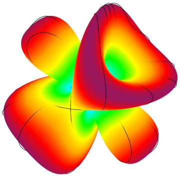

Assessing parametric regression models with directional predictors

Eduardo García–Portugués (eduardo.garcia@usc.es)
Ingrid Van Keilegom Rosa M. Crujeiras
Wenceslao González–Manteiga

Department of Statistics and Operations Research
University of Santiago de Compostela

Advances in Directional Statistics, May 20, 2014

Introduction

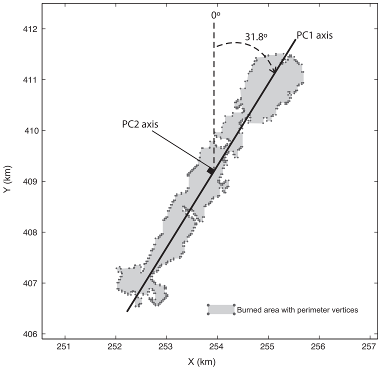

Figure : Fire orientation.

- Data: orientations $(\mathbf{X})$ and log–burnt areas (Y) of 26870 wildfires in Portugal during 1985–2005.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Introduction

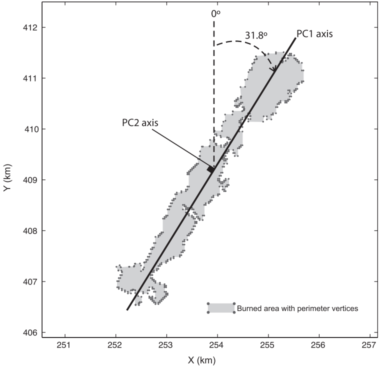

Figure : Fire orientation.

- Data: orientations $(\mathbf{X})$ and log–burnt areas (Y) of 26870 wildfires in Portugal during 1985–2005.
- What is the relationship m between $\mathbf{X}$ and Y?

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$$

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Introduction

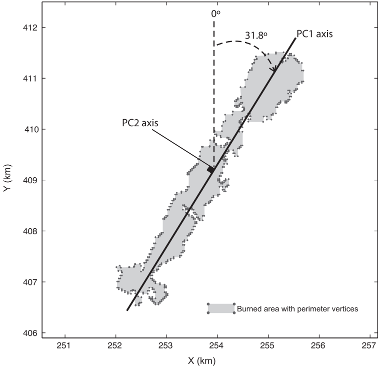

Figure : Fire orientation.

- Data: orientations $(\mathbf{X})$ and log–burnt areas (Y) of 26870 wildfires in Portugal during 1985–2005.
- What is the relationship m between $\mathbf{X}$ and Y?

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$$

- Two approaches for studying m:
  - Estimate m nonparametrically.
  - Check if m can be specified as a certain parametric model.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Aim of this work

Develop new nonparametric methods for the **regression** with directional predictor $\mathbf{X}$ and linear response Y:

1. Nonparametric **estimation** of the regression function m,

$$\{(\mathbf{X}_i, Y_i)\}_{i=1}^n \implies \widehat{m}_h.$$

2. **Goodness–of–fit** test for parametric regression models:

$$H_0 : m \in \mathcal{M}_{\boldsymbol{\Theta}} = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}.$$

Table of contents

Table of contents

Estimator

- Denote the q–sphere by $\Omega_q = \left\{\mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1\right\}$.
- Let $(\mathbf{X}, Y)$ be a random variable with support in $\Omega_q \times \mathbb{R}$.
- Consider the regression model

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon \quad \text{with} \quad \left\{ \begin{array}{ll} m(\mathbf{x}) & = \mathbb{E}\left[Y|\mathbf{X} = \mathbf{x}\right], \\ \sigma^2(\mathbf{x}) & = \mathbb{V}\mathrm{ar}\left[Y|\mathbf{X} = \mathbf{x}\right], \end{array} \right.$$

and with $\varepsilon \perp \mathbf{X}$, $\mathbb{E}\left[\varepsilon\right] = 0$ and $\mathbb{V}\mathrm{ar}\left[\varepsilon\right] = 1$.

Estimator

- Denote the q–sphere by $\Omega_q = \left\{\mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}|| = 1\right\}$.
- Let $(\mathbf{X}, Y)$ be a random variable with support in $\Omega_q \times \mathbb{R}$.
- Consider the regression model

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon \quad \text{with} \quad \begin{cases} m(\mathbf{x}) &= \mathbb{E}\left[Y|\mathbf{X} = \mathbf{x}\right], \\ \sigma^2(\mathbf{x}) &= \mathbb{V}\mathrm{ar}\left[Y|\mathbf{X} = \mathbf{x}\right], \end{cases}$$

  and with $\varepsilon \perp \mathbf{X}$, $\mathbb{E}\left[\varepsilon\right] = 0$ and $\mathbb{V}\mathrm{ar}\left[\varepsilon\right] = 1$.
- We want to estimate m nonparametrically from $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$.
- We will require Taylor expansions, so the first **condition** is

  LL1 Extend m and f from Ω_q to $\mathbb{R}^{q+1} \setminus \{\mathbf{0}\}$ by $m(\mathbf{x}) \equiv m(\mathbf{x}/||\mathbf{x}||)$ and $f(\mathbf{x}) \equiv f(\mathbf{x}/||\mathbf{x}||)$. m is third and f is twice continuously differentiable and f is bounded away from zero.

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right)$$

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \left(\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T\right)(\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right)$$

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$\begin{aligned} m(\mathbf{X}_i) &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_q \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2 \right) \\ &\approx \beta_0 + \boldsymbol{\beta}_{1:(q+1)}^T \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}), \end{aligned}$$

with $\mathbf{B}_q = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B}_q \mathbf{B}_q^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$, $\beta_0 = m(\mathbf{x})$ and $\boldsymbol{\beta}_{1:(q+1)} = \mathbf{B}_q^T \boldsymbol{\nabla} m(\mathbf{x})$.

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_q \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2\right)$$
$$\approx \beta_0 + \boldsymbol{\beta}_{1:(q+1)}^T \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}),$$

with $\mathbf{B}_q = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B}_q \mathbf{B}_q^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$, $\beta_0 = m(\mathbf{x})$ and $\boldsymbol{\beta}_{1:(q+1)} = \mathbf{B}_q^T \boldsymbol{\nabla} m(\mathbf{x})$.

- This motivates the weighted minimum least squares problem ($p = 0$ for local constant and $p = 1$ for local linear)

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{q+1}} \sum_{i=1}^n \left(Y_i - \beta_0 - \delta_{p,1} \left(\beta_1, \ldots, \beta_q\right)^T \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x})\right)^2 L_h(\mathbf{x}, \mathbf{X}_i).$$

- Let $\mathbf{x}, \mathbf{X}_i \in \Omega_q$. The one term Taylor expansion of m is:

$$\begin{aligned} m(\mathbf{X}_i) &= m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_q \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}) + \mathcal{O}\left(||\mathbf{X}_i - \mathbf{x}||^2 \right) \\ &\approx \beta_0 + \boldsymbol{\beta}_{1:(q+1)}^T \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}), \end{aligned}$$

with $\mathbf{B}_q = (\mathbf{b}_1, \ldots, \mathbf{b}_q)_{(q+1)\times q}$ such that $\mathbf{B}_q \mathbf{B}_q^T = \mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T$, $\beta_0 = m(\mathbf{x})$ and $\boldsymbol{\beta}_{1:(q+1)} = \mathbf{B}_q^T \boldsymbol{\nabla} m(\mathbf{x})$.

- This motivates the weighted minimum least squares problem ($p = 0$ for local constant and $p = 1$ for local linear)

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{q+1}} \sum_{i=1}^n \left(Y_i - \beta_0 - \delta_{p,1} \left(\beta_1, \ldots, \beta_q \right)^T \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x}) \right)^2 L_h(\mathbf{x}, \mathbf{X}_i).$$

- The solution is given by

$$\widehat{m}_{h,p}(\mathbf{x}) = \hat{\beta}_0 = \mathbf{e}_{1,p}^T \left(\boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},p} \right)^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x},p}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \mathbf{Y} = \sum_{i=1}^n W_p^n \left(\mathbf{x}, \mathbf{X}_i \right) Y_i,$$

where $\boldsymbol{\mathcal{W}}_{\mathbf{x}} = \mathrm{diag}\left(L_h(\mathbf{x}, \mathbf{X}_1), \ldots, L_h(\mathbf{x}, \mathbf{X}_n) \right)$ and

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{\mathcal{X}}_{\mathbf{x},0} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \mathbf{B}_q \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \mathbf{B}_q \end{pmatrix}_{n \times q}.$$

Alternative proposal

▸ Technical details

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data *J. Amer. Statist. Assoc.* (to appear).

