mais variáveis independentes e uma variável dependente. O modelo mais simples e de extensa utilização é o modelo de regressão linear, o qual supõe que a relação entre as variáveis explicativas e a variável resposta é linear. Podemos representar tal modelo para o i-ésimo indivíduo da seguinte forma:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p + x_{ip} + \epsilon_i, i = 1, \ldots, n,$$

onde Y denota a resposta do experimento, $X_1, \ldots, X_p$ denotam as variáveis explicativas, $\beta = (\beta_0, \ldots, \beta_p)$ é o vetor dos coeficientes de regressão relativo às variáveis explicativas (ou preditores) e ϵ_i é o erro aleatório.

Em algumas situações, através de gráficos de dispersão, podemos observar uma relação não linear entre uma ou mais variáveis explicativas e a variável resposta. Nesse caso, o modelo de regressão linear não é suficiente para estudar tais variáveis, sendo necessário considerar um modelo mais geral que inclua tanto relações lineares quanto não lineares na componente sistemática do modelo. Tais modelos são denominados modelos aditivos parcialmente lineares (APLM), cuja expressão é dada por

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + f_1(t_{i1}) + \ldots + f_s(t_{is}) + \epsilon_i, i = 1, \ldots, n,$$

onde $f_k(.)$ é uma função suave univariada que quantifica o efeito da k-ésima variável explanatória t_k $(k = 1, ..., s)$ que contribui não parametricamente para a variável resposta Y_i. Modelos semiparamétricos e modelos aditivos são um caso particular de APLM. No primeiro, existe apenas uma componente não-paramétrica, enquanto os modelos aditivos são formados apenas por componentes não-paramétricas. Para maiores informações, ver Wood (2017, Cap. 6).

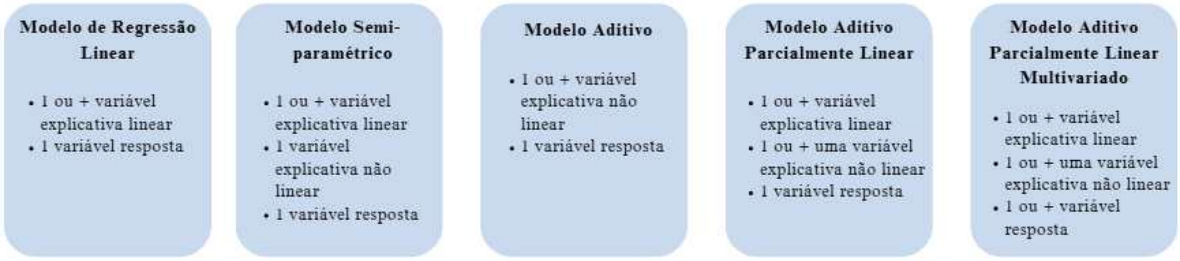
Modelos semiparamétricos têm sido aplicados em diversas situações. Ibacache-Pulgar, Paula e Cysneiros (2013) utilizam esses modelos para modelar a taxa de retorno de uma companhia de fundos de pensão do Chile através do índice de preços e tempo. Ferreira e Paula (2017) modelam a concentração de polén de Ambrosia usando como variáveis explicativas temperatura, velocidade do vento, tempo e ocorrência de chuva. Relvas e Paula (2017) modelam as temperaturas da cidade de Ubatuba (SP) correlacionando tempo e a temperatura da cidade de Cananéia (SP).

Modelos aditivos têm sido aplicados em diferentes áreas do conhecimento. Hastie e Tibshirani (1990, Cap. 4) modelam o nível de C-peptídeos através da idade e défict base. Ruppert, Wand e Carroll (2003) ajustam modelos aditivos em dados de temperatura nos Estados Unidos utilizando como variáveis explicativas não-paramétricas a latitude e a longitude. Racine, Su e Ullah (2014, Cap. 5) estimam a taxa de crescimento do produto interno bruto (PIB) entre países através da participação média do comércio e escolaridades média dos países. Faraway (2016) modela concentração de Ozônio na atmosfera através das componentes altura de base de inversão e temperatura superior de inversão.

Por fim, destacamos algumas aplicações em modelos aditivos parcialmente lineares. Liu, Wang e Liang (2011) aplicam APLM para estudar a relação entre beta-caroteno e tipos de câncer utilizando como componentes não-lineares a idade e índices de colesterol. Ibacache-Pulgar, Paula e Cysneiros (2013) modelam o preço de casas em Boston utilizando a taxa de imposto como relação linear e como não-lineares as variáveis porcentagem de menos status da população e distância ponderada a cinco centros de emprego da cidade. Wang et al. (2019) estudaram o valor médio aplicado pelo plano de saúde *Medicare* nos 50 estados dos Estados Unidos utilizando diversas variáveis demográficas de forma linear e como não lineares o número de beneficiários distintos do plano, número de serviços prestados e número de beneficiários distintos do Medicare por dias de serviço. Boente e Martinez (2023) modelam a concentração de ozônio utilizando como variáveis explicativas o mês, a temperatura, a velocidade do vento e a radiação solar.

Os modelos citados até agora tem como objetivo o estudo de somente uma variável resposta, mas, em muitos casos, vemos a necessidade do estudo de variáveis respostas correlacionadas. Tais modelos são chamados de multivariados, os quais lidam com a análise simultânea de múltiplas variáveis dependentes e independentes. Os mesmos podem capturar interações complexas que podem ser perdidas em análises univariadas, fornecendo uma compreensão mais completa do sistema.

Desse modo, no presente trabalho estudaremos os Modelos Aditivos Parcialmente Lineares Multivariados (MAPLM), extensão de um APLM, no qual a variável resposta é um vetor.



– Resumo dos modelos

O objetivo desse trabalho consiste na implementação de um algoritmo que possibilite a estimação dos parâmetros do modelo, o desenvolvimento de estudos de simulação afim de verificar as propriedades assintóticas dos estimadores de máxima verossimilhança e, por fim, o ajuste do modelo proposto em dados reais afim de ilustrar a utilidade do mesmo. Este trabalho se destaca por abordar o Modelo Aditivo Parcialmente Linear Multivariado, uma vez que, até o momento, não encontramos referências na literatura que explorem especificamente essa abordagem em um contexto multivariado.

.

2 MODELOS ADITIVOS PARCIALMENTE LINEARES MULTIVARIADOS

Modelos Aditivos Parcialmente Lineares Multivariados são uma extensão dos Modelos Aditivos Parcialmente Lineares, em que a variável resposta resposta passa a ser um vetor, permitindo, assim, o estudo de mais de uma variável dependente.

A representação geral de um MAPLM segue-se dessa forma:

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{1_{ij}} + \ldots + \beta_{pj}x_{p_{ij}} + f_{1_j}\left(t_{1_{ij}}\right) + \ldots + f_{q_j}\left(t_{q_{ij}}\right) + \epsilon_{ij}, \tag{2.1}$$

em que, temos que $i = 1, \ldots, n$ e n representa o número de observações; $j = 1, \ldots, m$, sendo m o número de variáveis resposta; Y_{ij} é a observação i da variável resposta j; $x_{1_{ij}}, x_{2_{ij}}, \ldots, x_{p_{ij}}$ são as variáveis preditoras que têm relação linear com Y_{ij}; $t_{1_{ij}}, t_{2_{ij}}, \ldots, t_{q_{ij}}$ são as variáveis preditoras que têm relação não linear com Y_{ij}; $\beta_{0_j}, \beta_{1_j}, \ldots, \beta_{p_j}$ são os coeficientes associados às variáveis $x_{1_{ij}}, x_{2_{ij}}, \ldots, x_{p_{ij}}$; $f_{1_j}\left(t_{1_{ij}}\right), ..., f_{q_j}\left(t_{q_{ij}}\right)$ são funções não lineares que modelam as relações entre as variáveis $t_{1_{ij}}, t_{2_{ij}}, ..., t_{q_{ij}}$ e as variáveis respostas; e ϵ_{ij} é o termo do erro, tal que $\epsilon_{ij} \sim N(0, \sigma_j)$.

