Revista Colombiana de Estadística
Special Issue: Colombian and Brazil Graduate Programs
December 2025, volume 48, issue 3, pp. 319–348
http://doi.org/10.15446/rce.v48n3.122634

Conditional Mode: An Approach via Smoothed Quantile Regression

Moda condicional: un enfoque vía regresión cuantílica suavizada

Artur Ongaratto[1,a], Eduardo Horta[2,b]

[1]Sicredi, Brazil

[2]Department of Statistics, Institute of Mathematics and Statistics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Abstract

Recently, it has been proposed to estimate the conditional mode of a response, given a vector of covariates, using a computationally scalable estimator derived from the linear quantile regression model. Alternatively, we propose to estimate the conditional mode by maximizing a smoothed conditional density estimator. This approach offers at least two benefits: computational efficiency and good asymptotic behavior which, in particular, bypasses the curse of dimensionality.

Keywords: Convolution; Mode regression.

Resumen

Recientemente, se ha propuesto estimar la moda condicional de una respuesta, dado un vector de covariables, mediante un estimador computacionalmente escalable derivado del modelo de regresión cuantílica lineal. Como alternativa, proponemos estimar la moda condicional maximizando un estimador de densidad condicional suavizada. Este enfoque ofrece al menos dos ventajas: eficiencia computacional y un buen comportamiento asintótico que, en particular, evita la maldición de la dimensionalidad.

Palabras clave: Convolución; Regresión modal.

[a]Masters in Statistics. E-mail: artur_ongaratto@sicredi.com.br
[b]Ph.D. E-mail: eduardo.horta@ufrgs.br

1. Introduction

The advantages of using the mode as a measure of central tendency in a univariate scenario are well known. In addition to being insensitive to outliers, when dealing with skewed or heavy-tailed distributions, the mode allows one to capture information often ignored by the mean or the median. However, in practice, the unconditional configuration is quite limited. In many scenarios, the value of a variable of interest (the response) is partially determined by the values of other variables (the covariates or regressors), these known to the researcher. In this milieu, regression analysis is possibly the most used statistical method to model the dependency structure between a response and covariates. In particular, mode regression specifies the functional form of the conditional mode of a response variable Y, given a vector of covariates X. When the random variables in question are continuous, the conditional mode, or "conditionally most likely" value, is nothing more than the location parameter that maximizes the conditional probability density function. This form of regression has been used in several areas, such as economics, environment Ullah et al. (2021), astronomy Bamford et al. (2008), machine learning Feng et al. (2020), medicine Wang et al. (2017) and geology Chacón (2020). In the literature, there are different approaches to estimate the conditional mode. The best known are non-parametric methods, as in Chen et al. (2016), based on a kernel density estimator, and linear mode regression, where the conditional mode is assumed to be linear in the covariates, as in Lee (1989) and Yao & Li (2014). However, the first alternative has a low convergence rate (for example, with complexity $n^{-2/(d+7)}$ in the case of Chen et al. (2016), where d represents the number of covariates in the model), while the second, despite circumventing the curse of dimensionality, involves a multidimensional, non-convex optimization problem.

To avoid the problems of these two approaches, recently Ota et al. (2019) proposed to estimate the conditional mode through a conditional quantile density estimator, obtained via numerical differentiation of the Koenker & Bassett (1978) estimator for the conditional quantile function. With this method, the multidimensional optimization problem becomes a one-dimensional problem that can be solved by a simple grid search and, consequently, avoids the curse of dimensionality. Furthermore, the proposal of the authors allows greater flexibility in the functional form of the conditional mode, assuming linearity only in the conditional quantile function.

To illustrate the practical relevance of conditional mode regression, consider an example from biomedical research. In medical studies, the mode can represent the most likely physiological or biochemical response of a patient, even when the outcome distribution is skewed or affected by outliers (see Wang et al., 2017). This example highlights that conditional mode regression may provide a more robust and interpretable characterization of the response–covariate relationship in the presence of asymmetry or heavy tails.