- Taylor expansion based on the tangent–normal decomposition $\mathbf{X}_i = \mathbf{x}\cos(\theta_{\mathbf{x},i}) + \boldsymbol{\xi}_{\mathbf{x},i}\sin(\theta_{\mathbf{x},i})$, where $\theta_{\mathbf{x},i} \in [0, 2\pi)$, $\mathbf{x}, \boldsymbol{\xi}_{\mathbf{x},i} \in \Omega_q$ and $\boldsymbol{\xi}_{\mathbf{x},i} \perp \mathbf{x}$:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}^T \boldsymbol{\nabla} m(\mathbf{x}) + \mathcal{O}\left(\theta_{\mathbf{x},i}^2\right).$$

Alternative proposal ▸ Technical details

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data *J. Amer. Statist. Assoc.* (to appear).

- Taylor expansion based on the tangent–normal decomposition $\mathbf{X}_i = \mathbf{x}\cos(\theta_{\mathbf{x},i}) + \boldsymbol{\xi}_{\mathbf{x},i}\sin(\theta_{\mathbf{x},i})$, where $\theta_{\mathbf{x},i} \in [0, 2\pi)$, $\mathbf{x}, \boldsymbol{\xi}_{\mathbf{x},i} \in \Omega_q$ and $\boldsymbol{\xi}_{\mathbf{x},i} \perp \mathbf{x}$:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}^T \boldsymbol{\nabla} m(\mathbf{x}) + \mathcal{O}\left(\theta_{\mathbf{x},i}^2\right).$$

- Weighted minimum least squares problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{q+2}} \sum_{i=1}^n \left(Y_i - \beta_0 - (\beta_1, \ldots, \beta_{q+1})^T \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}\right)^2 K_\kappa(\cos(\theta_{\mathbf{x},i})),$$

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & \theta_{\mathbf{x},1}\boldsymbol{\xi}_{\mathbf{x},1}^T \\ \vdots & \vdots \\ 1 & \theta_{\mathbf{x},n}\boldsymbol{\xi}_{\mathbf{x},n}^T \end{pmatrix}_{n\times(q+1)}, \boldsymbol{\mathcal{W}}_{\mathbf{x}} = \operatorname{diag}\left(K_\kappa(\cos(\theta_{\mathbf{x},1})), \ldots, K_\kappa(\cos(\theta_{\mathbf{x},n}))\right).$$

Alternative proposal

Technical details

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data *J. Amer. Statist. Assoc.* (to appear).

- Taylor expansion based on the tangent–normal decomposition $\mathbf{X}_i = \mathbf{x}\cos(\theta_{\mathbf{x},i}) + \boldsymbol{\xi}_{\mathbf{x},i}\sin(\theta_{\mathbf{x},i})$, where $\theta_{\mathbf{x},i} \in [0, 2\pi)$, $\mathbf{x}, \boldsymbol{\xi}_{\mathbf{x},i} \in \Omega_q$ and $\boldsymbol{\xi}_{\mathbf{x},i} \perp \mathbf{x}$:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}^T \boldsymbol{\nabla} m(\mathbf{x}) + \mathcal{O}\left(\theta_{\mathbf{x},i}^2\right).$$

- Weighted minimum least squares problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{q+2}} \sum_{i=1}^n \left(Y_i - \beta_0 - (\beta_1, \ldots, \beta_{q+1})^T \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}\right)^2 K_\kappa(\cos(\theta_{\mathbf{x},i})),$$

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & \theta_{\mathbf{x},1}\boldsymbol{\xi}_{\mathbf{x},1}^T \\ \vdots & \vdots \\ 1 & \theta_{\mathbf{x},n}\boldsymbol{\xi}_{\mathbf{x},n}^T \end{pmatrix}_{n\times(q+1)}, \boldsymbol{\mathcal{W}}_{\mathbf{x}} = \mathrm{diag}\left(K_\kappa(\cos(\theta_{\mathbf{x},1})), \ldots, K_\kappa(\cos(\theta_{\mathbf{x},n}))\right).$$

- $\boldsymbol{\mathcal{X}}_{\mathbf{x},1}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},1}$ is singular, so a pseudo–inverse is required.

Alternative proposal ▸ Technical details

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data *J. Amer. Statist. Assoc.* (to appear).

- Taylor expansion based on the tangent–normal decomposition $\mathbf{X}_i = \mathbf{x}\cos(\theta_{\mathbf{x},i}) + \boldsymbol{\xi}_{\mathbf{x},i}\sin(\theta_{\mathbf{x},i})$, where $\theta_{\mathbf{x},i} \in [0, 2\pi)$, $\mathbf{x}, \boldsymbol{\xi}_{\mathbf{x},i} \in \Omega_q$ and $\boldsymbol{\xi}_{\mathbf{x},i} \perp \mathbf{x}$:

$$m(\mathbf{X}_i) = m(\mathbf{x}) + \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}^T \boldsymbol{\nabla} m(\mathbf{x}) + \mathcal{O}\left(\theta_{\mathbf{x},i}^2\right).$$

- Weighted minimum least squares problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{q+2}} \sum_{i=1}^n \left(Y_i - \beta_0 - (\beta_1, \ldots, \beta_{q+1})^T \theta_{\mathbf{x},i}\boldsymbol{\xi}_{\mathbf{x},i}\right)^2 K_\kappa(\cos(\theta_{\mathbf{x},i})),$$

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & \theta_{\mathbf{x},1}\boldsymbol{\xi}_{\mathbf{x},1}^T \\ \vdots & \vdots \\ 1 & \theta_{\mathbf{x},n}\boldsymbol{\xi}_{\mathbf{x},n}^T \end{pmatrix}_{n\times(q+1)}, \boldsymbol{\mathcal{W}}_{\mathbf{x}} = \mathrm{diag}\left(K_\kappa(\cos(\theta_{\mathbf{x},1})), \ldots, K_\kappa(\cos(\theta_{\mathbf{x},n}))\right).$$

- $\boldsymbol{\mathcal{X}}_{\mathbf{x},1}^T \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x},1}$ is singular, so a pseudo–inverse is required.
- Results for bias, variance and normality are equivalent to ours.

Remarkable cases

1. **Local constant** $(p = 0)$, the Nadaraya–Watson estimator:

$$\hat{m}_{h,0}(\mathbf{x}) = \sum_{i=1}^{n} W_p^n(\mathbf{x}, \mathbf{X}_i)\, Y_i = \frac{\sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i) Y_i}{\sum_{j=1}^{n} L_h(\mathbf{x}, \mathbf{X}_j)}.$$

Wang, X., Zhao, L., and Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

Remarkable cases

1. **Local constant** $(p = 0)$, the Nadaraya–Watson estimator:

$$\hat{m}_{h,0}(\mathbf{x}) = \sum_{i=1}^{n} W_p^n(\mathbf{x}, \mathbf{X}_i)\, Y_i = \frac{\sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i) Y_i}{\sum_{j=1}^{n} L_h(\mathbf{x}, \mathbf{X}_j)}.$$

Wang, X., Zhao, L., and Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

2. **Circular case** $(q = 1)$: set $\mathbf{x} = (\cos\theta, \sin\theta)$, $\mathbf{X}_i = (\cos\Theta_i, \sin\Theta_i)$ and $\kappa = h^{-2}$, for $\theta, \Theta_i \in [0, 2\pi)$. For $p = 1$, $\mathbf{B}_1 = (-\sin\theta, \cos\theta)^T$ and $\mathbf{B}_1 \perp \mathbf{x}$:

$$\boldsymbol{\mathcal{X}}_{\mathbf{x},1} = \begin{pmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \mathbf{B}_1 \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \mathbf{B}_1 \end{pmatrix} = \begin{pmatrix} 1 & \sin(\theta - \Theta_1) \\ \vdots & \vdots \\ 1 & \sin(\theta - \Theta_n) \end{pmatrix}.$$

Remarkable cases

1. **Local constant** $(p = 0)$, the Nadaraya–Watson estimator:

$$\widehat{m}_{h,0}(\mathbf{x}) = \sum_{i=1}^{n} W_p^n(\mathbf{x}, \mathbf{X}_i)\, Y_i = \frac{\sum_{i=1}^{n} L_h(\mathbf{x}, \mathbf{X}_i) Y_i}{\sum_{j=1}^{n} L_h(\mathbf{x}, \mathbf{X}_j)}.$$

Wang, X., Zhao, L., and Wu, Y. (2000). Distribution free laws of the iterated logarithm for kernel estimator of regression function based on directional data. *Chinese Ann. Math. Ser. B*, 21(4):489–498.