Podemos escrever o modelo (2.1) na forma matricial

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{N}_i\boldsymbol{\gamma} + \boldsymbol{\epsilon}_i,$$

em que $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ é o vetor de variáveis respostas; $\mathbf{X}_i$ é a matriz $m \times p$ de variáveis preditoras que se relacionam linearmente com $\mathbf{Y}_i$; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ é o vetor de coeficientes associado à matriz $\mathbf{X}_i$; $\boldsymbol{N}_i$ é uma matriz $m \times l$ de incidência bloco diagonal, sendo $l = \sum_{j=1}^{m}\sum_{s=1}^{q_j} d_{j_s}$, e d_{j_s} o número de nós associado à s-ésima variável explicativa da j-ésima componente da variável resposta; $\boldsymbol{\gamma} = \left(\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_s^\top\right)^\top$ é o coeficiente associado à matriz $\boldsymbol{N}_i$, onde $\boldsymbol{\gamma}_j = \left(\gamma_{j1}, \ldots, \gamma_{jd_j}\right)^T$, $j = 1, \ldots, s$; e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im})$ é o termo do erro, tal que $\boldsymbol{\epsilon}_i \sim N_m(\mathbf{0}, \boldsymbol{\Sigma})$.

Nas seções ESTUDO DE SIMULAÇÃO e APLICAÇÃO, daremos uma visão mais detalhada da construção da matriz $\boldsymbol{N}_i$.

O principal desafio quando usamos métodos não paramétricos é a escolha da função de suavização e o parâmetro subjacente. No presente trabalho foi empregado a representação por *splines*, que será discutido na seção a seguir. A estimação dos parâmetros será apresentada na seção 4.

3 *SPLINES*

Splines são uma técnica matemática usada em estatística e análise numérica para aproximar ou suavizar uma função desconhecida, dividindo-a em segmentos menores, mais simples e suaves. Esses segmentos individuais são polinômios de baixa ordem (diferente do que é feito nos modelos polinomiais, que utilizam polinômios com alto grau para obter o ajuste) que são combinados de forma contínua para formar uma curva suave que se ajusta aos dados. Existem vários tipos de *splines*, sendo mais comuns os *splines* lineares, cúbicos e naturais. Nesse trabalho, utilizamos *B-splines* cúbicos. Segundo Wood (2017), uma *spline* cúbica é uma curva construída a partir de seções de polinômios cúbicos unidas de modo que a curva seja contínua até a segunda derivada.

3.1 *B-splines*

Os *Splines* básicos, mais conhecidos como *B-splines*, foram propostos por De Boor (1978). Esses *splines* são construídos a partir de pedaços de polinômios que se unem de forma suavizada em pontos chamados de nós, evitando que entre esses pontos ocorram mudanças abruptas na função.

Um *B-spline* de grau q possui as seguintes propriedades:

- Consiste de d+1 pedaços polinomiais, cada um de grau q.
- Os pedaços polinomiais se juntam nos d nós interiores.
- Nos pontos de junção, derivadas de ordem igual ou inferior a q-1 são contínuas.
- O *B-spline* é positivo em um domínio abrangido por d+2 nós; caso contrário é 0.
- Exceto nos limites, ele sobrepõe com $2d$ pedaços polinomiais vizinhos (d à direita e d à esquerda).

Os *B-splines*, $B_j^{(q)}(x)$, são definidos recursivamente como

$$B_j^{(0)}(x) = \begin{cases} 1, \text{ se } t_j \leq x \leq t_{j+1} \\ 0, \text{ caso contrário} \end{cases}$$

$$B_j^{(q)}(x) = \frac{x - t_j}{t_{j+q} - t_j} B_j^{(q-1)}(x) + \frac{t_{j+q+1} - x}{t_{j+q+1} - t_{j+1}} B_{j+1}^{(q-1)}(x),$$

em que $x \in [a, b]$, d é o número de nós sendo que $a < t_1, t_2, \ldots, t_d < b$, $q = 1, 2, \ldots$ é o grau dos *B-splines*, $d = l + q + 1$ é o número de nós e $B_j^{(q)}(x)$ é o valor do j-ésimo *B-spline* de grau q no ponto x para uma malha de nós equidistantes, $j = 1, \ldots, l$.

A escolha de quantos nós serão utilizados no ajuste e seus respectivos posicionamentos é muito importante, pois essa escolha determinará a qualidade do ajuste do modelo. Green e Silverman (1994) usam como nós todos os distintos pontos enquanto Eilers e Marx (1996) adotam a metodologia de escolher d pontos igualmente espaçados, com d < n. Yu e Ruppert (2002) propõe que os nós sejam equidistantes e posicionados nos quantis das variáveis preditoras. Já Vanegas e Paula (2016) definem o número de nós como $d = n^{\frac{1}{3}} + 3$, sendo estes posicionados nos quantis $\left(q\left(t, \frac{1}{d+1}\right), \ldots, q\left(t, \frac{d}{d+1}\right)\right)$ de t_i. Para mais detalhes de como selecionar o número de nós e onde posicioná-los, consultar Ruppert, Wand e Carroll (2003). Nesse trabalho, utilizamos 3 nós para cada curvatura identificada no gráfico de dispersão entre a variável resposta e a variável preditora.

3.2 Critério de penalização

A utilização de um grande número de nós pode deixar a curva obtida superajustada aos dados, isto é, a qualidade do ajuste será ótima, porém a curva terá muita variabilidade. Em contrapartida, se a curva for muito suave haverá perda de informação. Com isso, faz-se necessário a introdução de uma *penalty* que irá controlar esse *trade-off* de viés e variância, afim de encontrar o melhor ajuste do modelo aos dados. Neste trabalho, utilizamos a *penalty*

$$J\left(f_j\right) = \int_{a_j}^{b_j} f_j''\left(t\right)^2 dt, \quad j = 1, ..., s,$$

onde $f_j''\left(t\right)$ é a segunda derivada de $f_j\left(t\right)$, sendo que o intervalo $[a_j, b_j]$ contém todos os nós. Dependendo da escolha dos nós, podemos ter expressões analíticas (e matriciais) para $J\left(f_j\right)$. Green e Silverman (1994) mostram que podemos expressar a penalidade acima da seguinte forma:

$$J\left(f_j\right) = \boldsymbol{\gamma}_j^T \mathbf{K}_j \boldsymbol{\gamma}_j,$$

onde $\mathbf{K}_j$ é a matriz *penalty* obtida como função apenas dos nós.

3.3 Escolha do parâmetro de penalização via Validação Cruzada

O principal desafio quando usamos métodos não-paramétricos é a escolha da função de suavização e o parâmetro subjacente. Gu e Ma (2005a,b) propuseram validação cruzada generalizada utilizando *splines* para a função de suavização. Xu e Zhu (2005) propuseram validação cruzada para o kernel baseado em modelos com efeitos mistos não paramétricos. Manteiga et al. (2013) comparam diferentes métodos de estimação não paramétricos para estes modelos. Neste trabalho, utilizamos o método de Validação Cruzada para selecionar o valor ótimo dos parâmetros de suavização.

Essa *penalty* que será incorporada na verossimilhança terá um parâmetro que controlará a quantidade de penalização ou suavidade da curva, dado por $\boldsymbol{\lambda}$. O método de validação cruzada (ver, por exemplo, Hastie, Tibshirani e Friedman (2009)), utiliza parte dos dados para o ajuste do modelo e o restante para avaliar a adequabilidade do ajuste. Definimos esse método por

$$VC\left(\boldsymbol{\lambda}\right) = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2,$$

em que as estimativas são obtidas por meio do processo iterativo dado um valor fixo de $\boldsymbol{\lambda}$. Nesse método, o conjunto de dados é dividido em k subconjuntos, e em cada iteração k-1 subconjuntos são utilizados para treinamento, isto é, cálculo das estimativas dos parâmetros do modelo, e um subconjunto é utilizado para validação, calculando VC($\boldsymbol{\lambda}$). Esse processo se repete k vezes, cada vez utilizando um subconjunto diferente para validação. Assim, o $\boldsymbol{\lambda}$ ótimo escolhido será aquele que minimize $VC\left(\boldsymbol{\lambda}\right)$.

4 ESTIMAÇÂO POR MÁXIMA VEROSSIMILHANÇA PENALIZADA

4.1 ESTIMAÇÃO DOS PARÂMETROS

O método de máxima verossimilhança é uma das técnicas mais utilizadas para estimação de parâmetros devido às suas propriedades estatísticas desejáveis. Ele é consistente, garantindo que o estimador converge para o verdadeiro valor do parâmetro conforme o tamanho da amostra cresce, e assintoticamente normal, permitindo a realização de inferências estatísticas. Além disso, é invariante, ou seja, se $\hat{\theta}$ é o estimador de máxima verossimilhança para um parâmetro θ, então, para qualquer transformação contínua $g(\theta)$, o estimador correspondente será dado por $g(\hat{\theta})$. Isso significa que o método de máxima verossimilhança preserva as relações entre parâmetros quando há reparametrizações. Por fim, é eficiente assintoticamente, atingindo a menor variância possível entre os estimadores não viciados. Essas características fazem da máxima verossimilhança um método robusto e amplamente aplicado em modelagem estatística.