Inspired by the approach of Ota et al. (2019), we propose a new estimator for the conditional mode, obtained by minimizing a conditional quantile density computed not by numerical differentiation, but through the smoothing methods of

Fernandes et al. (2021), which allow one to obtain a differentiable estimator for the conditional quantile function and, *a fortiori*, of the conditional quantile density. Compared with the estimator proposed by Ota et al. (2019), our approach provides both theoretical and computational advantages. By relying on the smoothed quantile regression framework of Fernandes et al. (2021), the proposed estimator avoids numerical differentiation of the quantile regression coefficients, yielding a continuously differentiable objective function. This property ensures numerical stability, facilitates analytical derivations, and results in estimators with lower variance and mean squared error. Moreover, the smoothing scheme allows the use of data-driven and possibly τ-dependent bandwidths, which enhances flexibility while maintaining computational efficiency. Overall, these features make the proposed estimator a competitive and robust alternative for conditional mode estimation.

The estimator for the quantile function introduced by these authors has good asymptotic properties: it is asymptotically unbiased and normally distributed and has lower variance and mean squared error compared to the canonical estimator. We start from the premise that these good properties must be inherited by the estimator we are proposing, and we perform Monte Carlo simulations to study their asymptotic behavior in five different data generating process scenarios, comparing our results to those obtained using the Ota et al. (2019) estimator. Regarding the mean squared error, our proposed estimator was systematically superior to the other in most scenarios. Leaving the "theoretical world", we also employ our estimator to reproduce the application to real data from Ota et al. (2019). In this application, data from a power plant were collected, and conditional modes for the production of electricity were estimated. Then, using the split conformal prediction procedure of Lei et al. (2018), we created prediction intervals for the conditional mode, for which once again the results were compared with those of Ota et al. (2019).

The layout of the article is as follows: In Section 2, we formally introduce our estimator for the conditional mode, based on the estimator for the conditional density of Fernandes et al. (2021), and we discuss the smoothing parameter for the kernel. We also present the estimator of Ota et al. (2019). In Section 3, we conduct a Monte Carlo study to compare the performances of these two estimators with respect to bias and mean squared error. Then, in the Section 4, we reproduce the application of Ota et al. (2019), which used data collected from a power plant to estimate conditional densities (and, later, conditional modes) of net hourly electrical energy output given exhaust vacuum, as well as for creating prediction intervals. Finally, in Section 5, we conclude with a general analysis of what will be shown over the next few pages, as well as suggestions for future work.

2. Methodology

Let Y be a scalar random variable, and let $X \in \mathbb{R}^d$ be a vector of regressors. The τ-th conditional quantile of Y given $X = x$ is defined, for $\tau \in (0, 1)$ and $x \in \text{support}(X)$, as the real number $Q(\tau|x)$ given by

$$Q(\tau|x) := \inf\{y \in \mathbb{R} : F(y|x) \geq \tau\}, \tag{1}$$

where $F(\cdot|x)$ denotes the conditional cumulative distribution function of the response variable Y given $X = x$. The map $\tau \mapsto Q(\tau|x)$ is called the conditional quantile function of Y given $X = x$. Mimicking the classical linear regression equation and aiming at parsimony, Koenker & Bassett (1978) proposed, in place of the fully nonparametric quantile regression Model (1), the linear specification

$$Q(\tau|x) = x'\beta(\tau), \quad \tau \in (0,1),\, x \in \text{support}(X), \tag{2}$$

where $\beta\colon (0,1) \to \mathbb{R}^d$ is the functional parameter of interest. In this context — and assuming henceforth that the linear Model (2) above holds — the following relation is particularly important

$$f\big(Q(\tau\,|\,x)\big|x\big) = \frac{1}{q(\tau|x)} \equiv \frac{1}{x'\beta^{(1)}(\tau)}, \quad \tau \in (0,1),\, x \in \text{support}(X), \tag{3}$$