2. **Circular case** $(q = 1)$: set $\mathbf{x} = (\cos\theta, \sin\theta)$, $\mathbf{X}_i = (\cos\Theta_i, \sin\Theta_i)$ and $\kappa = h^{-2}$, for $\theta, \Theta_i \in [0, 2\pi)$. For $p = 1$, $\mathbf{B}_1 = (-\sin\theta, \cos\theta)^T$ and $\mathbf{B}_1 \perp \mathbf{x}$:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 \sin(\theta - \Theta_i))^2 L_\kappa(\kappa(1 - \cos(\theta - \Theta_i))).$$

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Properties

Theorem (Conditional bias and variance)

Under assumptions LL1–LL4, for $\mathbf{x} \in \Omega_q$ *(uniformly),*

$$\mathbb{E}\left[\hat{m}_{h,p}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] = m(\mathbf{x}) + b_q(L)B_p(\mathbf{x})h^2 + o_{\mathbb{P}}\left(h^2\right),$$

$$\mathbb{V}\mathrm{ar}\left[\hat{m}_{h,p}(\mathbf{x})|\mathbf{X}_1,\ldots,\mathbf{X}_n\right] = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q f(\mathbf{x})}\sigma^2(\mathbf{x}) + o_{\mathbb{P}}\left((nh^q)^{-1}\right),$$

where $B_p(\mathbf{x}) = \begin{cases} \frac{2}{q}\frac{\boldsymbol{\nabla} f(\mathbf{x})^T}{f(\mathbf{x})}\boldsymbol{\nabla} m(\mathbf{x}) + \frac{1}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right], & p=0, \\ \frac{1}{q}\mathrm{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right], & p=1. \end{cases}$

- **Conditions**

LL2 σ^2 is uniformly continuous and bounded away from zero.
LL3 The kernel $L : [0,\infty) \to [0,\infty)$ is a bounded function with exponential decay ($L(r) \leq Me^{-\alpha r}$).
LL4 The positive sequence $h = h_n$ satisfies $h \to 0$ and $nh^q \to \infty$.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22(3):1346–1370.

- From the conditional bias and variance, expansions for conditional MSE, MISE and h_{AMISE} bandwidth follow. ▸ Simulations bandwidth selection

Corollary (Equivalent kernel, as in Fan and Gijbels (1996))

Under assumptions LL1–LL4, the weights in the estimator $\hat{m}_{h,p}(\mathbf{x}) = \sum_{i=1}^n W_p^n(\mathbf{x}, \mathbf{X}_i) Y_i$ *at* $\mathbf{x} \in \Omega_q$ *satisfy (uniformly)*

$$W_p^n(\mathbf{x}, \mathbf{X}_i) = \frac{1}{nh^q \lambda_q(L) f(\mathbf{x})} L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right)(1 + o_{\mathbb{P}}(1)).$$

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall.

Theorem (Asymptotic normality)

Under assumptions LL1–LL4, for $\mathbf{x} \in \Omega_q$ *such that for a* $\delta > 0$ $\mathbb{E}\left[(Y - m(\mathbf{X}))^{2+\delta} | \mathbf{X} = \mathbf{x}\right] < \infty$,

$$\sqrt{nh^q}\left(\hat{m}_{h,p}(\mathbf{x}) - m(\mathbf{x}) - b_q(L) B_p(\mathbf{x}) h^2\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{f(\mathbf{x})}\sigma^2(\mathbf{x})\right).$$

Table of contents

Testing a parametric model

- We want to check $H_0 : m \in \mathcal{M}_\Theta = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$.
- No test available in the literature for checking this hypothesis.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST*, 22(3):361–411.

Testing a parametric model

- We want to check $H_0 : m \in \mathcal{M}_\Theta = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$.
- No test available in the literature for checking this hypothesis.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST*, 22(3):361–411.

- We consider as statistic the smoothed weighted $\mathcal{L}^2$–distance between $\widehat{m}_{h,p}$ and $m_{\widehat{\boldsymbol{\theta}}}$:

$$T_n = \int_{\Omega_q} \left(\widehat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\widehat{\boldsymbol{\theta}}}(\mathbf{x})\right)^2 \widehat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}).$$

where $\mathcal{L}_{h,p} m(\mathbf{x}) = \sum_{i=1}^n W_n^p(\mathbf{x}, \mathbf{X}_i)\, m(\mathbf{X}_i)$.

Alcalá, J. T., Cristóbal, J. A., and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Probab. Lett.*, 42(1):39–46.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947.

Limit distribution

Theorem (Limit distribution of T_n) ‣ Sketch of the proof

Under conditions LL1–LL4, GF1–GF2 and under $H_0 : m \in \mathcal{M}_\Theta$,

$$nh^{\frac{q}{2}} \left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q} \int_{\Omega_q} \sigma^2(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}) \right) \stackrel{d}{\longrightarrow} \mathcal{N}\left(0, 2\nu^2\right).$$

- **Conditions**

GF1 Under H_0, there exists a $\widehat{\boldsymbol{\theta}}$ such that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.

GF2 $m_{\boldsymbol{\theta}}$ is continously differentiable as a function of $\boldsymbol{\theta}$, being this derivative also continuous for $\mathbf{x} \in \Omega_q$.

Limit distribution

Theorem (Limit distribution of T_n) ‣ Sketch of the proof

Under conditions LL1–LL4, GF1–GF2 and under $H_0 : m \in \mathcal{M}_\Theta$,

$$nh^{\frac{q}{2}}\left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right) \xrightarrow{d} \mathcal{N}\left(0, 2\nu^2\right).$$

- **Conditions**
- **Constants**

$$\nu^2 = \int_{\Omega_q}\sigma^4(\mathbf{x})w(\mathbf{x})^2\,\omega_q(d\mathbf{x})$$

$$\times\, \gamma_q\lambda_q(L)^{-4}\int_0^\infty r^{\frac{q}{2}-1}\left\{\int_0^\infty \rho^{\frac{q}{2}-1}L(\rho)\varphi_q(r,\rho)\,d\rho\right\}^2 dr$$

$$\varphi_q(r,\rho) = \begin{cases} L\left(r+\rho-2(r\rho)^{\frac{1}{2}}\right) + L\left(r+\rho+2(r\rho)^{\frac{1}{2}}\right), & q=1,\\ \int_{-1}^{1}\left(1-u^2\right)^{\frac{q-3}{2}} L\left(r+\rho-2u(r\rho)^{\frac{1}{2}}\right)du, & q\geq 2.\end{cases}$$

Limit distribution

Theorem (Limit distribution of T_n) ▸ Sketch of the proof

Under conditions LL1–LL4, GF1–GF2 and under $H_0 : m \in \mathcal{M}_\Theta$,

$$nh^{\frac{q}{2}}\left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\xrightarrow{d}\mathcal{N}\left(0,2\nu^2\right).$$

- **Conditions**
- **Constants**

$$\nu^2 = \int_{\Omega_q}\sigma^4(\mathbf{x})w(\mathbf{x})^2\,\omega_q(d\mathbf{x})$$
$$\times\,\gamma_q\lambda_q(L)^{-4}\int_0^\infty r^{\frac{q}{2}-1}\left\{\int_0^\infty \rho^{\frac{q}{2}-1}L(\rho)\varphi_q(r,\rho)\,d\rho\right\}^2 dr$$

- If L is the **von Mises kernel**, $\nu^2 = \int_{\Omega_q}\sigma^4(\mathbf{x})w(\mathbf{x})^2\,\omega_q(d\mathbf{x}) \times (8\pi)^{-\frac{q}{2}}$.

Power under local alternatives

- We consider a function $g:\Omega_q \longrightarrow \mathbb{R}$ such that $g \notin \mathcal{M}_\Theta$ and the Pitman local alternative:

$$H_{1L}: m(\mathbf{x}) = m_{\boldsymbol{\theta}_0}(\mathbf{x}) + \left(nh^{\frac{q}{2}}\right)^{-\frac{1}{2}} g(\mathbf{x}),\ \forall \mathbf{x} \in \Omega_q.$$

Theorem (Power under local alternatives)

Under conditions LL1–LL4, GF2–GF4 and under H_{1L},

$$nh^{\frac{q}{2}}\left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right) \xrightarrow{d} \mathcal{N}\left(\int_{\Omega_q} g(\mathbf{x})^2 f(\mathbf{x}) w(\mathbf{x})\,\omega_q(d\mathbf{x}), 2\nu^2\right).$$

- **Conditions**

GF3 Under H_{1L}, there exists a $\widehat{\boldsymbol{\theta}}$ such that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.

GF4 The function g is continuous.

