

Modelos Aditivos Parcialmente Lineares

Camila G. Toledo¹, Joysce S. Lopes¹, Clécio S. Ferreira^{†2}

¹Graduanda em Estatística, Universidade Federal de Juiz de Fora (UFJF).

²Docente do Departamento de Estatística, Universidade Federal de Juiz de Fora (UFJF).

Resumo: No presente trabalho são apresentados os modelos aditivos parcialmente lineares, isto é, uma extensão dos modelos lineares que incorporam componentes não lineares para modelar relacionamentos mais complexos entre variáveis independentes e a variável resposta. O objetivo do trabalho é estudar tais modelos e implementar um algoritmo que possibilite a estimação dos parâmetros. São utilizados os B-Splines para descrever as componentes não paramétricas. Considerando a função de verossimilhança penalizada, obtemos os estimadores de máxima verossimilhança bem como a matriz de informação de Fisher, sendo esta utilizada para obtenção das estimativas dos desvios-padrão dos estimadores dos parâmetros. Além disso, é utilizado o critério de informação Bayesiano para seleção dos parâmetros de suavização. Foram realizados estudos de simulação para verificar as propriedades assintóticas dos estimadores de máxima verossimilhança. Por fim, para ilustrar a utilidade do modelo proposto, ajustamos o mesmo a um conjunto de dados sobre mortalidade na cidade de Milão (Itália) utilizando como variáveis explicativas não lineares temperatura e umidade e como lineares o número de partículas em suspensão no ar e o tempo (em número de dias).

Palavras-chave: Modelos Aditivos Parcialmente Lineares; B-Splines; Máxima Verossimilhança Penalizada.

Partially Linear Additive Models

Abstract: In the present work, partially linear additive models are presented. This means an extension of linear models that incorporate non-linear components to model more complex relationships between independent variables and the response variable. The goal of the work is to study these models and implement an algorithm that enables parameter estimation. B-Splines are used to describe the non-parametric components. Considering the penalized likelihood function, we obtain maximum likelihood estimators as well as the Fisher information matrix, which is used to obtain standard error estimates for the parameter estimators. Additionally, the Bayesian information criterion is used for parameter smoothing selection. Simulation studies were conducted to verify the asymptotic properties of the maximum likelihood estimators. Finally, to illustrate the utility of the proposed model, it was applied to a dataset on mortality in the city of Milan, Italy. Non-linear explanatory variables include temperature and humidity, while linear variables include the number of suspended particles in the air and time (in days).

Keywords: Partially Linear Additive Models; B-Splines; Penalized Maximum Likelihood.

Introdução

Modelos de regressão linear (MRL) tem recebido especial atenção nos últimos anos, em diversas áreas do conhecimento. Tais modelos assumem que o relacionamento entre a variável resposta e as variáveis explicativas para o i -ésimo indivíduo é representado da seguinte forma

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \epsilon_i \quad (1)$$

onde Y denota a resposta do experimento, X_1, \dots, X_p denotam as variáveis explicativas, $\beta = (\beta_0, \dots, \beta_p)$ é o vetor dos coeficientes de regressão relativo às variáveis explicativas (ou preditores) e ϵ é o erro aleatório. Em algumas situações, algumas variáveis explicativas podem apresentar relações não lineares com a variável resposta. Neste caso é 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

[†]Autor correspondente: clecio.ferreira@ufjf.br.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + f_1(t_{i1}) + \dots + f_s(t_{is}) + \epsilon_i \quad (2)$$

onde $f_k(\cdot)$ é uma função suave univariada que quantifica o efeito da k -ésima variável explanatória t_k ($k=1, \dots, s$) 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 (WOOD, 2017, Cap. 6).

Modelos semiparamétricos têm sido aplicados em diversas situações. Ibacache-Pulgar e Paula (2011) 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 (2016) modelam a concentração de polén de Ambrosia usando como variáveis explicativas temperatura, velocidade do vento, tempo e ocorrência de chuva. Relvas e Paula (2016) modelam as temperaturas da cidade de Ubatuba (SP) correlacionando tempo e a temperatura da cidade de Cananéia (SP).