where $q(\tau|x) := \partial/\partial\tau Q(\tau|x)$ and $\beta^{(1)}(\tau) = \partial/\partial\tau\beta(\tau)$, valid as long as $F(\cdot|x)$ has a continuous density $f(\cdot|x)$. In this case, for τ and x as above, setting

$$\mathtt{f}(\tau|x) := \left(Q(\tau|x),\, \frac{1}{q(\tau|x)}\right) \equiv \left(x'\beta(\tau),\, \frac{1}{x'\beta^{(1)}(\tau)}\right), \tag{4}$$

we see that the range of the path $\tau \mapsto \mathtt{f}(\tau|x)$ coincides with the graph of $\big(y, f(y|x)\big)$. In particular, one can define the conditional mode of Y given $X = x$ to be the function $\mathrm{m}(x) := Q(\tau_x|x)$, $x \in \text{support}(X)$, where τ_x minimizes, with respect to $\tau \in (0,1)$, the quantile density $q(\tau|x)$.[1] Now, for each $\tau \in (0,1)$ fixed, the vector $\beta(\tau)$ in (2) solves a minimization problem analogous to the one found in the case of the conditional expectation in linear regression, and its objective function can be expressed as

$$R\,(b;\tau) := \mathbb{E}\left[\rho_\tau\left(Y - X'b\right)\right], \quad \tau \in (0,1),\, b \in \mathbb{R}^d, \tag{5}$$

where $\rho_\tau(u) := u \cdot (\tau - \mathbb{I}(u \leq 0))$, for $\tau \in (0,1)$ and $u \in \mathbb{R}$, is called the check function. While (5) represents the population objective function, its sample equivalent, as proposed by Koenker & Bassett (1978), has the form

$$\widehat{R}(b;\tau) := \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(Y_i - X_i'b\right), \quad \tau \in (0,1),\, b \in \mathbb{R}^d, \tag{6}$$

where (Y_i, X_i), $i \in \{1, \ldots, n\}$ is a random sample of (Y, X). For $\tau \in (0,1)$, the quantile regression canonical estimator is the vector $\widehat{\beta}(\tau)$ given by

$$\widehat{\beta}(\tau) = \arg\min_{b \in \mathbb{R}^d} \widehat{R}\,(b;\tau). \tag{7}$$

As shown by Bassett & Koenker (1982, p. 409), the empirical conditional quantile function $\tau \mapsto x'\widehat{\beta}(\tau)$, estimated using (7), has jumps, which makes it difficult to

[1] Of course, some regularity assumptions about $f(\cdot|x)$ are required to ensure that m is well defined.

explore the relationship in (3) to estimate the conditional density and, *a fortiori*, the conditional mode. To avoid this problem, Fernandes et al. (2021) suggest estimating the parameter $\beta(\tau)$ in (2) differently, replacing the check function in (6) by a smoothed version through a kernel, similar to Nadaraya (1964). As a result, instead of the sample objective function $\widehat{R}(b;\tau)$, we have a smoothed version

$$\widehat{R}_h(b;\tau) := \frac{1}{n}\sum_{i=1}^{n} k_h * \rho_\tau(Y_i - X_i'b), \quad \tau \in (0,1),\, b \in \mathbb{R}^d, \tag{8}$$

where $*$ denotes the convolution operation[2] and where k_h is a kernel with smoothing parameter $h > 0$ (h tends to 0 as the sample size increases). Analogously to (7), for $\tau \in (0,1)$, the smoothed quantile regression estimator is the vector $\widehat{\beta}_h(\tau)$ given by

$$\widehat{\beta}_h(\tau) = \arg\min_{b\in\mathbb{R}^d} \widehat{R}_h(b;\tau). \tag{9}$$

As shown in Fernandes et al. (2021), unlike the objective function in (6), the one expressed in (8) is continuously differentiable, and this has at least two benefits. The first is that the smoothed estimator in (9) inherits the regularity of the smoothed objective function (8). The second is that differentiability allows us to estimate the asymptotic covariance matrix of the slope coefficients canonically — for example, see Newey & McFadden (1994).