Segundo Casella e Berger (2002), se $W_1, \ldots, W_n$ são uma amostra independente e identicamente distribuída de uma população com função densidade de probabilidade $f\left(w|\theta_1, \ldots, \theta_k\right)$, a função de verossimilhança é definida por

$$L\left(\boldsymbol{\theta} \mid \mathbf{w}\right) = \prod_{i=1}^{n} f\left(w_i \mid \theta_1, \ldots, \theta_k\right),$$

e o estimador de máxima verossimilhança (EMV) é definido como o ponto do espaço paramétrico para qual a amostra observada é a mais provável. Se a função de verossimilhança é diferenciável (em θ_i), possíveis candidatos para o EMV são os valores de $\theta_1, \ldots, \theta_k$ que solucionam

$$\frac{\partial}{\partial \theta_i} L\left(\boldsymbol{\theta}|\mathbf{w}\right) = 0.$$

Pontos nos quais a primeira derivada é zero podem ser um mínimo local ou global, um máximo local ou global ou, ainda, pontos de inflexão. Nesse sentido, afim de comprovar que a solução da equação acima é um ponto de máximo global, devemos verificar que

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} L\left(\boldsymbol{\theta}|\mathbf{w}\right)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} < 0.$$

Neste trabalho, avaliamos se a matriz hessiana é definida negativa. Conforme descrito na seção 1, supomos que os erros são independentes e seguem distribuição Normal multivariada, dessa forma, obtemos a seguinte função de log-verossimilhança

$$\ell\left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}\right) = -\frac{np}{2} log\left(2\pi\right) - \frac{n}{2} log\left(|\boldsymbol{\Sigma}|\right) - \frac{1}{2} \sum_{i=1}^{n} \left(\mathbf{Y_i} - \mathbf{X_i}\boldsymbol{\beta} - \mathbf{N}_i\boldsymbol{\gamma}\right)^T \boldsymbol{\Sigma}^{-1} \left(\mathbf{Y_i} - \mathbf{X_i}\boldsymbol{\beta} - \mathbf{N}_i\boldsymbol{\gamma}\right), \tag{4.1}$$

onde $\boldsymbol{\alpha}$ é o vetor de itens da matriz $\boldsymbol{\Sigma}^{1/2}$, utilizados para efeito de cálculos. Maximizar a equação (4.1) sem impor restrições sobre as funções $f_1, ..., f_s$ pode causar sobre-ajuste ou não-identificabilidade no parâmetro $\boldsymbol{\beta}$; vide, por exemplo, Green e Silverman (1994) e Ibacache-Pulgar, Paula e Cysneiros (2013). Um procedimento bem conhecido é o de estimação de máxima verossimilhança penalizada, que consiste em incorporar em $\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$ uma função *penalty*, denotada anteriormente como $J(f_j)$.

Dessa maneira, temos que a função de log-verossimilhança penalizada é dada por

$$\ell_p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) - \frac{1}{2}\boldsymbol{\gamma}^T \mathbf{K}^* \boldsymbol{\gamma},$$

onde $\mathbf{K}^*$ é uma matriz bloco diagonal formada pelas matrizes $\mathbf{K}_1 * \lambda_1, \mathbf{K}_2 * \lambda_2, \ldots, \mathbf{K}_s * \lambda_s$ e λ_j é o parâmetro de suavização. Quando $\lambda_j \to 0$, temos que a penalização tem pouca influência no modelo, de forma que a prioridade seja o ajuste ótimo da curva. Em contrapartida, quando $\lambda_j \to \infty$ a suavidade da curva é priorizada, resultando em uma reta.

Afim de encontrar os estimadores de máxima verossimilhança para $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$, derivamos $\ell_p(\boldsymbol{\theta})$ em relação a cada um dos parâmetros, assim temos as seguintes escores

$$\frac{\partial}{\partial \boldsymbol{\beta}}\ell_p(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \boldsymbol{N}_i\boldsymbol{\gamma}),$$

$$\frac{\partial}{\partial \boldsymbol{\gamma}}\ell_p(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{N}_i^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\boldsymbol{\gamma}) - \mathbf{K}^*\boldsymbol{\gamma} \text{ e}$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}}\ell_p(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{n}\left[2\boldsymbol{\Sigma} - diag(\boldsymbol{\Sigma}) - (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\hat{\boldsymbol{\gamma}})(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\hat{\boldsymbol{\gamma}})^T\right].$$

Dessa forma, os estimadores de máxima verossimilhança penalizada são

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{N}_i\boldsymbol{\gamma}),$$

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{i=1}^{n} \mathbf{N}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{N}_i + \mathbf{K}^*\right)^{-1} \sum_{i=1}^{n} \mathbf{N}_i^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \text{ e}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{N}_i\hat{\boldsymbol{\gamma}}\right)\left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{N}_i\hat{\boldsymbol{\gamma}}\right)^T.$$

Os estimadores são obtidos através de um método iterativo, descrito pelos seguintes passos:

1. Definir os valores iniciais

   $$\hat{\boldsymbol{\beta}}^{(0)} = \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{Y}_i,$$

   dado por $\hat{\boldsymbol{\beta}}$, retirando $\boldsymbol{\Sigma}^{-1}$ e $\mathbf{N}_i\boldsymbol{\gamma}$ da equação,

   $$\hat{\boldsymbol{\gamma}}^{(0)} = \left(\sum_{i=1}^{n} \mathbf{N}_i^T \mathbf{N}_i + \mathbf{K}^*\right) \sum_{i=1}^{n} \mathbf{N}_i^T \left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}^{(0)}\right),$$

   dado por $\hat{\boldsymbol{\gamma}}$, retirando $\boldsymbol{\Sigma}^{-1}$ da equação e utilizando $\hat{\boldsymbol{\beta}}^{(0)}$ no lugar de $\hat{\boldsymbol{\beta}}$ e

$$\hat{\Sigma}^{(0)} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}^{(0)} - \mathbf{N}_i\hat{\boldsymbol{\gamma}}^{(0)}\right)\left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}^{(0)} - \mathbf{N}_i\hat{\boldsymbol{\gamma}}^{(0)}\right)^T,$$

dado por $\hat{\boldsymbol{\Sigma}}$, utilizando $\hat{\boldsymbol{\beta}}^{(0)}$ e $\hat{\boldsymbol{\gamma}}^{(0)}$ no lugar de $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$.

2. Calcular os valores dos estimadores

$$\hat{\boldsymbol{\beta}}^{(k)} = \left(\sum_{i=1}^{n}\mathbf{X}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}_i - \mathbf{N}_i\boldsymbol{\gamma}\right),$$

$$\hat{\boldsymbol{\gamma}}^{(k)} = \left(\sum_{i=1}^{n}\mathbf{N}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{N}_i + \mathbf{K}^*\right)^{-1}\sum_{i=1}^{n}\mathbf{N}_i^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \text{ e}$$

$$\hat{\boldsymbol{\Sigma}}^{(k)} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{N}_i\hat{\boldsymbol{\gamma}}\right)\left(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \mathbf{N}_i\hat{\boldsymbol{\gamma}}\right)^T$$

até que a distância $\sqrt{\sum_i\left(\hat{\theta}_i^{(k+1)} - \hat{\theta}_i^{(k)}\right)^2}$ seja suficientemente pequena, por exemplo, menor que 10^{-5}, ou se 5000 iterações tenham sido realizadas.

4.2 MATRIZ DE INFORMAÇÃO DE FISHER

A matriz de informação de Fisher $\boldsymbol{I}_F(\boldsymbol{\theta})$ para uma função de log-verossimilhança diferenciável é dada por

$$\boldsymbol{I}_F(\boldsymbol{\theta}) = -E\left(\frac{\partial^2\ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^2}\right).$$

Reescrevemos a função de log-verossimilhança da seguinte forma

$$\ell_p(\boldsymbol{\theta}) = -\frac{n}{2}\Delta - \frac{1}{2}\sum_{i=1}^{n}w_i - \frac{1}{2}g,$$

sendo $\Delta = ln|\boldsymbol{\Sigma}|$, $w_i = \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\boldsymbol{\gamma}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{N}_i\boldsymbol{\gamma}\right)$ e $g = \boldsymbol{\gamma}^T\mathbf{K}^*\boldsymbol{\gamma}$.