Modelos aditivos têm sido aplicado em diferentes áreas do conhecimento. Hastie e Tibshirani (1990, Cap. 4) modelam o nível de C-peptídeos através da idade e déficit base. Ruppert et al. (2003) ajustam modelos aditivos em dados de temperatura nos Estados unidos utilizando como variáveis explicativas não-paramétricas a latitude e a longitude. Racine et al. (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 et al. (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 et al. (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.

O objetivo deste trabalho é desenvolver os estimadores de máxima verossimilhança penalizada para os parâmetros do APLM, bem como a matriz de informação de Fisher para calcular as estimativas de desvio-padrão dos estimadores dos parâmetros. As demais seções estão organizadas da seguinte forma: uma seção sobre a teoria de Splines, outra sobre estimação de máxima penalizada dos parâmetros do modelo, uma terceira sobre a matriz de informação de Fisher. Posteriormente, são apresentados um estudo de simulação para verificar as propriedades assintóticas dos estimadores e uma aplicação a dados de mortalidade na cidade de Milão (Itália).

Splines

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.

Um B-spline de grau q consiste de pedaços polinomiais, conectados de uma forma especial pelos chamados nós, com as seguintes propriedades:

- Consiste de $q+1$ pedaços polinomiais, cada um de grau q .
- Os pedaços polinomiais se juntam nos q 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 $q+2$ nós; caso contrário é 0.

- Exceto nos limites, ele sobrepõe com 2q pedaços polinomiais vizinhos (q à direita e q à esquerda).

Sejam t_{min} e t_{max} o domínio da função e r nós interiores, existindo então $r+1$ intervalos. Cada intervalo será coberto de $q+1$ B-splines de grau q . O número total de nós será $q+r+1+2q$ e o número total de B-splines será $r+q$. Seja $B_j(x; q)$ o valor do j -ésimo B-spline de grau q no ponto x . Assim,

$$f_j(t_{ji}) = \sum_{l=1}^{q'} \theta_l B_l(t_{ji})$$

Para maiores informações sobre B-splines, ver de Boor (2001). Podemos escrever o modelo (2) na forma matricial

$$Y = \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^s \mathbf{N}_j \boldsymbol{\gamma}_j + \boldsymbol{\epsilon},$$

onde N_j é uma matriz de incidência $n \times q_j$, com $\boldsymbol{\gamma}_j = (f_j(t_{1j}), \dots, f_j(t_{nj}))$.

Estimação de Máxima Verossimilhança Penalizada

Sob suposição de normalidade dos erros aleatórios, a função de log-verossimilhança associada ao modelo (2) é dada por

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s) \approx -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1}^s \mathbf{N}_j \boldsymbol{\gamma}_j)^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1}^s \mathbf{N}_j \boldsymbol{\gamma}_j). \quad (3)$$

Maximizar a equação (3) sem impor restrições sobre as funções f_1, \dots, f_s pode causar sobreajuste ou não-identificabilidade no parâmetro $\boldsymbol{\beta}$. Um procedimento bem conhecido é o de estimação de máxima verossimilhança penalizada, que consiste em incorporar em $\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s)$ uma função penalty, denotada por $J(f)$. Neste trabalho utilizamos

$$J(f_j) = \int_a^b f_j''(t)^2 dt, \quad j = 1, \dots, s,$$

onde $f_j''(t)$ é a segunda derivada de $f_j(t)$ onde o intervalo $[a_j, b_j]$ contém todos os nós. Dependendo da escolha dos nós, podemos ter expressões analíticas (e matriciais) para $J(f_j)$. Por exemplo, Green and Silverman (1994) usam como nós todos os distintos pontos enquanto Eilers and Marx (1996) adotam a metodologia de escolher k pontos igualmente espaçados, com $k < n$. As duas formas levam a penalty à seguinte forma: $J(f_j) = \boldsymbol{\gamma}_j^T \mathbf{K}_j \boldsymbol{\gamma}_j$.

Assim, para um modelo com erros normais, a função de log-verossimilhança penalizada é dada por

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s, \boldsymbol{\alpha}) \approx \ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s) - 1/2 \sum_{j=1}^s \alpha_j \boldsymbol{\gamma}_j^T \mathbf{K}_j \boldsymbol{\gamma}_j$$

em que \mathbf{K}_j é a matriz penalty obtida como função apenas dos nós.