Incidentally, the properties mentioned above allow one to obtain, in a natural way, an estimator for the conditional probability density function of Y given $X = x$: the second derivative of the smoothed sampling objective function (8) of Fernandes et al. (2021) is given by

$$\widehat{R}_h^{(2)}(b;\tau) = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' k_h\big(-(Y_i - X_i'b)\big), \quad \tau \in (0,1),\, b \in \mathbb{R}^d.$$

Furthermore, since $\widehat{\beta}_h(\tau)$ satisfies the condition $\widehat{R}_h^{(1)}\big(\widehat{\beta}_h(\tau);\tau\big) = 0$, where $\widehat{R}_h^{(1)}(b;\tau)$ denotes the gradient of $\widehat{R}_h(b;\tau)$ with respect to b, it follows by the implicit function theorem in Marsden & Tromba (2012) that $\widehat{\beta}_h(\tau)$ is continuously differentiable with respect to τ, with

$$\frac{\partial}{\partial\tau}\widehat{\beta}_h(\tau) =: \widehat{\beta}_h^{(1)}(\tau) = \Big[\widehat{R}_h^{(2)}\Big(\widehat{\beta}_h(\tau);\tau\Big)\Big]^{-1}\bar{X}. \tag{10}$$

The authors use this fact to estimate $\mathtt{f}(\tau|x)$, replacing $\beta(\tau)$ by $\widehat{\beta}_h(\tau)$ in (4) to get

$$\widehat{\mathtt{f}}_h(\tau|x) := \left(x'\widehat{\beta}_h(\tau), \frac{1}{x'\widehat{\beta}_h^{(1)}(\tau)}\right), \quad \tau \in (0,1), x \in \text{support}(X). \tag{11}$$

By Proposition 1 in Fernandes et al. (2021), this estimator is consistent for $\mathtt{f}$, uniformly with respect to τ and h.

[2]Given two real functions φ and ψ satisfying an integrability condition, the convolution of φ and ψ is defined by the equality $\varphi * \psi(u) = \int_{\mathbb{R}} \varphi(v)\psi(u-v)\,\mathrm{d}v$.

2.1. Estimator for Conditional Mode

We propose to employ an adaptation of the estimator $\widehat{\mathtt{f}}_h(\tau|x)$ of Fernandes et al. (2021) as an intermediate step to estimate the conditional mode $\mathrm{m}(x)$. The idea is to "plug in" a smoothing parameter that is determined by the data (and possibly τ-dependent) in place of the bandwidth h in (11). In this direction, suppose that $\eta := \{\eta(\tau)\colon \tau \in \mathscr{T}\}$ is a stochastic process with state-space $\big[h_{[n]}, h^{[n]}\big]$ (note that implicitly η depends on n), where $\mathscr{T}$ is a compact subset of the unit interval $(0,1)$ specified by the user and where $1/h_{[n]} = o\big((n/\log n)^{1/3}\big)$ and $h^{[n]} = o(1)$. We define the smoothed conditional mode estimator via

$$\widehat{\mathrm{m}}_\eta(x) := x'\widehat{\beta}_{\eta(\widehat{\tau}_{x\eta})}\big(\widehat{\tau}_{x\eta}\big), \quad x \in \mathrm{support}(X), \tag{12}$$

where $\widehat{\tau}_{x\eta} = \arg\min_{\{\tau\in\mathscr{T}\}} x'\widehat{\beta}^{(1)}_{\eta(\tau)}(\tau)$, with $\widehat{\beta}^{(1)}_h$ defined as in (10). This estimator appears as an alternative to the one proposed by Ota et al. (2019), who also explore the relationship in (3) to estimate the conditional mode by