Primeiramente, encontramos a matriz Escore, dada por

$$U(\boldsymbol{\theta}) = \frac{\partial\ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \tag{4.2}$$

$$= -\frac{n}{2}\frac{\partial\Delta}{\partial\boldsymbol{\theta}} - \frac{1}{2}\sum_{\hat{\imath}=1}^{n}\frac{\partial w_i}{\partial\boldsymbol{\theta}} - \frac{1}{2}\frac{\partial g}{\partial\boldsymbol{\theta}}. \tag{4.3}$$

Esses resultados estão disponíveis no Apêndice A.

Assim, a matriz Hessiana é dada por

$$H(\boldsymbol{\theta}) = \frac{\partial^2\ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} \tag{4.4}$$

$$= -\frac{n}{2}\frac{\partial^2\Delta}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 w_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} - \frac{1}{2}\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}. \tag{4.5}$$

Dessa forma, temos que a matriz de informação de Fisher é igual a

$$\boldsymbol{I}_F(\boldsymbol{\theta}) = \frac{n}{2}E\left(\frac{\partial^2\Delta}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) + \frac{1}{2}\sum_{i=1}^{n}E\left(\frac{\partial^2 w_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) + \frac{1}{2}E\left(\frac{\partial^2 g}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right).$$

Resolvendo os seguintes valores esperados, obtemos o resultado final

$$E\left(\frac{\partial^2 w_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right) = 2\mathbf{X}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_i,$$
$$E\left(\frac{\partial^2 w_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}^T}\right) = 2\mathbf{X}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{N}_i,$$
$$E\left(\frac{\partial^2 w_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\top}\right) = 2\mathbf{N}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{N}_i,$$
$$E\left(\frac{\partial^2 w_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}}\right) = \mathbf{0},$$
$$E\left(\frac{\partial^2 w_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\alpha}^\top}\right) = \mathbf{0},$$
$$E\left(\frac{\partial^2 g}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T}\right) = 2\mathbf{K}^* \text{ e}$$
$$E\left(\tfrac{\partial^2\Delta}{\partial\alpha_j\partial\alpha_k^T}\right) = -\operatorname{Tr}\left(\mathbf{B}^{-1}\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j\right).$$

Note que

$$\frac{\partial^2 w_i}{\partial\alpha_j\partial\alpha_k} = \boldsymbol{e}_i^T\boldsymbol{A}_{jk}\boldsymbol{e}_i,$$

onde $\boldsymbol{A}_{jk} = \mathbf{B}^{-1}(\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j\mathbf{B}^{-1} + \dot{\mathbf{B}}_j\mathbf{B}^{-1}\dot{\mathbf{B}}_k\mathbf{B}^{-1} + \dot{\mathbf{B}}_j\mathbf{B}^{-2}\dot{\mathbf{B}}_k + \dot{\mathbf{B}}_k\mathbf{B}^{-2}\dot{\mathbf{B}}_j + \mathbf{B}^{-1}\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j$ $+\mathbf{B}^{-1}\dot{\mathbf{B}}_j\mathbf{B}^{-1}\dot{\mathbf{B}}_k)\mathbf{B}^{-1}$, tal que $\dot{\boldsymbol{B}}_k$ é uma matriz de zeros $m \times m$ com 1 na posição da matriz $\boldsymbol{B}(\boldsymbol{\alpha})$ relativa a α_k.

Teorema 1 *Seja $\boldsymbol{X}$ um vetor de variáveis aleatórias com vetor de médias $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{V}$, positiva definida. Seja também uma matriz $\boldsymbol{A}$, simétrica, de constantes reais. Então,*

$$E[\boldsymbol{X'AX}] = \operatorname{Tr}(\boldsymbol{AV}) + \boldsymbol{\mu'A\mu}.$$

Este resultado é encontrado em Montgomery et al. (2012, Pg. 580).

Pelo Teorema 1 e dado que $E(\boldsymbol{e}_i) = \mathbf{0}$, temos

$$E(\boldsymbol{e}_i\boldsymbol{A}_{jk}\boldsymbol{e}_i) = tr(\boldsymbol{A}_{jk}Cov(\boldsymbol{e}_i)) \quad (4.6)$$
$$= tr(\boldsymbol{A}_{jk}\boldsymbol{\Sigma}) \quad (4.7)$$

Assim,

$$\boldsymbol{I}_F(\boldsymbol{\theta}) = \sum_{i=1}^{n} \begin{pmatrix} \mathbf{X}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_i & X_i^T \boldsymbol{\Sigma}^{-1} \mathbf{N}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_i \boldsymbol{\Sigma}^{-1} \mathbf{N}_i + \mathbf{K}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{\boldsymbol{B}} \end{pmatrix},$$

onde $I_{\boldsymbol{B}} = \begin{pmatrix} I_{\alpha_1\alpha_1} & \cdots & I_{\alpha_1\alpha_{p_0}} \\ \vdots & \ddots & \vdots \\ I_{\alpha_{p_0}\alpha_1} & \cdots & I_{\alpha_{p_0}\alpha_{p_0}} \end{pmatrix}$ e $I_{\alpha_j\alpha_k} = \frac{1}{n}\left(tr\left(\mathbf{B^{-1}}\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j\right) + tr\left(A_{jk}\boldsymbol{\Sigma}\right)\right)$, $j, k = 1, \ldots, p_0$, onde $p_0 = m(m+1)/2$.

Para mais informações, ver Apêndice A.

Nos casos em que a matriz de informação de Fisher não é inversível, uma outra abordagem é utilizada para o cálculo dos erros padrão das estimativas dos parâmetros, o *bootstrap* paramétrico. Utilizando esta técnica, são feitas reamostragens dos erros aleatórios sob os parâmetros estimados e a partir das estimativas obtidas nestas reamostragens, são calculados os erros padrão de cada parâmetro. Para maiores informações sobre a técnica *bootstrap*, ver Davison e Hinkley (1997), por exemplo.

4.3 ANÁLISE DE RESÍDUOS

A análise de resíduos é uma técnica estatística usada para avaliar a adequação de um modelo aos dados observados. Ela envolve a investigação dos resíduos do modelo, que são as diferenças entre os valores observados e os valores preditos da variável resposta, chamado resíduo ordinário. Neste trabalho utilizamos os resíduos ordinários padronizados, denominados de distâncias de Mahalanobis, sendo calculados da seguinte forma

$$r_i = (\mathbf{y}_i - \hat{\mathbf{y}}_i)^T \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{y}_i - \hat{\mathbf{y}}_i)$$

onde r_i segue distribuição $\chi^2_{(m)}$.

Utilizamos a distância de Mahalanobis para construção de gráficos envelopes simulados afim de verificar a suposição de normalidade dos erros. Além disso, plotamos os valores de r_i adicionando um ponto de corte dado pelo quantil $100(1-\alpha)\,\%$ de uma $\chi^2_{(m)}$, para verificar possíveis pontos influentes ou mal ajustados.

5 ESTUDO DE SIMULAÇÃO

Este estudo tem como objetivo a estimação por Máxima Verossimilhança Penalizada (EMVP) de um Modelo Aditivo Parcialmente Linear Multivariado considerando o modelo (1) utilizando a teoria de *B-Splines*. A simulação foi realizada através do método de Monte Carlo com 1000 replicações para os tamanhos amostrais iguais a 500, 1000 e 2000. O modelo simulado é composto de duas variáveis explicativas não lineares e quatro variáveis explicativas lineares. Isto é,

$$\begin{aligned} Y_{1i} &= \beta_1 x_{2i} + \beta_2 x_{3i} + \beta_3 x_{4i} + f_1\left(x_{1_i}\right) + \epsilon_{1i}, \\ Y_{2i} &= \beta_4 x_{1i} + \beta_5 x_{3i} + \beta_6 x_{4i} + f_2\left(x_{2_i}\right) + \epsilon_{2i}, \\ Y_{3i} &= \beta_7 x_{1i} + \beta_8 x_{2i} + \beta_9 x_{3i} + \beta_{10} x_{4i} + \epsilon_{3i}, \end{aligned}$$

onde $i = 1, ..., n$, x_1 é uma sequência igualmente espaçada no intervalo $(0, 2\pi)$, x_2 é uma sequência igualmente espaçada no intervalo $(0.6, 1.6)$, $x_3 \sim N(0,1)$, $x_4 \sim U(0,1)$, $f_1(x_1) = cos(x_1)$, $f_2(x_2) = cos(4\pi x_2)\, e^{-x_2^2/2}$, $\boldsymbol{\beta} \sim N(0,1)$, $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\top \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Destacamos que uma mesma variável pode se relacionar linearmente com uma das variáveis resposta e não linearmente com uma outra variável resposta. Os valores de $\boldsymbol{\alpha}$, componentes da matriz $\boldsymbol{\Sigma}^{1/2}$, são gerados usando a função *genPositiveDefMat* do pacote **clusterGeneration**. Ademais, utilizamos 12 nós para ambas as componentes não lineares do modelo.