Através de um método iterativo chega-se aos estimadores de máxima verossimilhança penalizada de $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_s)$, os quais são dados por

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(k)} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{Y} - \sum_{j=1}^s \mathbf{N}_j \hat{\boldsymbol{\gamma}}_j^{(k)}) \\ \hat{\boldsymbol{\gamma}}_i^{(k)} &= (\mathbf{N}_i^T \mathbf{N}_i + \hat{\sigma}^2 \alpha_i \mathbf{K}_i)^{-1} \mathbf{N}_i^T (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k)} - \sum_{j=1, j \neq i}^s \mathbf{N}_j \hat{\boldsymbol{\gamma}}_j^{(k)}), \quad i = 1, \dots, s \\ \hat{\sigma}^2^{(k)} &= \frac{1}{n} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k)} - \sum_{j=1}^q \mathbf{N}_j \hat{\boldsymbol{\gamma}}_j^{(k)})^T (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k)} - \sum_{j=1}^q \mathbf{N}_j \hat{\boldsymbol{\gamma}}_j^{(k)}). \end{aligned}$$

As iterações são repetidas até que uma regra de convergência adequada seja satisfeita, e.g., $\sqrt{\sum_i (\theta_i^{(k+1)} - \theta_i^{(k)})^2}$ seja suficientemente menor que 10^{-5} ou se 5000 iterações tenham sido realizadas. Como valores iniciais utilizamos

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(0)} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \\ \widehat{\sigma}^2^{(0)} &= \frac{1}{n} (\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}^{(0)})^T (\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}^{(0)}), \\ \widehat{\boldsymbol{\gamma}}_i^{(0)} &= (\mathbf{N}_i^T \mathbf{N}_i + \widehat{\sigma}^2^{(0)} \alpha_i \mathbf{K}_i)^{-1} \mathbf{N}_i^T (\mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}^{(0)}), \quad i = 1, \dots, s.\end{aligned}$$

Além disso, o SIC é utilizado a fim de selecionar o valor de α_i , que é responsável por controlar o grau de penalização a ser aplicado na i -ésima curva não-paramétrica do modelo, durante o processo de estimação. Valores mais altos de α_i resultam em uma maior penalidade aplicada sobre as curvas, tornando-as mais suaves. À medida que os valores de α_i aumentam, tendendo ao ∞ , as curvas se tornam cada vez menos oscilantes assumindo as formas de retas. Por outro lado, valores menores de α_i , tornam as curvas menos suaves, pois a penalização aplicada sobre elas é menor. Com isso, de acordo que α_i diminuem, as funções não-paramétricas do modelo tendem a curvas que se ajustam de forma excessiva aos dados, aumentando também sua instabilidade. Sendo assim, a seleção de valores ideais para α_i é algo imprescindível para obter um modelo bem ajustado.

Matriz de Informação de Fisher

Os erros padrão das estimativas de β , $\boldsymbol{\gamma}_i$ e σ^2 podem ser obtidos por meio da relação: $\text{Ep}(\theta) = \text{diag}(\text{MI}^{-1})^{1/2}$, onde MI é a matriz de informação de Fisher.

$$\text{MI} = \begin{bmatrix} -\frac{\partial^2 l_p(\theta)}{\partial \beta \partial \beta^T} & -\frac{\partial^2 l_p(\theta)}{\partial \beta \partial \boldsymbol{\gamma}_j^T} & -\frac{\partial^2 l_p(\theta)}{\partial \beta \partial \sigma^2} \\ -\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \beta^T} & -\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_i^T} & -\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \sigma^2} \\ -\frac{\partial^2 l_p(\theta)}{\partial \sigma^2 \partial \beta} & -\frac{\partial^2 l_p(\theta)}{\partial \sigma^2 \partial \boldsymbol{\gamma}_j} & -\frac{\partial^2 l_p(\theta)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix}$$