Calibration in practice

Algorithm (Testing procedure) ▸ Bootstrap consistency

Let $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ *be a random sample. To test* $H_0 : m \in \mathcal{M}_{\boldsymbol{\Theta}}$*:*

1. *Compute* $\widehat{\boldsymbol{\theta}}$*, and set* $\widehat{\varepsilon}_i = Y_i - \widehat{m}_{h,p}(\mathbf{X}_i)$*,* $i = 1, \ldots, n$*.*
2. *Compute*
$$T_n = \int_{\Omega_q} \left(\widehat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p} m_{\widehat{\boldsymbol{\theta}}}(\mathbf{x})\right)^2 \widehat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}).$$
3. *Bootstrap strategy. For* $b = 1, \ldots, B$*:*
   - *Set* $Y_i^* = m_{\widehat{\boldsymbol{\theta}}}(\mathbf{X}_i) + \widehat{\varepsilon}_i V_i$*,* $i = 1, \ldots, n$*, where* V_i *are golden binary random variables.*
   - *Compute* $\widehat{\boldsymbol{\theta}}^*$ *from* $\{(\mathbf{X}_i, Y_i^*)\}_{i=1}^n$ *and set*
   $$T_n^{*b} = \int_{\Omega_q} \left(\widehat{m}_{h,p}^*(\mathbf{x}) - \mathcal{L}_{h,p} m_{\widehat{\boldsymbol{\theta}}^*}(\mathbf{x})\right)^2 \widehat{f}_h(\mathbf{x}) w(\mathbf{x})\, \omega_q(d\mathbf{x}).$$
4. *Estimate the* p*–value as* $\#\left\{T_n^{*b} \leq T_n\right\}/B$*.*

Simulation study

Scenarios: $Y = m_{\boldsymbol{\theta}}(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$.

- Density of $\mathbf{X}$: 6 different directional densities.
- Regression function $m_{\boldsymbol{\theta}}$: 12 new directional–linear models.
- Noise $\varepsilon \sim \mathcal{N}(0,1)$, $\sigma(\mathbf{x})$ constant and variable (heteroskedastic).
- Deviations from $H_0 : m \in \mathcal{M}_{\boldsymbol{\Theta}}$ constructed by sampling from $m_\delta(\mathbf{x}) = m_{\boldsymbol{\theta}_0}(\mathbf{x}) + \delta\Delta(\mathbf{x})$. Details

Simulation setting:

- Sample sizes $n = 100, 250, 500$, dimensions $q = 1, 2, 3$, local constant $(p = 0)$ and linear $(p = 1)$ estimators.
- $B = 1000$ bootstrap replicates and $M = 1000$ Monte Carlo trials for evaluating the empirical size/power of the test.
- Bandwidth grid for exploring its effect.

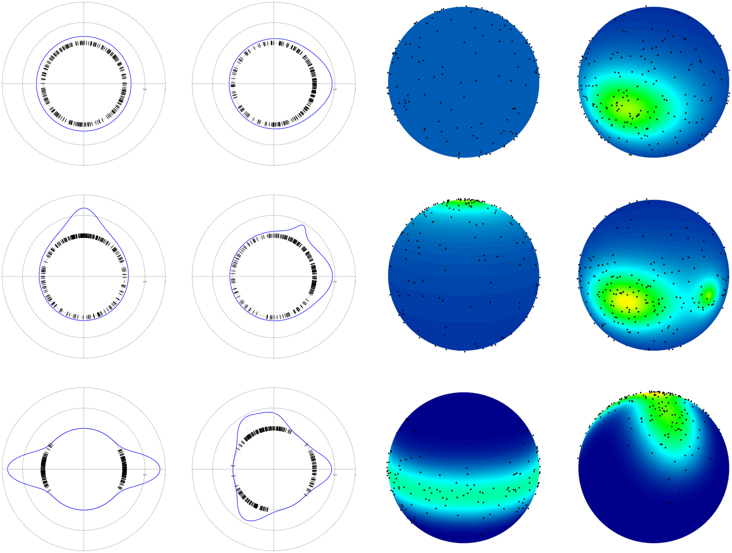

Figure : Densities D1 to D6 for the circular and spherical cases.

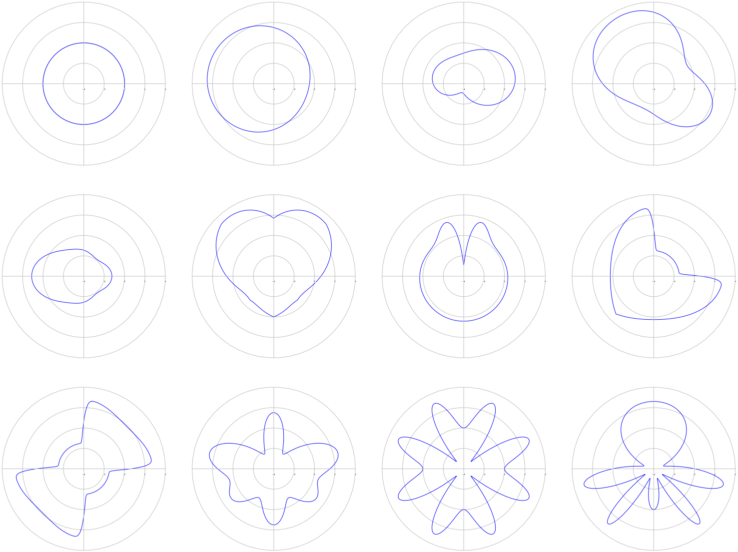

Figure : Models M1 to M12 for the circular case ($q = 1$). ▸ Details

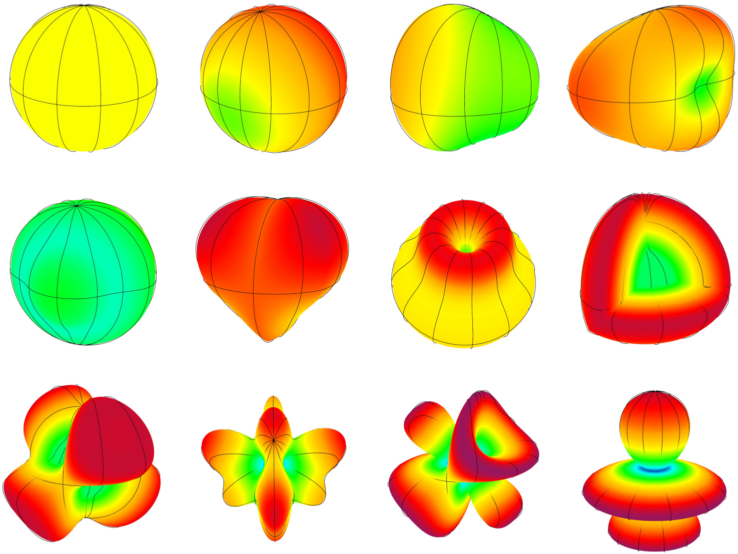

Figure : Models M1 to M12 for the spherical case ($q = 2$). ▸ Details

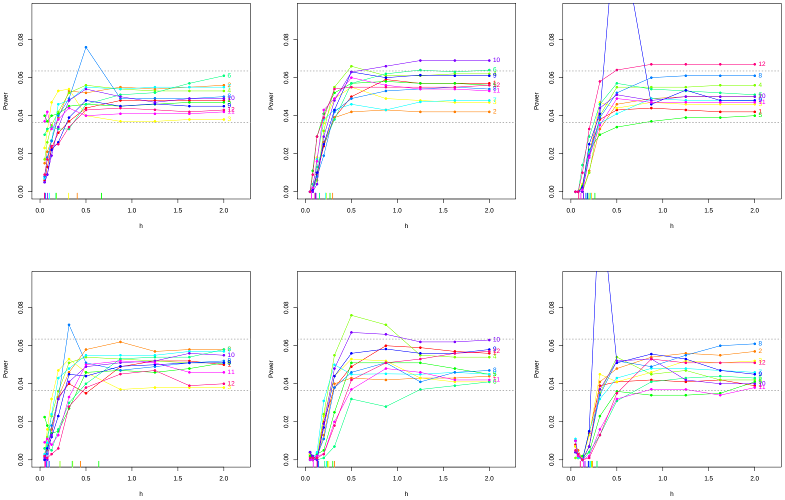

Figure : Empirical sizes for $\alpha = 0.05$ with $p = 0$ (upper row) and $p = 1$ (lower row). Columns: $q = 1, 2, 3$ and $n = 100, 250, 500$.