A matriz de incidência $\mathbf{N}_i$ e a matriz $\mathbf{X}_i$ são construídas da seguinte maneira

$$\mathbf{N}_i = \begin{pmatrix} \boldsymbol{n}_{i1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{n}_{i2} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{5.1}$$

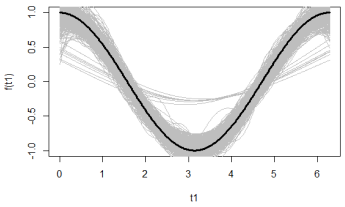
e

$$\mathbf{X}_i = \begin{pmatrix} x_{i2} & x_{i3} & x_{i4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{i1} & x_{i3} & x_{i4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_{i1} & x_{i2} & x_{i3} & x_{i4} \end{pmatrix}. \tag{5.2}$$

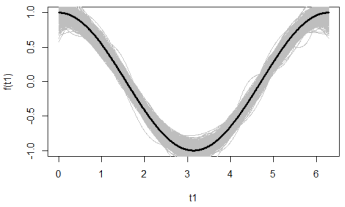
Cada linha das matrizes correspondem a cada variável resposta e as colunas correpondem às variáveis explicativas.

As Figuras 1, 2 e 3 abaixo mostram as 1000 curvas estimadas (em cinza) e a curva real $f_1(x_1)$ (em preto) para cada um dos três tamanhos amostrais.

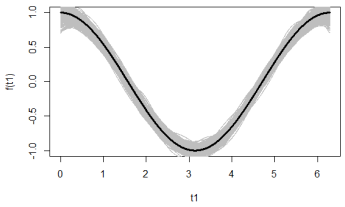
As Figuras 4, 5 e 6 abaixo exibem as 1000 curvas estimadas (em cinza) juntamente com a curva real $f_2(x_2)$ (em preto) para cada um dos três tamanhos amostrais.



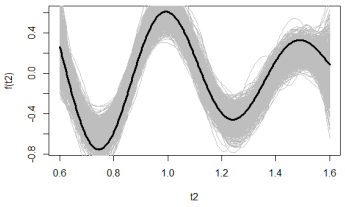
– Figura 1: Gráfico das curvas estimadas e da curva real $f_1(x_1) = cos(x_1)$ para n = 500.



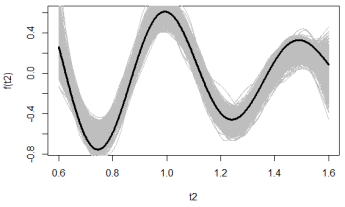
– Figura 2: Gráfico das curvas estimadas e da curva real $f_1(x_1) = cos(x_1)$ para n = 1000.



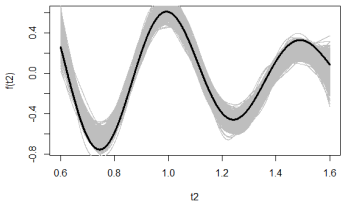
– Figura 3: Gráfico das curvas estimadas e da curva real $f_1(x_1) = cos(x_1)$ para n = 2000.



– Figura 4: Gráfico das curvas estimadas e da curva real $f_2(x_2) = cos(4\pi x_2)\, e^{-x_2^2/2}$ para n = 500.



– Figura 5: Gráfico das curvas estimadas e da curva real $f_2(x_2) = cos(4\pi x_2)\, e^{-x_2^2/2}$ para n = 1000.



– Figura 6: Gráfico das curvas estimadas e da curva real $f_2(x_2) = cos(4\pi x_2)\, e^{-x_2^2/2}$ para n = 2000.

Observa-se nas Figuras 1 a 6 que, em ambas as curvas, conforme o valor de n aumenta, as curvas estimadas convergem para a curva real. Esse comportamento indica um desempenho adequado do modelo.

Pela Tabela 1, nota-se que, conforme o valor amostral n aumenta, os vieses, dados por $\theta_k - \frac{\sum_{i=1}^{1000} \hat{\theta}_{ki}}{1000}$, e os desvios padrão das estimativas, dados por $\sqrt{\frac{\sum_{i=1}^{1000} \left(\hat{\theta}_{ki} - \frac{\sum_{i=1}^{1000} \hat{\theta}_{ki}}{1000} \right)^2}{1000}}$, diminuem.

Tabela 1 – Vieses e desvios padrão das estimativas.

		n = 500		n = 1000		n = 2000	
	Valor Real	Viés	DP	Viés	DP	Viés	DP
α_1	0.851	0.008	0.028	0.004	0.019	0.002	0.013
α_2	0.050	0.001	0.019	0.001	0.013	0.001	0.009
α_3	-0.006	-0.001	0.021	-0.001	0.015	0.000	0.010
α_4	0.859	0.008	0.028	0.003	0.019	0.000	0.013
α_5	-0.001	0.000	0.021	0.000	0.015	0.000	0.010
α_6	0.999	0.004	0.032	0.003	0.021	0.001	0.016
β_1	2.287	-0.003	0.060	-0.001	0.042	0.021	0.032
β_2	-1.197	-0.026	0.037	-0.006	0.028	-0.026	0.020
β_3	-0.694	-0.004	0.117	-0.007	0.085	-0.055	0.062
β_4	-0.412	-0.034	0.015	-0.034	0.011	-0.036	0.008
β_5	-0.971	0.009	0.039	0.016	0.027	0.002	0.019
β_6	-0.947	0.213	0.100	0.215	0.070	0.225	0.052
β_7	0.748	0.000	0.053	0.000	0.036	0.000	0.026
β_8	-0.117	0.001	0.204	0.000	0.140	-0.001	0.098
β_9	0.153	0.002	0.043	0.001	0.032	-0.001	0.022
β_{10}	2.190	-0.008	0.152	-0.004	0.110	0.000	0.076

Tabela 2 – Média dos parâmetros de suavização.

	n = 500	n = 1000	n = 2000
λ_1	0.128	0.0332	0.0204
λ_2	0.008	0.006	0.004

6 APLICAÇÃO

A aplicação do modelo MAPLM foi realizada em uma amostra de 1795 observações do PISA de 2003 referente ao Brasil. O Programa Internacional de Avaliação de Estudantes (PISA) é um estudo mundial realizado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) em países membros e não-membros. O objetivo é avaliar os sistemas educacionais medindo o desempenho de alunos de 15 anos em matemática, ciências e leitura.

Como variáveis resposta utilizamos o desempenho dos alunos nas três disciplinas e como variáveis explicativas selecionamos duas com relação linear com as variáveis resposta (HISEI e ESCS) e três com relação não linear com as variáveis resposta (RATCOMP, RMHMWK e PROPQPED). Isto é,

$$MATEMATICA_i = \beta_1 HISEI_i + \beta_2 ESCS_i + f_1\left(RATCOMP_i\right) + f_2\left(RMHMWK_i\right) + f_3\left(PROPQPED_i\right) + \epsilon_{1i},$$

$$CIENCIAS_i = \beta_1 HISEI_i + \beta_2 ESCS_i + f_1\left(RATCOMP_i\right) + f_2\left(RMHMWK_i\right) + f_3\left(PROPQPED_i\right) + \epsilon_{2i},$$

$$LEITURA_i = \beta_1 HISEI_i + \beta_2 ESCS_i + f_1\left(RATCOMP_i\right) + f_2\left(RMHMWK_i\right) + f_3\left(PROPQPED_i\right) + \epsilon_{3i}.$$

A matriz de incidência $\mathbf{N}_i$ e a matriz $\mathbf{X}_i$ são construídas do seguinte modo:

$$\mathbf{N}_i = \begin{pmatrix} \boldsymbol{n}_{i1} & \boldsymbol{n}_{i2} & \boldsymbol{n}_{i3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{n}_{i1} & \boldsymbol{n}_{i2} & \boldsymbol{n}_{i3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{n}_{i1} & \boldsymbol{n}_{i2} & \boldsymbol{n}_{i3} \end{pmatrix} \text{ e}$$

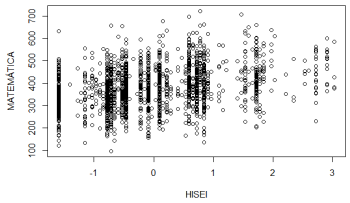
$$\mathbf{X}_i = \begin{pmatrix} x_{i1} & x_{i2} & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{i1} & x_{i2} & 0 & 0 \\ 0 & 0 & 0 & 0 & x_{i1} & x_{i2} \end{pmatrix}.$$

Cada linha das matrizes correspondem a cada variável resposta e as colunas correpondem às variáveis explicativas.