$$\widehat{\mathrm{m}}^{\mathrm{o}}_h(x) := x'\widehat{\beta}\big(\widehat{\tau}^{\mathrm{o}}_{xh}\big), \quad x \in \mathrm{support}(X), \tag{13}$$

where $\widehat{\tau}^{\mathrm{o}}_{xh}$ satisfies the inequality

$$\frac{x'\big(\widehat{\beta}(\widehat{\tau}^{\mathrm{o}}_{xh}+h) - \widehat{\beta}(\widehat{\tau}^{\mathrm{o}}_{xh}-h)\big)}{2h} \leq \inf_{\tau\in\mathscr{T}} \frac{x'\big(\widehat{\beta}(\tau+h) - \widehat{\beta}(\tau-h)\big)}{2h} + o\left(\frac{1}{\sqrt[2/3]{nh^2}}\right),$$

and where h is a bandwidth that will be discussed, as well as the process η used for our estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$, in Section 3. For more details about the estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ in (13), see Ota et al. (2019).

Regarding the benefits of estimating the conditional mode using $\widehat{\mathrm{m}}_\eta(x)$, we emphasize that the smoothed estimator $\widehat{\beta}_h(\tau)$ is asymptotically unbiased and normally distributed, having lower variance and asymptotic mean squared error than the canonical estimator $\widehat{\beta}(\tau)$, uniformly with respect to $\tau \in \mathscr{T}$ and $h \in \big[h_{[n]}, h^{[n]}\big]$ — see Theorems 1, 3 and 5, and also Proposition 1 in Fernandes et al. (2021). Uniformity with respect to the smoothing parameter is particularly important as it allows random and possibly τ-dependent bandwidths to be used, like the η process introduced above. Furthermore, the map $\tau \mapsto x'\widehat{\beta}_h(\tau)$ is continuous with probability tending to one, while $\tau \mapsto x'\widehat{\beta}(\tau)$ is a step function — see Theorem 2 in Fernandes et al. (2021). The properties mentioned above give us clues that the estimator $\widehat{\mathrm{m}}_\eta(x)$ must dominate $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ with respect to the asymptotic mean squared error.[3]

3. Monte Carlo Study

In this section we describe the results of a Monte Carlo study conducted to compare the estimators introduced in Section 2. All analyses in this section, as well

[3] A positive side effect of the estimator $\widehat{\mathtt{f}}_h(\tau|x)$ is to avoid the curse of dimensionality, as it uses the linear structure of quantile regression. This aspect will not be explored in the Monte Carlo study described in Section 3.

as those performed in Section 4, were performed using the R software, version 4.0.3 (R Core Team, 2021). For the computation of $\widehat{\beta}_h(\tau)$ (using a Gaussian kernel), we made use of the package `conquer`, from He et al. (2020). The implementation of the algorithm in R to compute the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}$, was kindly provided by Hirofumi Ota.

3.1. Experiment Design

Inspired by the generating processes employed in Fernandes et al. (2021), we generate the data from the model $Y = X'\beta + Z$, where $X = (1 \quad \tilde{X})'$, with $\tilde{X} \sim \text{Uniform}[1,5]$ and $\beta = (1 \quad 1)'$. Note that for $\tau \in (0,1)$ and $x' = (1 \quad \tilde{x})$, where $\tilde{x} \in [1,5]$, the pair (X,Y) satisfies the linear quantile regression model $Q(\tau|x) = x'\beta(\tau)$, with $\beta(\tau) = \beta + \big(Q_Z(\tau|x) \quad 0\big)'$. Like the authors, we consider five different distributions for the error term Z, as described in items 1 to 5 below:

1. **Exponential error term.** With this specification, we want to evaluate the estimators in an "extreme" asymmetry scenario ($\text{skew}(Z) = 2$), where the conditional mode is a value at the boundary of the support of $f(\cdot|x)$. Here, the error term Z follows a Exponential$(1/\sqrt{2})$ distribution, independent of X.