More sizes

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

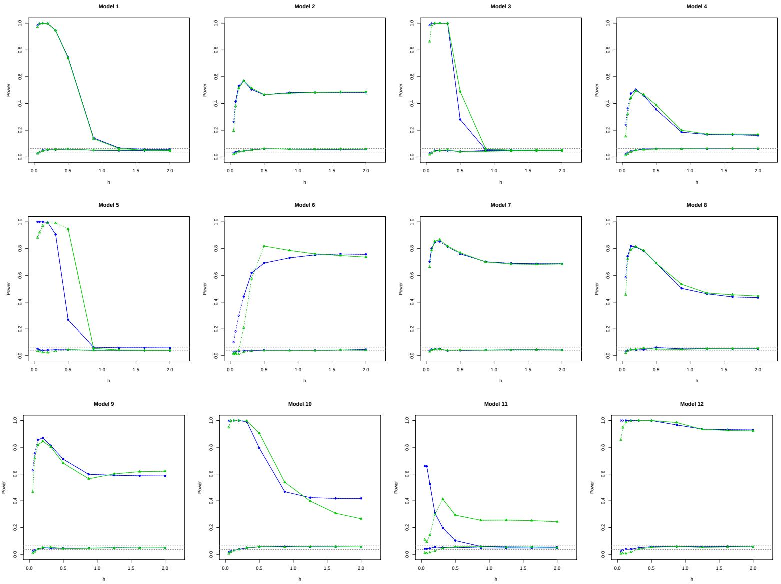

Figure : Powers for $\alpha = 0.05$ with $p = 0$ (blue) and $p = 1$ (green), for $q = 1$ and $n = 250$. ▸ More powers

Table of contents

Wildfire orientation

- Data: 102 averaged orientations and log–burnt areas of the wildfires in each watershed of Portugal in 1985–2005.

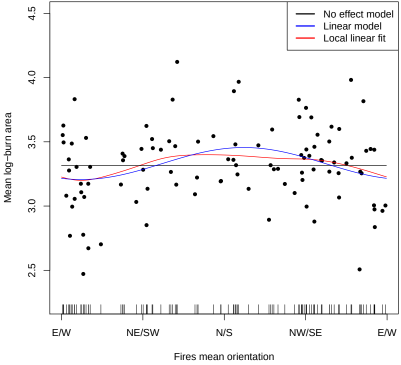

Figure : Wildfires data with two parametric fits and a local linear fit.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Wildfire orientation

- Data: 102 averaged orientations and log–burnt areas of the wildfires in each watershed of Portugal in 1985–2005.
- No effect model:

$$m(\mathbf{x}) = c.$$

- FDR p–values: 0.027 (loc. const.) and 0.047 (loc. lin.).
- **Evidence to reject the model.**

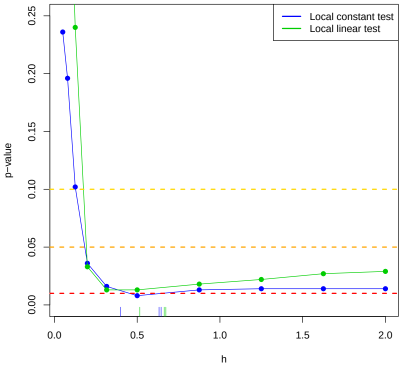

Figure : p–values of the goodness–of–fit test for no effect model.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Wildfire orientation

- Data: 102 averaged orientations and log–burnt areas of the wildfires in each watershed of Portugal in 1985–2005.
- Linear model:

$$m(\mathbf{x}) = c + \boldsymbol{\beta}^T \mathbf{x}$$

- FDR p–values: 0.230 (loc. const.) and 0.204 (loc. lin.).
- **No evidence to reject the model**.

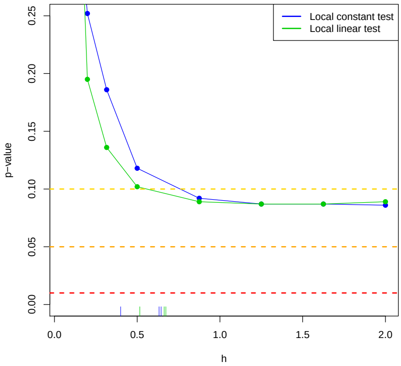

Figure : p–values of the goodness–of–fit test for linear model.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Wildfire orientation

- Data: 102 averaged orientations and log–burnt areas of the wildfires in each watershed of Portugal in 1985–2005.
- Linear model:

$$m(\theta) = c + \beta_1 \cos(\theta) + \beta_2 \sin(\theta)$$

- FDR p–values: 0.230 (loc. const.) and 0.204 (loc. lin.).
- **No evidence to reject the model**.

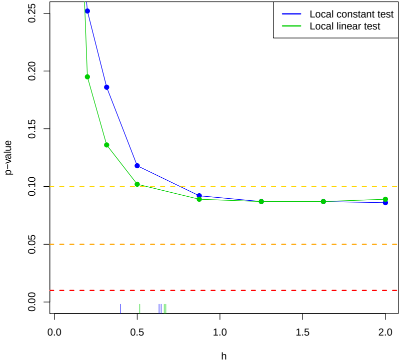

Figure : p–values of the goodness–of–fit test for linear model.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Main references

Alcalá, J. T., Cristóbal, J. A., and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Probab. Lett.*, 42(1):39–46.

Barros, A. M. G., Pereira, J. M. C., and Lund, U. J. (2012). Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecol. Manag.*, 264, 98–107.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.*, 79(19):2066–2075.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, 7:1655–1685.

García-Portugués, E., Crujeiras, R., and González-Manteiga, W. (2014). Central limit theorems for directional and linear data with applications. *arXiv:1402.6836.*

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22(3):1346–1370.

Some notes of caution in smoothing

1. **Holes in the support and local linear estimation.** Local linear smoothing can be worse than local constant smoothing when there are holes in the support.

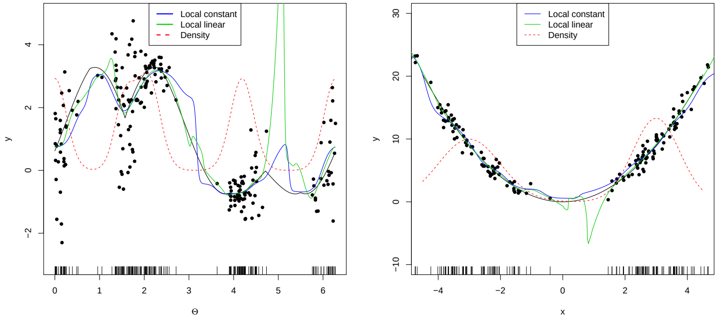

Figure : Local constant and linear estimators with predictor for samples with areas with low density.

Some notes of caution in smoothing

1. **Holes in the support and local linear estimation.**
2. **Nonparametric wild bootstrap:** $\hat{\varepsilon}_i = Y_i - \hat{m}_{h,p}(\mathbf{X}_i)$. Versus the parametric one that we have used: $\hat{\varepsilon}_i = Y_i - m_{\hat{\boldsymbol{\theta}}}(\mathbf{X}_i)$.

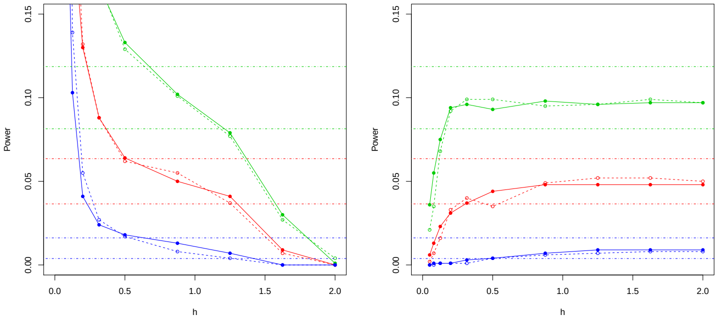

Figure : Sizes for nonparametric and parametric wild bootstraps.

Some notes of caution in smoothing

1. **Holes in the support and local linear estimation.**
2. **Nonparametric wild bootstrap:** $\hat{\varepsilon}_i = Y_i - \hat{m}_{h,p}(\mathbf{X}_i)$.
3. **Estimation bandwidths for testing.** Bandwidths arising from estimation criteria are not always a reasonable option.

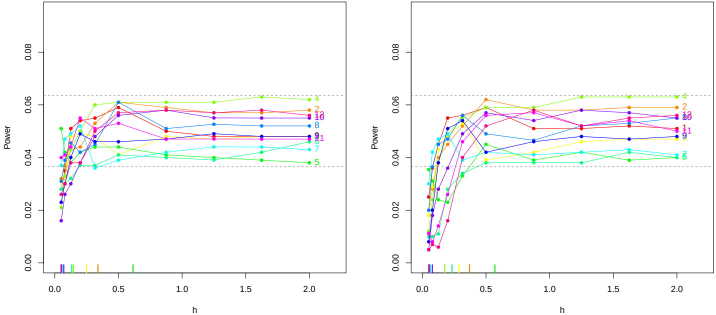

Figure : Sizes for averaged estimation bandwidths for $p = 0$ and $p = 1$.