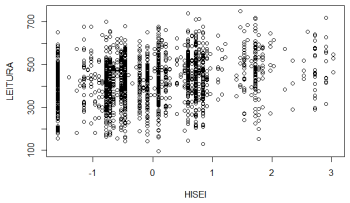
Abaixo, temos o dicionário dessas variáveis e em seguida, os gráficos de dispersão contra as variáveis respostas.

- HISEI: Mais alto índice sócio-econômico (do pai ou da mãe),
- ESCS: Índice de status cultural e sócio-econômico,
- RATCOMP: Nº total de computadores / tamanho de escola,

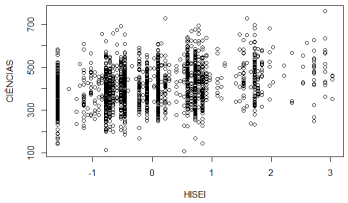
- RMHMWK: Tempo gasto com lição de casa e
- PROPQPED: Proporção de professores com ISCED 5A em Pedagogia.



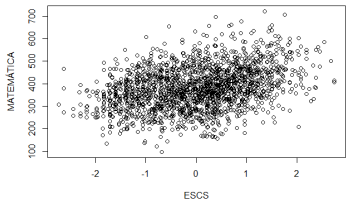
– Figura 7: Gráfico de dispersão mostrando a relação entre as variáveis *matematica* e *hisei*.



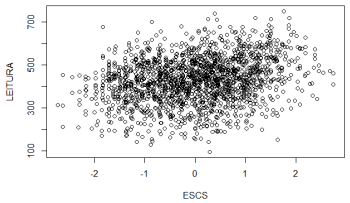
– Figura 8: Gráfico de dispersão mostrando a relação entre as variáveis *leitura* e *hisei*.



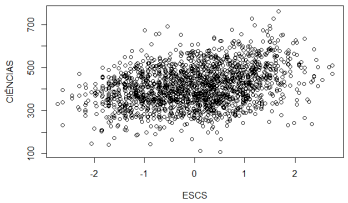
– Figura 9: Gráfico de dispersão mostrando a relação entre as variáveis *ciencias* e *hisei*.



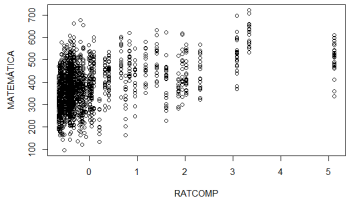
– Figura 10: Gráfico de dispersão mostrando a relação entre as variáveis *matematica* e *escs*.



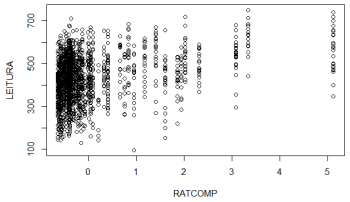
– Figura 11: Gráfico de dispersão mostrando a relação entre as variáveis *leitura* e *escs*.



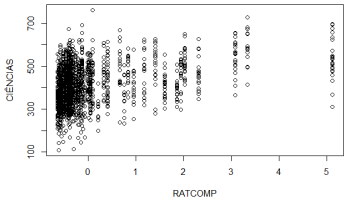
– Figura 12: Gráfico de dispersão mostrando a relação entre as variáveis *ciencias* e *escs*.



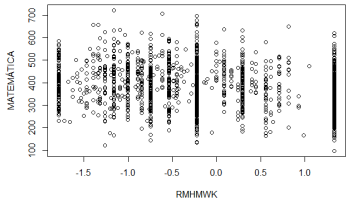
– Figura 13: Gráfico de dispersão mostrando a relação entre as variáveis *matematica* e *ratcomp*.



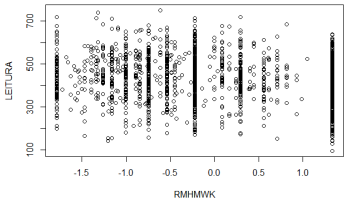
– Figura 14: Gráfico de dispersão mostrando a relação entre as variáveis *leitura* e *ratcomp*.



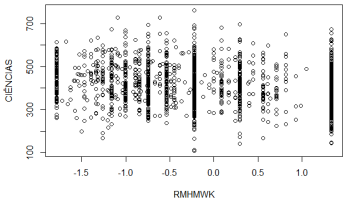
– Figura 15: Gráfico de dispersão mostrando a relação entre as variáveis *ciencias* e *ratcomp*.



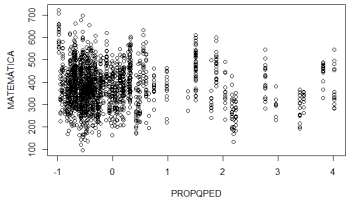
– Figura 16: Gráfico de dispersão mostrando a relação entre as variáveis *matematica* e *rmhmwk* .



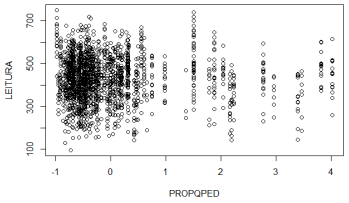
– Figura 17: Gráfico de dispersão mostrando a relação entre as variáveis *leitura* e *rmhmwk*.



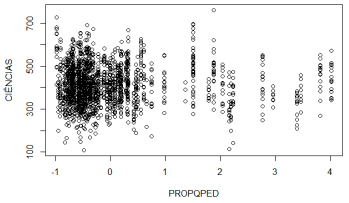
– Figura 18: Gráfico de dispersão mostrando a relação entre as variáveis *ciencias* e *rmhmwk*.



– Figura 19: Gráfico de dispersão mostrando a relação entre as variáveis *matematica* e *propqped*.



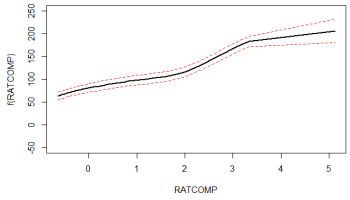
– Figura 20: Gráfico de dispersão mostrando a relação entre as variáveis *leitura* e *propqped*.



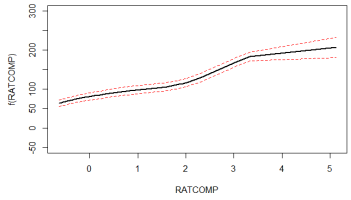
– Figura 21: Gráfico de dispersão mostrando a relação entre as variáveis *ciencias* e *propqped*.

Para a estimação das curvas, padronizamos as variáveis explicativas, utilizamos 6 nós para a variável RATCOMP e 12 para cada uma das variáveis RMHMWK e PROPQPED. Além disso, para o vetor $\boldsymbol{\lambda}$, ou seja, os parâmetros de suavização, utilizamos o valor mínimo igual a 0.00001, valor máximo igual a 0.001 e valor inicial igual a 0.0001 na função *optim* do software R Core R Core Team (2023), isto é, uma função de otimização numérica, que é usada para encontrar os valores dos parâmetros que minimizam ou maximizam uma função de objetivo. Nesse caso queremos encontrar os valores de $\boldsymbol{\lambda}$ que minimizam o valor de VC($\boldsymbol{\lambda}$), abordado na seção 3.2.

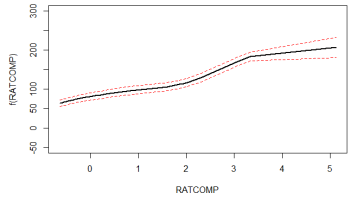
Abaixo temos os gráficos das curvas estimadas com suas respectivas bandas de confiança. A Tabela 2 das estimativas e erros padrão de $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ e a Tabela 3 das estimativas de $\boldsymbol{\lambda}$. As curvas estimadas representam bem a relação entre as variáveis respostas e as variáveis explicativas, uma vez que as mesmas se assemelham aos gráficos de dispersão apresentados acima. Além disso, os erros padrão das estimativas assumem valores baixos, mostrando a eficácia do modelo.



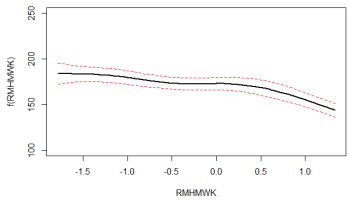
– Figura 22: Gráfico da curva estimada $f\,(ratcomp)$ para a variável resposta *matematica* .



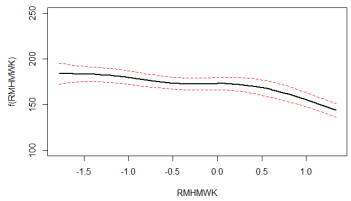
– Figura 23: Gráfico da curva estimada $f\,(ratcomp)$ para a variável resposta *leitura* .