2. **Gumbel error term.** The purpose of this specification is to evaluate the estimators under a "low" asymmetry scenario ($\text{skew}(Z) \approx 1.44$). Here, the error term Z follows a Gumbel$(0, \sqrt{12}/\pi)$ distribution, independent of X.

3. **χ^2 error term.** In this specification, we want to evaluate the estimators in an "intermediate" asymmetry scenario ($\text{skew}(Z) = 1.63$). Here, we have the error term $Z = W - 1$, where $W \sim \chi^2_3$ is independent of X.

4. **Student's t error term.** In this specification, we want to evaluate the estimators in a heavy-tailed scenario. Here, we have the error term $Z = W/\sqrt{1.5}$, where $W \sim \boldsymbol{t}(3)$ is independent of X.

5. **Heteroskedastic error term.** Finally, in this specification, we want to evaluate the estimators in a symmetric scenario, but with heteroskedasticity in the error term. Here, we have the heteroskedastic error term $Z = \frac{1}{4}(1 + \tilde{X})W$, where $W \sim \text{N}\,(0, {}^{24}/_{13})$ is independent of $\tilde{X}$.

In all these specifications, for $x' = (1 \quad \tilde{x})$ with $\tilde{x} \in [1,5]$, we clearly have $\mathrm{m}_Z(x) = 0$ and, consequently, in all simulated models, the identity $\mathrm{m}(x) = 1 + \tilde{x}$, is valid for x as above. Furthermore, except in the χ^2 case, we have $\text{Var}(Z) = 2$. For each of the five specifications described above, we sample $n \in \{100, 250, 500, 1000\}$ observations, resulting in a total of 20 different scenarios. In order to evaluate them, we generated $n_{\text{repl}} := 10,000$ replications of each one. We condition on $X = x$ with $x = (1 \quad 1.5)'$ and $x = (1 \quad 3)'$,[4] in each replication $j \in \{1, \ldots, n_{\text{repl}}\}$,

[4] In preliminary simulations, we also conditioned on $x = (1 \quad 4.5)'$. However, in all scenarios there was no practical distinction between this case and the case where $x = (1 \quad 1.5)'$, except for the heteroskedastic error term.

we compute the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}[j]}(x)$ of Ota et al. (2019) and our estimator $\widehat{\mathrm{m}}_\eta^{[j]}(x)$. In the simulations, we take $\mathscr{T} = \{0.01, 0.02, \ldots, 0.98, 0.99\}$.

In each scenario, for a generic estimator $m \in \{\widehat{\mathrm{m}}_\eta, \widehat{\mathrm{m}}_h^{\mathrm{o}}\}$ and x as above, we calculate the absolute bias $\mathrm{Bias}_{\mathrm{mc}}(m, x)$ expressed by

$$\mathrm{Bias}_{\mathrm{mc}}(m, x) = \left| \mathrm{m}(x) - \frac{1}{n_{\mathrm{repl}}} \sum_{j=1}^{n_{\mathrm{repl}}} m^{[j]}(x) \right|.$$

Similarly, we compute the mean squared error $\mathrm{MSE}_{\mathrm{mc}}(m, x)$ by

$$\mathrm{MSE}_{\mathrm{mc}}(m, x) = \frac{1}{n_{\mathrm{repl}}} \sum_{j=1}^{n_{\mathrm{repl}}} \big(\mathrm{m}(x) - m^{[j]}(x) \big)^2.$$

3.2. Bandwidth Selection

As mentioned in Section 2, to estimate $\widehat{\mathrm{m}}_\eta(x)$ in (12), it is necessary to select the process η. Following Fernandes et al. (2021), we use Silverman's bandwidth Silverman (1986), also called *Silverman's rule-of-thumb bandwidth*, which can be expressed, in the present context, by

$$\eta^*(\tau) = \frac{1.06}{\sqrt[5]{n}} \, \widehat{s}(\tau), \quad \tau \in \mathscr{T}, \tag{14}$$

where n is the sample size and, for each quantile level $\tau \in \mathscr{T}$, a different $\widehat{s}(\tau)$ is computed using the following algorithm:

1. The canonical estimator $\widehat{\beta}(\tau)$ is calculated, and the corresponding residuals $\widehat{Z}_i(\tau) := Y_i - X_i' \widehat{\beta}(\tau)$ are computed for $i \in \{1, \ldots, n\}$;
2. The sample interquartile range $\mathrm{iq}(\tau)$ is calculated, corresponding to the residuals $\widehat{Z}_1(\tau), \ldots, \widehat{Z}_n(\tau)$;
3. The sample standard deviation $\widehat{\sigma}(\tau)$, from $\widehat{Z}_1(\tau), \ldots, \widehat{Z}_n(\tau)$, is calculated;
4. Finally, $\widehat{s}(\tau)$ is given by

$$\widehat{s}(\tau) := \min \big\{ 0.7199528 \times \mathrm{iq}(\tau), \, \widehat{\sigma}(\tau) \big\}.$$

Note that the bandwidth η^* in (14) is τ-dependent, that is, we have a different smoothing parameter for each $\tau \in \mathscr{T}$. Furthermore, η^* is data-driven, that is, it can (and will) vary with each replication, as it depends on the generated sample. We emphasize that, by construction, the bandwidth η^* is indicated for cases where there is normality in the data — see Silverman (1986) — which does not occur in our simulations. Despite that, Fernandes et al. (2021) signal that this choice of smoothing parameter value can have satisfactory performance even when the Gaussianity assumption is violated. In order to explore the estimator performance for other smoothing parameters, we chose to evaluate

a series of bandwidths proportional to η^*: we consider the estimator $\widehat{\mathrm{m}}_\eta$ computed with $\eta \in \{\eta^*/3, \eta^*/2, \eta^*, 2\eta^*, 3\eta^*\}$. Exceptionally, for the scenario where $Z \sim \text{Exponential}(1/\sqrt{2})$, $\tilde{x} = 3$ and $n = 1000$, we noticed (in preliminary simulations) that smaller η's produced estimates with smaller MSE_{mc} and Bias_{mc}. Because of that, in this case, we also consider $\eta = \eta^*/8$ and $\eta = \eta^*/5$.

As with $\widehat{\mathrm{m}}_\eta(x)$, the estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ of Ota et al. (2019) depends on the definition of a bandwidth h. In the Monte Carlo study and the application to real data, the authors select it according to the following procedure: first, one lets

$$h^{\mathrm{km}}(\tau) = n^{-1/3} z_\alpha^{2/3} \times \left(1.5 \frac{\phi(\Phi^{-1}(\tau))}{2\Phi^{-1}(\tau)^2 + 1}\right)^{1/3},$$

with ϕ and Φ being, respectively, the probability density function and the cumulative distribution function corresponding to the distribution $\mathrm{N}(0,1)$, and where $z_\alpha = \Phi^{-1}(1 - \alpha/2)$. Then select the bandwidth h^* through steps 1 to 4 described below:

1. User determines the level $\alpha \in (0,1)$. In the present case, $\alpha = 0.05$;
2. Choose x in the support of X;
3. Use the bandwidth $h^{\mathrm{pilot}} = n^{1/6}\, h^{\mathrm{km}}(0.5) \propto n^{-1/6}$ as a pilot to build the preliminary estimator of τ_x, say $\widehat{\tau}_x^{\mathrm{pre}}$; specifically $\widehat{\tau}_x^{\mathrm{pre}} = \arg\min_{\tau \in \mathscr{T}} (2h)^{-1} x' \big(\widehat{\beta}(\tau + h) - \widehat{\beta}(\tau - h)\big)$ with $h = h^{\mathrm{pilot}}$;
4. Finally, $h^* \equiv h_n^*(x) := n^{1/6}\, h^{\mathrm{km}}(\widehat{\tau}_x^{\mathrm{pre}})$ (note that h^* is independent of τ).