Concept	Our proposal	Di Marzio *et al.* (2014)
Taylor expansion	$m(\mathbf{X}_i) \approx m(\mathbf{x}) + \boldsymbol{\nabla} m(\mathbf{x})^T \mathbf{B}_q \mathbf{B}_q^T (\mathbf{X}_i - \mathbf{x})$	$m(\mathbf{X}_i) \approx m(\mathbf{x}) + \theta_{\mathbf{x},i} \boldsymbol{\xi}_{\mathbf{x},i}^T \boldsymbol{\nabla} m(\mathbf{x})$, $\mathbf{X}_i = \mathbf{x}\cos(\theta_{\mathbf{x},i}) + \boldsymbol{\xi}_{\mathbf{x},i} \sin(\theta_{\mathbf{x},i})$
$\boldsymbol{\mathcal{X}}_{\mathbf{x},1}$'s i–th row	$\left(1,\, \mathbf{B}_q(\mathbf{X}_i - \mathbf{x})^T\right)_{q+1}$	$\left(1,\, \theta_{\mathbf{x},i} \boldsymbol{\xi}_{\mathbf{x},i}^T\right)_{q+2}$
Bandwidth	$h = \kappa^{-\frac{1}{2}}$	$\kappa = h^{-2}$
Kernel	$L_h(\mathbf{x}, \mathbf{X}_i) = c_{h,q}(L) L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right)$	$K_\kappa(\cos(\theta_i)) = K_\kappa(\mathbf{x}^T \mathbf{X}_i)$
Case $q = 1$	Di Marzio *et al.* (2009) estimator. If $\mathbf{x} = (\cos\theta, \sin\theta)$, $\mathbf{X}_i = (\cos\Theta_i, \sin\Theta_i)$, $\mathbf{B}_1 = (-\sin\theta, \cos\theta)^T$, then $\mathbf{B}_1(\mathbf{X}_i - \mathbf{x}) = \sin(\theta - \Theta_i)$	Different from Di Marzio *et al.* (2009) (usual vs. tangent–normal Taylor expansions)
Bias	$\frac{b_q(L)}{q} \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right] h^2$	$\frac{b_2(\kappa)}{2q} \operatorname{tr}\left[\boldsymbol{\mathcal{H}}_m(\mathbf{x})\right]$
Variance	$\frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q f(\mathbf{x})} \sigma^2(\mathbf{x})$	$\frac{\nu_0(\kappa)}{n f(\mathbf{x})} \sigma^2(\mathbf{x})$

Table : Proposals for local linear smoothing. ◂ Alternative proposal

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Model	Circular case $(q=1)$						Spherical case $(q=2)$					
	Local constant			Local linear			Local constant			Local linear		
	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}
M1	**0.009**	**0.009**	0.013	**0.009**	**0.009**	0.013	**0.003**	**0.003**	0.007	0.003	0.003	0.007
M2	0.027	0.026	0.041	0.025	**0.023**	0.068	0.025	0.024	0.063	0.022	**0.021**	0.068
M3	0.039	0.037	0.063	0.033	**0.031**	0.038	0.038	0.038	0.067	0.030	**0.029**	0.047
M4	0.056	0.056	0.253	**0.046**	**0.045**	0.274	0.044	0.044	0.045	0.041	**0.040**	0.043
M5	0.040	0.044	0.044	0.027	0.032	**0.025**	**0.030**	**0.030**	0.065	0.034	0.036	0.042
M6	0.051	0.050	0.059	**0.046**	0.048	0.114	**0.030**	**0.030**	0.057	0.031	0.031	0.033
M7	0.043	0.042	0.320	0.043	**0.040**	0.317	0.062	0.062	0.062	0.061	0.061	**0.060**
M8	0.048	0.048	0.586	0.047	**0.047**	0.484	**0.161**	0.163	0.803	0.167	0.178	0.779
M9	0.058	0.057	0.793	0.054	**0.052**	1.068	0.130	0.124	1.095	0.109	**0.109**	1.368
M10	0.036	0.036	0.102	0.032	**0.031**	0.079	0.111	0.107	0.527	0.107	**0.105**	0.676
M11	0.030	0.029	0.094	0.202	**0.022**	0.040	0.165	0.160	1.272	0.158	**0.130**	0.679
M12	0.059	0.059	0.293	0.081	**0.042**	0.604	0.061	0.060	0.073	0.059	**0.050**	0.054

Table : Empirical ASEs with $n = 250$. Significative best combinations of estimator and bandwidth selector are marked in bold. ◂ Bandwidth selection

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Model	Circular case $(q=1)$						Spherical case $(q=2)$					
	Local constant			Local linear			Local constant			Local linear		
	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}
M1	**0.008**	**0.008**	0.017	**0.008**	0.009	0.018	**0.001**	**0.001**	0.003	**0.001**	**0.001**	0.003
M2	0.046	0.046	0.093	0.044	**0.042**	0.087	0.015	0.015	0.042	0.013	**0.013**	0.070
M3	0.065	0.066	0.098	0.056	**0.054**	0.090	0.024	0.024	0.044	**0.018**	**0.018**	0.028
M4	0.077	0.077	0.076	0.074	**0.074**	0.080	0.029	0.029	0.029	0.026	**0.026**	0.026
M5	0.056	**0.055**	0.113	0.064	0.066	0.084	**0.019**	0.019	0.042	0.021	0.023	0.023
M6	**0.048**	0.049	0.081	0.050	**0.049**	0.054	**0.021**	**0.021**	0.040	0.022	0.021	0.022
M7	0.088	0.088	**0.082**	0.087	0.087	0.084	0.046	0.046	0.050	0.045	**0.045**	0.049
M8	**0.232**	0.254	0.802	0.247	0.279	0.753	**0.121**	0.121	0.756	0.122	0.125	0.766
M9	0.199	0.175	1.104	0.170	**0.162**	1.201	0.092	0.090	1.113	0.077	**0.076**	1.392
M10	0.167	**0.153**	0.679	0.183	0.155	0.703	0.080	0.079	0.380	0.074	**0.073**	0.555
M11	0.246	0.217	1.053	0.542	**0.192**	0.564	0.121	0.120	1.370	0.115	**0.095**	0.731
M12	0.099	0.096	0.118	0.121	**0.083**	0.096	0.042	0.042	0.052	0.040	**0.035**	0.037

Table : Empirical ASEs with $n = 500$. Significative best combinations of estimator and bandwidth selector are marked in bold. Bandwidth selection

Model	Circular case $(q=1)$						Spherical case $(q=2)$					
	Local constant			Local linear			Local constant			Local linear		
	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}	h_{CV}	h_{GCV}	h_{PI}
M1	**0.022**	0.024	0.033	**0.023**	0.023	0.034	**0.005**	**0.005**	0.007	**0.005**	0.005	0.007
M2	0.063	0.058	0.078	0.055	**0.052**	0.086	0.016	0.015	0.027	0.014	**0.013**	0.076
M3	0.085	0.079	0.124	0.074	**0.068**	0.083	0.022	0.021	0.037	0.019	**0.017**	0.022
M4	0.119	0.117	0.344	**0.100**	**0.100**	0.330	0.034	0.033	0.146	0.027	**0.026**	0.220
M5	0.071	0.067	0.093	**0.058**	0.062	**0.059**	0.022	0.030	0.026	0.015	0.019	**0.014**
M6	0.106	0.105	0.116	**0.091**	0.094	0.183	0.031	0.030	0.036	**0.028**	0.029	0.087
M7	0.165	0.162	0.309	**0.152**	**0.149**	0.304	0.023	0.023	0.321	**0.022**	**0.022**	0.320
M8	0.091	**0.088**	0.814	0.099	0.092	0.617	0.030	0.030	0.377	0.028	**0.028**	0.367
M9	0.102	**0.097**	0.854	0.119	0.100	1.052	0.040	0.040	0.761	0.034	**0.034**	1.041
M10	0.071	0.070	0.180	0.067	**0.064**	0.134	0.021	0.021	0.059	0.018	**0.018**	0.051
M11	0.091	0.047	0.147	0.331	**0.041**	0.081	0.023	0.023	0.062	0.082	**0.015**	0.025
M12	0.092	0.091	0.409	0.249	**0.073**	0.800	0.045	0.044	0.202	0.042	**0.029**	0.478

Table : Empirical ASEs with $n = 500$. Significative best combinations of estimator and bandwidth selector are marked in bold. Bandwidth selection

Sketch of the proof

Limit distribution of T_n

$$nh^{\frac{q}{2}}\left(T_n-\frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\stackrel{d}{\longrightarrow}\mathcal{N}\left(0,2\nu^2\right)$$