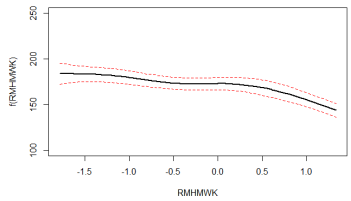
– Figura 24: Gráfico da curva estimada $f\,(ratcomp)$ para a variável resposta *ciencias* .



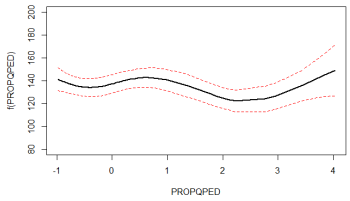
– Figura 25: Gráfico da curva estimada $f\,(rmhmwk)$ para a variável resposta *matematica* .



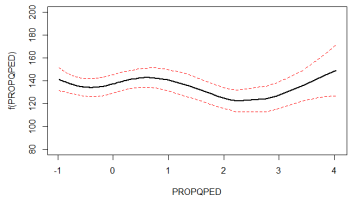
– Figura 26: Gráfico da curva estimada $f\,(rmhmwk)$ para a variável resposta *leitura*.



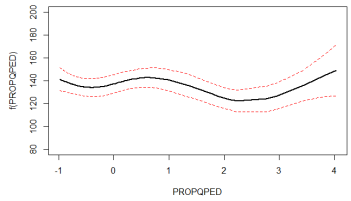
– Figura 27: Gráfico da curva estimada $f\,(rmhmwk)$ para a variável resposta *ciencias*.



– Figura 28: Gráfico da curva estimada $f\,(propqped)$ para a variável resposta *matematica*.



– Figura 29: Gráfico da curva estimada $f\,(propqped)$ para a variável resposta *leitura*.



– Figura 20: Gráfico da curva estimada $f\,(propqped)$ para a variável resposta *ciencias*.

Salienta-se que os erros padrão, nesse caso, foram obtidos via bootstrap, uma vez que não foi possível inverter a matriz de informação de Fisher.

Tabela 3 – Estimativas dos parâmetros e seus erros padrão.

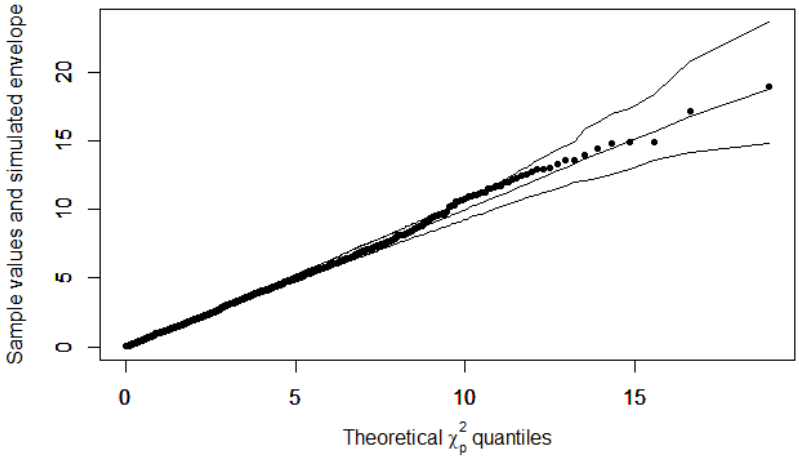
Parâmetro	Estimativa	Erro Padrão
α_1	74.996	1.278
α_2	18.539	0.929
α_3	21.349	0.923
α_4	91.917	1.499
α_5	23.746	0.937
α_6	75.426	1.361
β_1	11.693	2.800
β_2	10.425	3.036
β_3	10.727	3.695
β_4	8.299	4.076
β_5	11.070	2.902
β_6	9.941	3.421

Podemos afirmar que as variáveis preditoras com relações lineares contribuem de forma positiva para as variáveis resposta. As variáveis não lineares *ratcomp* e *propqped* possuem uma relação crescente com as variáveis resposta e a variável *rmhmwk* possui uma relação decrescente com as variáveis resposta.

Tabela 4 – Estimativas do parâmetro de suavização $\boldsymbol{\lambda}$

	HISEI	ESCS	RATCOMP	RMHMWK	PROPQPED
Matemática	-	-	0.00001	0.0001	0.0001
Leitura	-	-	0.0001	0.0001	0.0001
Ciências	-	-	0.0001	0.0001	0.0001

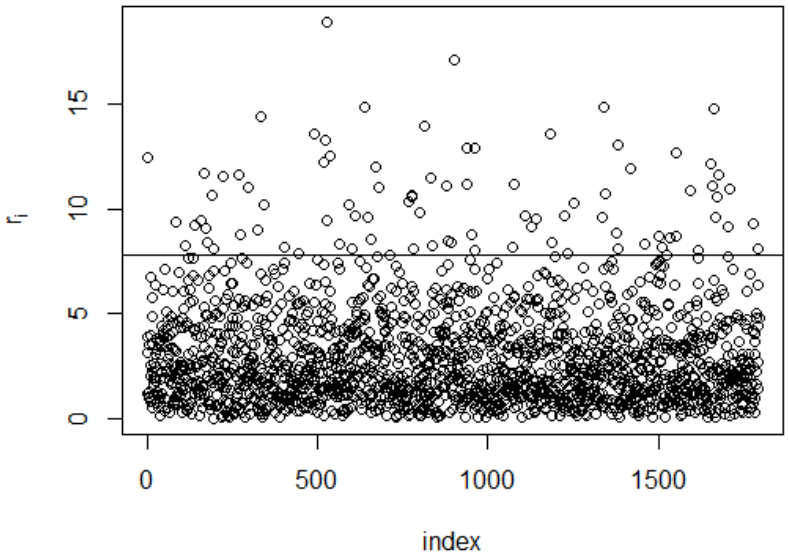
Como forma de avaliarmos o ajuste do modelo, abaixo apresentamos o gráfico envelope.



– Figura 31 - Gráfico Envelope.

Foram identificados 25 pontos fora dos limites na Figura 31. Considerando que o tamanho da amostra é de 1975 observações, isso representa cerca de 1.27% dos pontos. Esse resultado sugere que o modelo está bem ajustado aos dados.

O gráfico de resíduos apresentado na Figura 32 inclui uma linha de corte no valor crítico correspondente ao quantil 95% da distribuição qui-quadrado com 3 graus de liberdade. Observamos que a maioria dos resíduos está abaixo dessa linha, sugerindo que o modelo se ajusta bem à maior parte dos dados. No entanto, identificamos alguns pontos que ultrapassam esse limite, indicando potenciais outliers ou dados com grande influência. No total foram 87 pontos acima dessa linha, isto é, aproximadamente 4.8% do total. Uma alternativa é retirar tais pontos influentes do conjunto de dados, estimar novamente os parâmetros do modelo e avaliar a mudança nas estimativas. Se os pontos influentes causam variações significativas nas estimativas, isso pode ser um sinal de que esses pontos não são consistentes com o restante dos dados.



– Figura 32 - Gráfico dos resíduos com uma linha de corte dado pelo quantil 95% de uma $\chi^2_{(3)}$

.

7 CONCLUSÃO

Neste trabalho, exploramos a implementação e a aplicação de Modelos Aditivos Parcialmente Lineares Multivariados (MAPLM) como uma abordagem eficaz para a análise de dados. Através da utilização de *B-Splines*, conseguimos suavizar e ajustar as curvas estimadas, permitindo uma melhor representação das relações não lineares entre as variáveis explicativas e as variáveis resposta. Utilizamos o método de Máxima Verossimilhança Penalizada afim de encontrarmos os estimadores dos parâmetros de um MAPLM. Além disso, demonstramos a construção da matriz de informação para obtenção dos erros padrão dos estimadores e desenvolvemos um algoritmo de *bootstrap*, caso não seja possível a inversão da matriz. Para estimação do parâmetro de suavização, utilizamos a técnica de Validação Cruzada, evitando o problema de superajuste e mantendo um equilíbrio adequado entre viés e variância.

Os resultados das simulações demonstraram que quanto maior a amostra, melhor é o ajuste do modelo, avaliando a eficiência e comprovando as propriedades do EMVP. Na aplicação, também verificamos um bom comportamento do modelo e a análise de resíduos confirmou a adequação do mesmo, evidenciando que as suposições estatísticas foram atendidas.