Com o objetivo de facilitar a implementação e diminuir o custo computacional, a matriz de Fisher utilizada foi a matriz esperada. Sendo assim, temos os seguintes resultados:

$$\begin{aligned}E\left(-\frac{\partial^2 l_p(\theta)}{\partial \beta \partial \beta^T}\right) &= \frac{1}{\sigma^2} \mathbf{X}_i^T \mathbf{X}_i & E\left(-\frac{\partial^2 l_p(\theta)}{\partial \beta \partial \boldsymbol{\gamma}_j^T}\right) &= \frac{1}{\sigma^2} \mathbf{X}_i^T \mathbf{N}_i \\ E\left(-\frac{\partial^2 l_p(\theta)}{\partial \beta \partial \sigma^2}\right) &= 0 & E\left(-\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \beta^T}\right) &= \frac{1}{\sigma^2} (\mathbf{X}_i^T \mathbf{N}_i)^T \\ E\left(-\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_i^T}\right) &= \frac{1}{\sigma^2} \mathbf{N}_j^T \mathbf{N}_i + \alpha_i \mathbf{K}_i \quad (i = j) & E\left(-\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \boldsymbol{\gamma}_i^T}\right) &= \frac{1}{\sigma^2} \mathbf{N}_j^T \mathbf{N}_i \quad (i \neq j) \\ E\left(-\frac{\partial^2 l_p(\theta)}{\partial \boldsymbol{\gamma}_j \partial \sigma^2}\right) &= 0 & E\left(-\frac{\partial^2 l_p(\theta)}{\partial \sigma^2 \partial \beta}\right) &= 0 \\ E\left(-\frac{\partial^2 l_p(\theta)}{\partial \sigma^2 \partial \boldsymbol{\gamma}_j}\right) &= 0 & E\left(-\frac{\partial^2 l_p(\theta)}{\partial \sigma^2 \partial \sigma^2}\right) &= -\frac{n}{2\sigma^4}\end{aligned}$$

Em casos onde 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.

Estudo de Simulação

Este estudo tem como objetivo a estimação por Máxima Verossimilhança Penalizada (EMVP) de um Modelo Aditivo Parcialmente Linear considerando o modelo (2) 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 200, 500 e 1000. O modelo simulado é composto de duas variáveis explicativas não lineares e uma variável explicativa linear. Isto é

$$Y_i = \beta X_i + f_1(t_{1i}) + f_2(t_{2i}) + \epsilon_i, \quad i = 1, \dots, n,$$

onde $X \sim U(0,1)$, $\beta = 5$, $f_1(t_1) = \cos(t_1)$, $f_2(t_2) = \cos(4\pi t_2)e^{-t_2^2/2}$, $\epsilon_i \sim N(0,0.05)$, $t_1 \sim U(2\pi, 6\pi)$, $t_2 \sim U(0.6, 1.6)$

A Figura 1 apresenta as 1000 curvas estimadas da curva 1, enquanto a Figura 2 apresenta as 1000 curvas estimadas da curva 2.

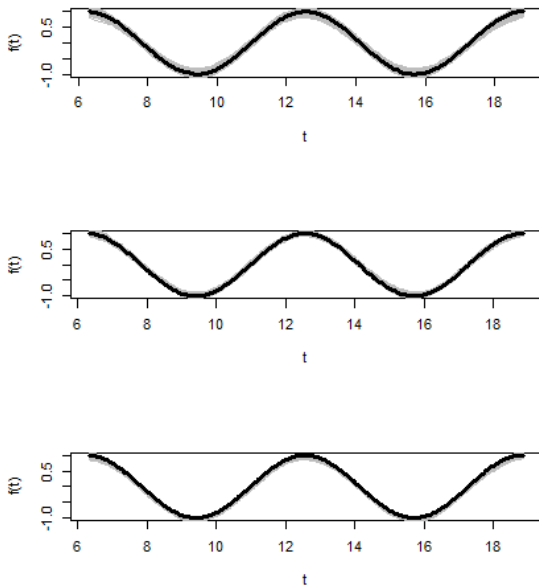


Figure 1: $f(t) = \cos(t)$

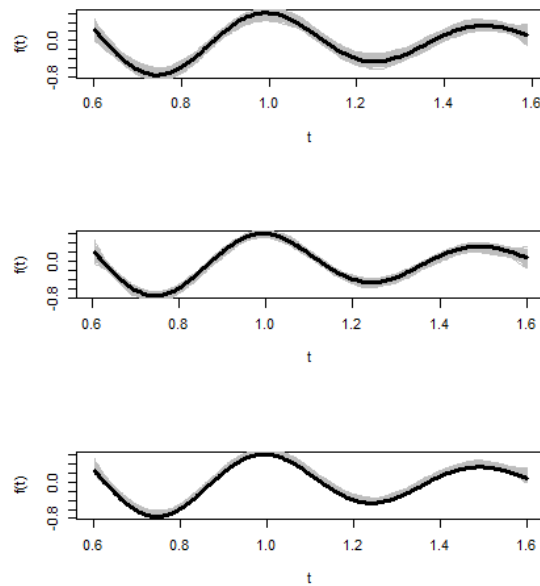


Figure 2: $f(t) = \cos(4\pi t)e^{-t^2/2}$

Table 1: Médias e desvios padrão das estimativas e média dos desvios padrão obtidos pela matriz de informação dos parâmetros.