1. Decompose $T_n=(T_{n_1}+T_{n_2}-2T_{n_3})(1+o_{\mathbb{P}}\left(1\right))$ using the equivalent kernel formulation:

$$T_{n_1}=\int_{\Omega_q}\bigg(\sum_{i=1}^n L_p^*\left(\mathbf{x},\mathbf{X}_i\right)\left(Y_i-m_{\boldsymbol{\theta}_0}(\mathbf{X}_i)\right)\bigg)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}),$$

$$T_{n_2}=\int_{\Omega_q}\left(\mathcal{L}_{h,p}\left(m_{\boldsymbol{\theta}_0}-m_{\hat{\boldsymbol{\theta}}}\right)(\mathbf{x})\right)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}),$$

$$T_{n_3}=\int_{\Omega_q}\left(\hat{m}_{h,p}(\mathbf{x})-\mathcal{L}_{h,p}m_{\boldsymbol{\theta}_0}(\mathbf{x})\right)\mathcal{L}_{h,p}\left(m_{\boldsymbol{\theta}_0}-m_{\hat{\boldsymbol{\theta}}}\right)(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}).$$

Sketch of the proof

Limit distribution of T_n

$$nh^{\frac{q}{2}}\left(T_n-\frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\stackrel{d}{\longrightarrow}\mathcal{N}\left(0,2\nu^2\right)$$

1. Decompose $T_n=(T_{n_1}+T_{n_2}-2T_{n_3})(1+o_{\mathbb{P}}\left(1\right))$ using the equivalent kernel formulation:

$$T_{n_1}=\int_{\Omega_q}\bigg(\sum_{i=1}^n L_p^*\left(\mathbf{x},\mathbf{X}_i\right)\left(Y_i-m_{\boldsymbol{\theta}_0}(\mathbf{X}_i)\right)\bigg)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}),$$

$$T_{n_2}=\int_{\Omega_q}\left(\mathcal{L}_{h,p}\left(m_{\boldsymbol{\theta}_0}-m_{\widehat{\boldsymbol{\theta}}}\right)(\mathbf{x})\right)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}),$$

$$T_{n_3}=\int_{\Omega_q}\left(\widehat{m}_{h,p}(\mathbf{x})-\mathcal{L}_{h,p}m_{\boldsymbol{\theta}_0}(\mathbf{x})\right)\mathcal{L}_{h,p}\left(m_{\boldsymbol{\theta}_0}-m_{\widehat{\boldsymbol{\theta}}}\right)(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}).$$

2. $nh^{\frac{q}{2}}T_{n_2}=o_{\mathbb{P}}\left(1\right)$ and $nh^{\frac{q}{2}}T_{n_3}=o_{\mathbb{P}}\left(1\right)$ by $\sqrt{n}$–consistency of $\hat{\boldsymbol{\theta}}$.

Sketch of the proof ◂ Limit distribution of T_n

$$nh^{\frac{q}{2}}\left(T_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right) \stackrel{d}{\longrightarrow} \mathcal{N}\left(0, 2\nu^2\right)$$

1. Decompose $T_n = (T_{n_1} + T_{n_2} - 2T_{n_3})(1 + o_{\mathbb{P}}(1))$ using the equivalent kernel formulation:

$$T_{n_1} = \int_{\Omega_q}\Big(\sum_{i=1}^n L_p^*(\mathbf{x}, \mathbf{X}_i)\left(Y_i - m_{\boldsymbol\theta_0}(\mathbf{X}_i)\right)\Big)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}),$$

$$T_{n_2} = \int_{\Omega_q}\left(\mathcal{L}_{h,p}\left(m_{\boldsymbol\theta_0} - m_{\widehat{\boldsymbol\theta}}\right)(\mathbf{x})\right)^2 f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}),$$

$$T_{n_3} = \int_{\Omega_q}\left(\widehat{m}_{h,p}(\mathbf{x}) - \mathcal{L}_{h,p}m_{\boldsymbol\theta_0}(\mathbf{x})\right)\mathcal{L}_{h,p}\left(m_{\boldsymbol\theta_0} - m_{\widehat{\boldsymbol\theta}}\right)(\mathbf{x})f(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x}).$$

2. $nh^{\frac{q}{2}}T_{n_2} = o_{\mathbb{P}}(1)$ and $nh^{\frac{q}{2}}T_{n_3} = o_{\mathbb{P}}(1)$ by $\sqrt{n}$–consistency of $\hat{\boldsymbol\theta}$.

3. Split $T_{n_1} = T_{n_1}^{(1)} + T_{n_1}^{(2)} + 2T_{n_1}^{(3)}$ where $nh^{\frac{q}{2}}T_{n_1}^{(2)} = o_{\mathbb{P}}(1)$ and $nh^{\frac{q}{2}}T_{n_1}^{(3)} = o_{\mathbb{P}}(1)$. Finally, by de Jong (1987)'s CLT,

$$nh^{\frac{q}{2}}\left(T_{n_1}^{(1)} - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right) \stackrel{d}{\longrightarrow} \mathcal{N}\left(0, 2\nu^2\right).$$

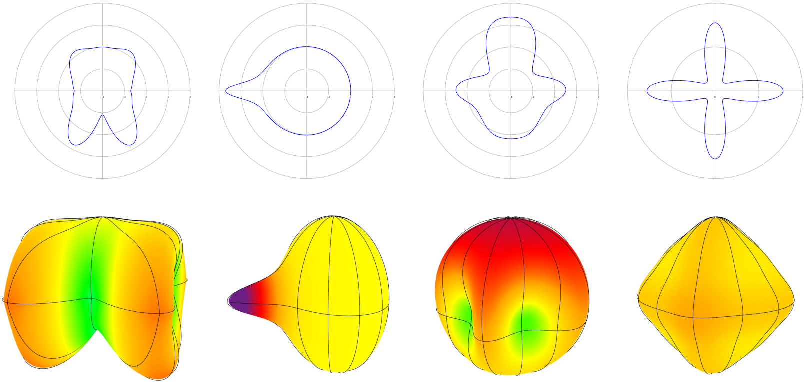

Figure : Deviations Δ_1, Δ_2 and Δ_3 (from left to right, first three columns) and conditional standard deviation for the heteroskedastic noise (fourth column). Simulation study

Theorem (Bootstrap consistency) Calibration in practice

Under conditions LL1–LL4, GF1–GF5 and under the hypothesis H_{1L}*,*

$$nh^{\frac{q}{2}}\left(T_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q}\int_{\Omega_q}\sigma^2(\mathbf{x})w(\mathbf{x})\,\omega_q(d\mathbf{x})\right)\xrightarrow{d}\mathcal{N}\left(0,2\nu^2\right)$$

with probability one.

- A **bootstrap** analogue of **condition** GF1 is required:

GF5 Under H_{1L}, and for $\mathbf{X}_1,\ldots,\mathbf{X}_n$, there exists an estimator $\widehat{\boldsymbol{\theta}}^*$ such that $\widehat{\boldsymbol{\theta}}-\widehat{\boldsymbol{\theta}}^*=\mathcal{O}_{\mathbb{P}^*}\left(n^{-\frac{1}{2}}\right)$.

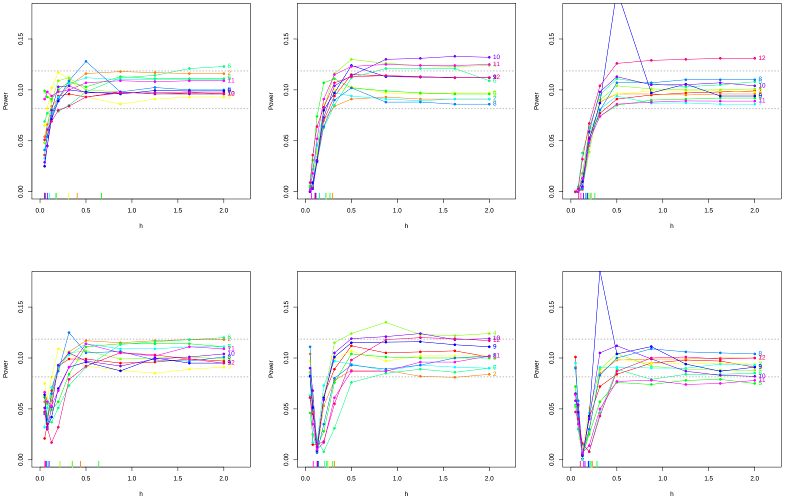

Figure : Empirical sizes for $\alpha = 0.10$ with $p = 0$ (upper row) and $p = 1$ (lower row). Columns: $q = 1, 2, 3$ and $n = 100, 250, 500$. Simulation