Por fim, este estudo não apenas valida a eficácia dos MAPLM na análise de dados educacionais, mas também sugere que essa abordagem pode ser aplicada em diversas áreas, como economia, saúde e ciências sociais, onde as relações entre variáveis são frequentemente complexas e não lineares.

REFERÊNCIAS

DAVISON, A. C.; HINKLEY, D. V. **Bootstrap Methods and their Application**. New York: Cambridge University Press, 1997.

BOOR, C. de. **A Practical Guide to Splines**. 2. ed. Berlin: Springer, 1978.

BOENTE, Graciela; MARTINEZ, Alejandra Mercedes. A robust spline approach in partially linear additive models. *Computational Statistics and Data Analysis*, v. 178, p. 107611, 2023. DOI: 10.1016/j.csda.2023.107611.

CASELLA, George; BERGER, Roger L. **Statistical Inference**. 2. ed. Pacific Grove, CA: Duxbury Press, 2002.

EILERS, P. H. C.; MARX, B. D. Flexible Smoothing with B-splines and Penalties. *Statistical Science*, v. 11, n. 2, p. 89–121, 1996.

FARAWAY, J. J. **Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models**. 2. ed. Boca Raton: CRC Press, 2016.

GREEN, P. J.; SILVERMAN, B. W. **Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach**. London: Chapman and Hall, 1994.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York, NY: Springer, 2009. Disponível em: `https://link.springer.com/book/10.1007/978-0-387-84858-7`.

HASTIE, T.; TIBSHIRANI, R. **Generalized Additive Models**. London: Chapman and Hall, 1990.

IBACACHE-PULGAR, G.; PAULA, G. A.; CYSNEIROS, F. J. A. Semiparametric Additive Models under Symmetric Distributions. *Test*, v. 22, p. 103–121, 2013.

LIU, X.; WANG, L.; LIANG, H. Estimation and Variable Selection for Semiparametric Additive Partial Linear Models. *Statistica Sinica*, v. 21, n. 3, p. 1225–1248, 2011.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2023. Disponível em: `http://www.R-project.org`.

RACINE, J. S.; SU, L.; ULLAH, A. **The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics**. New York: Oxford University Press, 2014.

RUPPERT, D.; WAND, M. P.; CARROLL, R. J. **Semiparametric Regression**. New York: Cambridge University Press, 2003.

WANG, B.; FANG, Y.; LIAN, H.; LIANG, H. Additive Partially Linear Models for Massive Heterogeneous Data. *Electronic Journal of Statistics*, v. 13, p. 391–431, 2019.

WOOD, S. N. **Generalized Additive Models: An Introduction with R**. 2. ed. Boca Raton: Chapman and Hall, 2017.

YU, Y.; RUPPERT, D. Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of the American Statistical Association*, v. 97, p. 1042–1054, 2002.

VANEGAS, L. H.; PAULA, G. A. An extension of Log-symmetric Regression Models: R Codes and Applications. *Journal of Statistical Simulation and Computation*, v. 86, p. 1709–1735, 2016.

MANTEIGA, W. G.; LOMBARDÍA, M. J.; MIREA, M. D. M.; SPERLICH, S. Kernel Smoothers and Bootstrapping for Semiparametric Mixed Effects Models. *Journal of Multivariate Analysis*, v. 114, p. 288–302, 2013.

GU, C.; MA, P. Generalized Nonparametric Mixed-effect Models: Computation and Smoothing Parameter Selection. *Journal of Computational and Graphical Statistics*, v. 14, p. 485–504, 2005.

GU, C.; MA, P. Optimal Smoothing in Nonparametric Mixed-effect Models. *Annals of Statistics*, v. 33, p. 1357–1379, 2005.

XU, W.; ZHU, L. Kernel-based Generalized Cross-validation in Non-parametric Regression. *Journal of Computational and Graphical Statistics*, 2005.

FERREIRA, C. S.; PAULA, G. A. Estimation and diagnostic for skew-normal partially linear models. *Journal of Applied Statistics*, v. 44, n. 16, p. 3033–3053, 2017.

RELVAS, C. M.; PAULA, G. A. Partially linear models with first-order autoregressive symmetric errors. *Statistical Papers*, v. 57, n. 3, p. 795–825, 2017.

Bíblia Sagrada: Nova Versão Internacional. São Paulo: Editora Vida, 2001. Tradução baseada nos textos originais hebraico, aramaico e grego.

Introduction to Linear Regression Analysis. 5th ed. Hoboken, NJ: Wiley, 2012. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining.

REFERÊNCIAS

APÊNDICE A – Matriz de Informação

Dado as seguintes definições$B\left(\boldsymbol{\alpha}\right) = \sum^{\frac{1}{2}}$, sendo $\boldsymbol{\alpha}$ os elementos da matriz $\sum$

$\dot{B}_j = \dfrac{\partial B\left(\boldsymbol{\alpha}\right)}{\partial \alpha_j}$, $j = 1, \ldots, \dfrac{r\left(r+1\right)}{2}$, onde rxr é a dimensão da matriz $\sum$

Temos os seguintes resultados para as primeiras derivadas

$$\frac{\partial \Delta}{\partial \boldsymbol{\beta}} = \frac{\partial \Delta}{\partial \boldsymbol{\gamma}} = \mathbf{0}$$

$$\frac{\partial \Delta}{\partial \alpha_j} = 2tr\left(B_{-1}\dot{B}_j\right)$$

$$\frac{\partial g}{\partial \boldsymbol{\beta}} = \frac{\partial g}{\partial \boldsymbol{\alpha}} = \mathbf{0}$$

$$\frac{\partial g}{\partial \boldsymbol{\gamma}} = 2\mathbf{K}^{*}\boldsymbol{\gamma}$$

E os seguintes resultados para as segundas derivadas

$$\frac{\partial^2 w_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = 2\mathbf{X}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{X}_i$$

$$\frac{\partial^2 w_i}{\partial \boldsymbol{\beta}\partial \alpha_j} = 2\mathbf{X}_i^T\mathbf{B}^{-1}\left(\dot{\mathbf{B}}_j\mathbf{B}^{-1} + \mathbf{B}^{-1}\dot{\mathbf{B}}_j\right)\mathbf{B}^{-1}\mathbf{e}_i$$

$$\frac{\partial^2 w_i}{\partial \boldsymbol{\beta}\partial \boldsymbol{\gamma}^T} = 2\boldsymbol{X}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{N}_i$$

$$\frac{\partial^2 w_i}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}} = 2\boldsymbol{N}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{N}_i$$

$$\frac{\partial^2 w_i}{\partial \alpha_j\partial \boldsymbol{\gamma}} = 2\boldsymbol{N}_i^T\boldsymbol{B}^{-1}\left(\dot{\boldsymbol{B}}_j\boldsymbol{B}^{-1} + \boldsymbol{B}^{-1}\boldsymbol{B}_j\right)\boldsymbol{B}^{-1}\mathbf{e}_i$$

$$\frac{\partial^2 w_i}{\partial \alpha_j\partial \alpha_k} = \mathbf{e}_i^T\mathbf{B}^{-1}\Big(\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j\mathbf{B}^{-1} + \dot{\mathbf{B}}_j\mathbf{B}^{-1}\dot{\mathbf{B}}_k\mathbf{B}^{-1} + \dot{\mathbf{B}}_j\mathbf{B}^{-2}\dot{\mathbf{B}}_k + \dot{\mathbf{B}}_k\mathbf{B}^{-2}\dot{\mathbf{B}}_j + \mathbf{B}^{-1}\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j + \mathbf{B}^{-1}\dot{\mathbf{B}}_j\mathbf{B}^{-1}\dot{\mathbf{B}}_k\Big)\mathbf{B}^{-1}\mathbf{e}_i$$

$$\frac{\partial^2 g}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}^T} = 2\mathbf{K}^{*}$$

$$\frac{\partial^2 \Delta}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^T} = \left[\frac{\partial^2 \Delta}{\partial \alpha_j\partial \alpha_k}\right]_{j,k=1,\ldots,p_0} = \left[-tr\left(\mathbf{B^{-1}}\dot{\mathbf{B}}_k\mathbf{B}^{-1}\dot{\mathbf{B}}_j\right)\right]_{j,k=1,\ldots,p_0}$$

$$\frac{\partial \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = \begin{pmatrix} \dfrac{\partial^2 \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} & \dfrac{\partial^2 \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\gamma}^T} & \dfrac{\partial^2 \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}^T} \\ \cdot & \dfrac{\partial^2 \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}^T} & \dfrac{\partial^2 \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\alpha}^T} \\ \cdot & \cdot & \dfrac{\partial^2 \ell_p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^T} \end{pmatrix}$$