Valor Real	n=200				n=500				n=1000			
	Média	DP	DP Emp		Média	DP	DP Emp		Média	DP	DP Emp	
σ^2	0.05	0.046	0.005	0.005	0.048	0.003	0.003		0.049	0.002	0.002	
β	5.00	4.995	0.056	0.055	5.001	0.036	0.034		5.000	0.024	0.025	

Pela Tabela 1, nota-se um comportamento satisfatório do modelo, uma vez que as curvas estimadas estão muito próximas da curva real. Além disso, observa-se que ao aumentar o tamanho amostral, os desvios padrão e os desvios padrão empíricos dos estimadores de β e σ diminuem, e a média de suas estimativas se aproximam de seus valores reais, alcançando as propriedades assintóticas do Estimador de Máxima Verossimilhança (EMV).

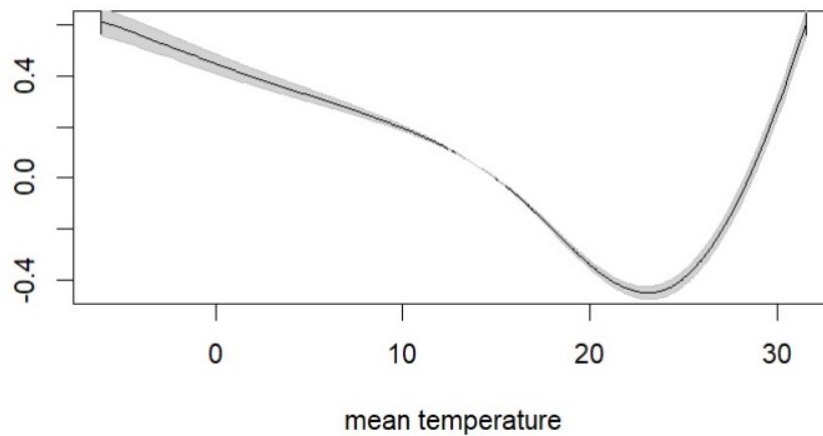
Aplicação

Para a aplicação foi utilizado o conjunto de dados “milan.mort” do pacote SemiPar disponível no software R, o qual contém dados de 3652 dias consecutivos (de 1^o de janeiro de 1980 a 30 de dezembro de 1989) para a cidade de Milão, Itália. O objetivo é modelar a variabilidade do total de mortes utilizando como variáveis explicativas não lineares a média de temperatura ($temperature_t$) e a humidade relativa ($humidity_t$) e como variáveis lineares partículas totais em suspensão no ar ambiente (TSP_t) e número de dias (t) desde 31 de dezembro de 1979.

$$\sqrt{mortality_t} = \beta_0 + \beta_1 TSP_t + \beta_2 t + f_1(temperature_t) + f_2(humidity_t) + \epsilon_t, \quad t = 1, \dots, 3652,$$

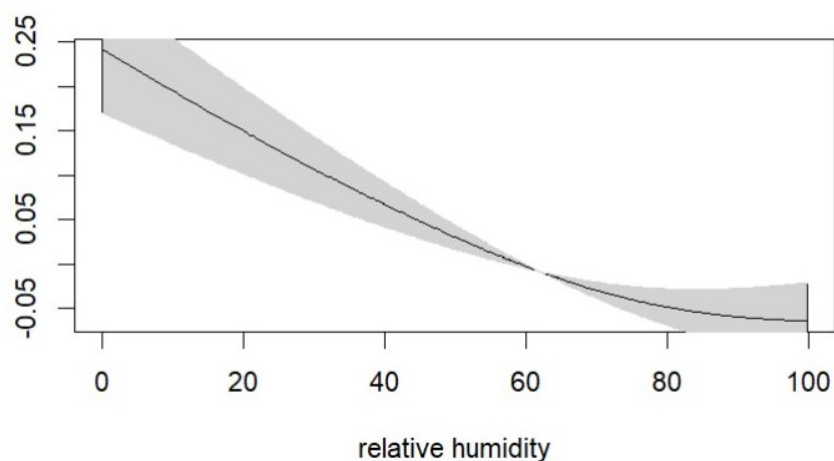
onde $mortality_t$ é o número total de mortes no tempo t . A Figura 3 apresenta a estimativa não paramétrica para a temperatura enquanto a Figura 4 apresenta a estimativa não paramétrica para a humidade relativa.