Model	Regression function	Density	Noise	Deviation
M1	$m(\mathbf{x}) = m$	D1	Het.	$0.5\Delta_1(\mathbf{x})$
M2	$m(\mathbf{x}) = m + \boldsymbol{\eta}^T\mathbf{x}$	D2	Het.	$-0.25\Delta_1(\mathbf{x})$
M3	$m(\mathbf{x}) = m + \boldsymbol{\eta}^T\mathbf{x} + \boldsymbol{\gamma}^T(x_1^2, \ldots, x_q^2)$	D3	Het.	$-\Delta_1(\mathbf{x})$
M4	$m(\mathbf{x}) = m + a(x_{q+1}^2 - x_q^3) + bx_1x_2$	D4	Het.	$-0.25\Delta_1(\mathbf{x})$
M5	$m(\mathbf{x}) = m + ax_1^4 + b\left(\frac{3}{2} - x_1\right)^{-\frac{1}{2}}$	D5	Het.	$-\Delta_2(\mathbf{x})$
M6	$m(\mathbf{x}) = ae^{bx_2}\log(cx_{q+1}\lvert x_1\rvert)$ $+d\max(\lvert x_1\rvert, \lvert x_2\rvert)$	D6	Het.	$-3\Delta_2(\mathbf{x})$
M7	$m(\mathbf{x}) = m + d_1 f_{vM}(\mathbf{x}, (\mathbf{0}_q, 1), \kappa_1)$ $-d_2 f_{Ca}(\mathbf{x}, (\mathbf{0}_q, 1), \kappa_2)$	D1	Hom.	$-0.3\Delta_2(\mathbf{x})$
M8	$m(\mathbf{x}) = m + af_{SN}(\max(-\mathbf{x}), b, c, d)$	D2	Hom.	$-0.4\Delta_2(\mathbf{x})$
M9	$m(\mathbf{x}) = m + af_{SN}(\prod_{i=1}^{q+1} x_i, b, c, d)$	D3	Hom.	$0.1\Delta_3(\mathbf{x})$
M10	$m(\mathbf{x}) = m + a\sin(2\pi x_2) + b\cos(2\pi x_1)$	D4	Hom.	$0.25\Delta_3(\mathbf{x})$
M11	$m(\mathbf{x}) = m + a\sin(2\pi b x_1 x_{q+1})$	D5	Hom.	$-1.5\Delta_3(\mathbf{x})$
M12	$m(\mathbf{x}) = m + a\sin\left(2\pi b(2 + x_{q+1})^{-1}\right)$	D6	Hom.	$0.3\Delta_3(\mathbf{x})$

Table : Simulation scenarios. ◂ Models

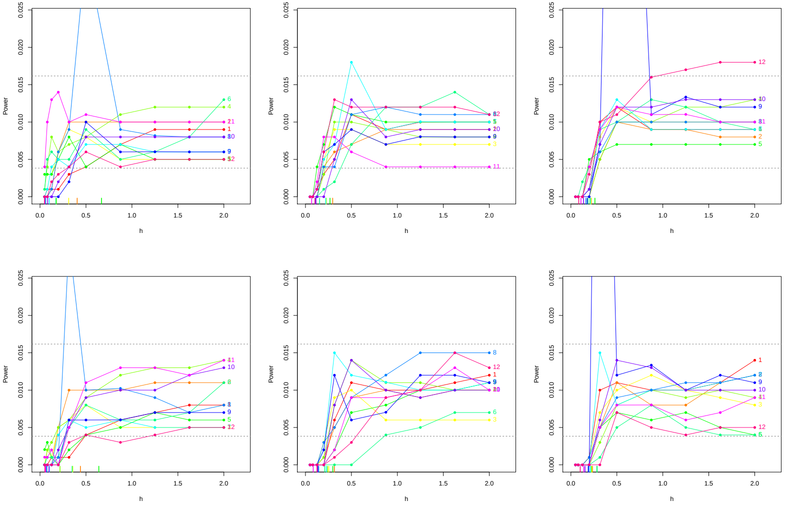

Figure : Empirical sizes for $\alpha = 0.01$ with $p = 0$ (upper row) and $p = 1$ (lower row). Columns: $q = 1, 2, 3$ and $n = 100, 250, 500$. Simulation

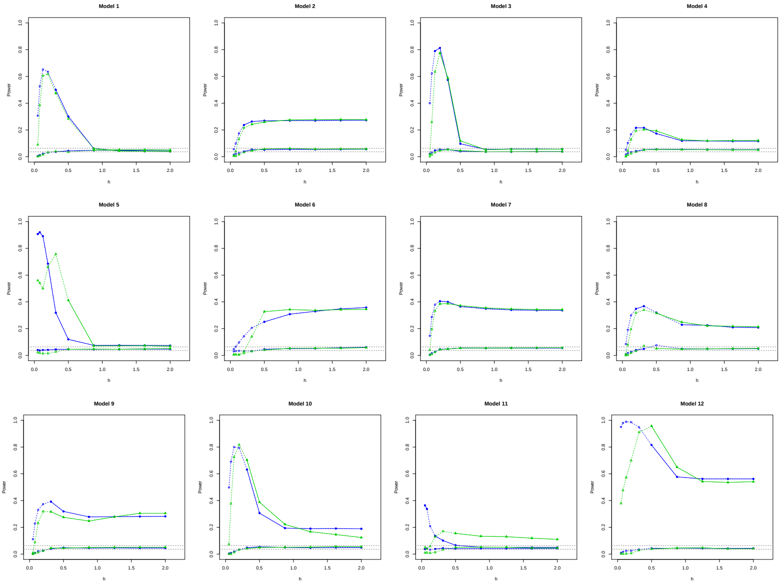

Figure : Empirical powers for $\alpha = 0.05$ with $p = 0$ (blue) and $p = 1$ (green), for $q = 1$ and $n = 100$. ◂ Simulation

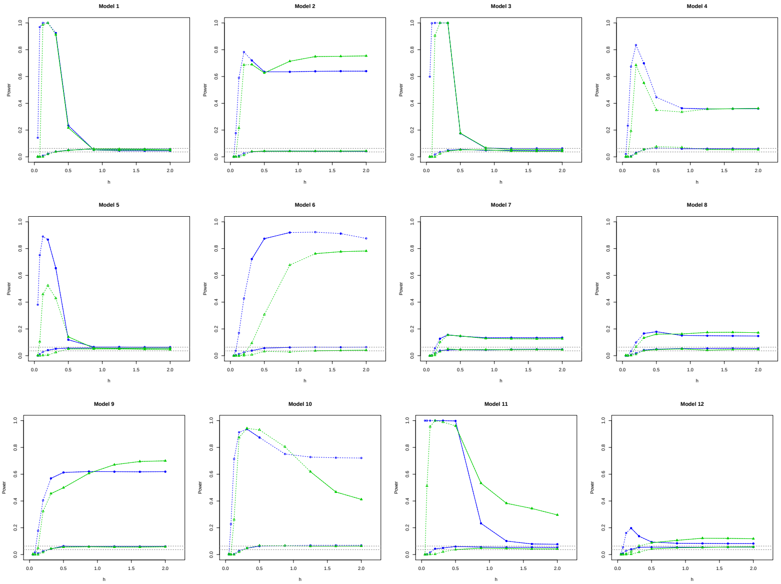

Figure : Empirical powers for $\alpha = 0.05$ with $p = 0$ (blue) and $p = 1$ (green), for $q = 2$ and $n = 250$. ◂ Simulation

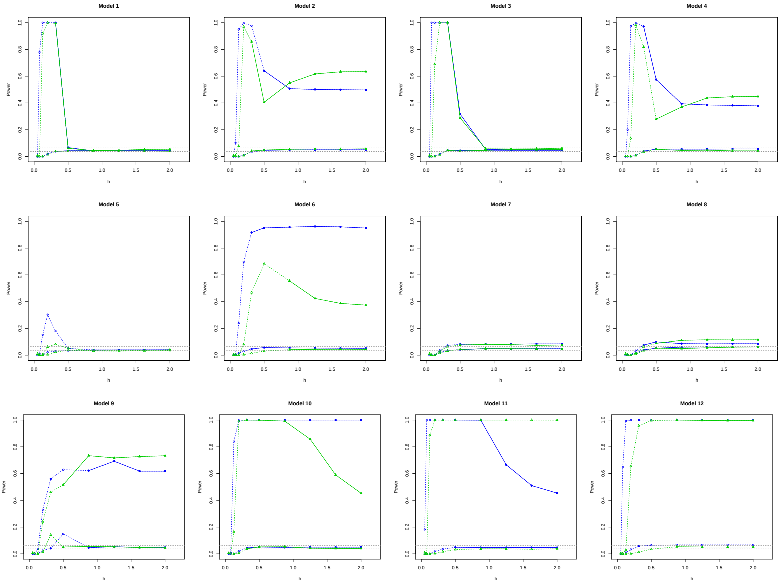

Figure : Empirical powers for $\alpha = 0.05$ with $p = 0$ (blue) and $p = 1$ (green), for $q = 3$ and $n = 500$. ◂ Simulation