Figure 3: Estimativa não paramétrica para temperatura.



Fonte: Autores.

Figure 4: Estimativa não paramétrica para humidade.



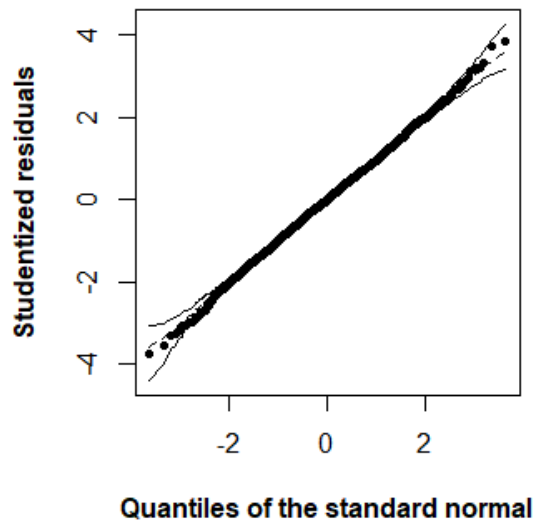
Fonte: Autores

Table 2: Mortalidade: Estimativas de máxima verossimilhança penalizada dos parâmetros e seus erros-padrão.

Parâmetro	Estimativa	Erro Padrão
σ^2	0.3256739	0.0265091
β_0	5.803937	0.0001389
β_1	0.000837	0.0000093
β_2	-0.000173	0.0490978
α_1	100	-
α_2	2.758055	-
$\ell(\boldsymbol{\theta})$	-3482.079	-
SIC	7070.462	-

Uma das maneiras de avaliarmos o ajuste do modelo é analisando o gráfico envelope, apresentado na Figura 5. Analisando o gráfico, temos evidências de que o modelo foi bem ajustado, pois os resíduos encontram-se entre as faixas do gráfico envelope.

Figure 5: Mortalidade: Gráfico de envelope simulado



Fonte: Autores.

Conclusões

Neste trabalho desenvolvemos os estimadores de máxima verossimilhança e a matriz de informação de Fisher para os modelos aditivos parcialmente lineares. Um estudo de simulação foi realizado para verificar as propriedades assintóticas do EMVP. O modelo foi aplicado a um conjunto de dados reais sobre mortalidade na cidade de Milão (Itália), mostrando a eficiência do modelo. Extensões deste trabalho podem ser desenvolvidas considerando erros aleatórios que levem em conta assimetria e/ou cauda pesada.

Agradecimentos

Clécio da Silva Ferreira e Camila Galhardo Toledo foram parcialmente financiados por FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), Grant CEX-APQ-00409-22. Clécio da Silva Ferreira foi parcialmente financiado por CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Grant 302666/2022-2.

References

- DAVISON, A. C.; HINKLEY, D. V. *Bootstrap methods and their application*. New York: Cambridge University Press, 1997.
- DE BOOR, C. *A practical guide to splines*, revised. Berlin: Springer, 2001.
- EILERS, P. H. C.; MARX, B. D. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89-121, 1996.
- FARAWAY, J. J. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, sec. 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. *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*, 22, 103-121, 2013.
- LIU, X.; WANG, L.; LIANG, H. Estimation and Variable Selection for Semiparametric Additive Partial Linear Models. *Stat Sin.*, 21 (3), 1225-1248, 2011.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 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.; CARROL, R. J. *Semiparametric Regression*. New York: Cambridge Univ. Press, 2003.
- WANG, B.; FANG, Y.; LIAN, H.; LIANG, H. Additive partially linear models for massive heterogeneous data. *Electronic Journal of Statistics*, 13, 391-431, 2019.
- WOOD, S. N. *Generalized Additive Models: an introduction with R*. Second Edition. Boca Raton: Chapman and Hall, 2017.