Ota et al. (2019) argue that, despite not being optimal, this bandwidth selection worked well in their simulations. In our Monte Carlo study, seeking to follow an isonomy criterion, we compute the estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_h$ with $h \in \{h^*/2, h^*, 2h^*\}$. In each case, we present the results for $\widehat{\mathrm{m}}_{h*}$ and for the estimator with the best performance between $\widehat{\mathrm{m}}_{2h^*}$ and $\widehat{\mathrm{m}}_{h^*/2}$, (except when $Z \sim \boldsymbol{t}$, for which the performance of $\widehat{\mathrm{m}}_{h*}$ was notably superior to that of $\widehat{\mathrm{m}}_{2h^*}$ and $\widehat{\mathrm{m}}_{h^*/2}$, resulting in both suppression).

3.3. Description of the Monte Carlo Study Results

3.3.1. Exponential Error Term

In Table 1 we see that, in the extreme asymmetry scenario implied by an exponential error term, our estimator $\widehat{\mathrm{m}}_\eta(x)$ was systematically surpassed by the estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$. However, the good news is that as the sample size increases, we see a considerable reduction in Bias_{mc} and MSE_{mc}. In Figures 1 and 2, it is important to note that as n grew, $\widehat{\mathrm{m}}_\eta(x)$ produced better results with smaller η's, that is, a lower smoothing of $\widehat{\beta}_h(\tau)$ was more appropriate in this scenario. Especially when $n = 1000$ and $\tilde{x} = 3$, the idea of using $\widehat{\mathrm{m}}_{\eta^*/8}$ was rewarding (resulting in the smaller Bias_{mc} and MSE_{mc} among all $\widehat{\mathrm{m}}_\eta(x)$ estimators).

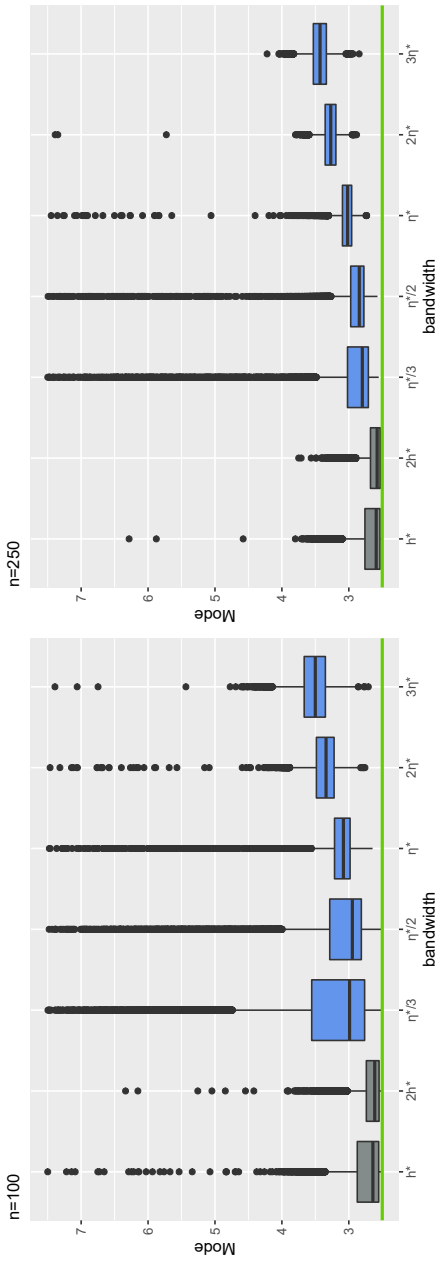

FIGURE 1: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^o(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim$ Exponential.

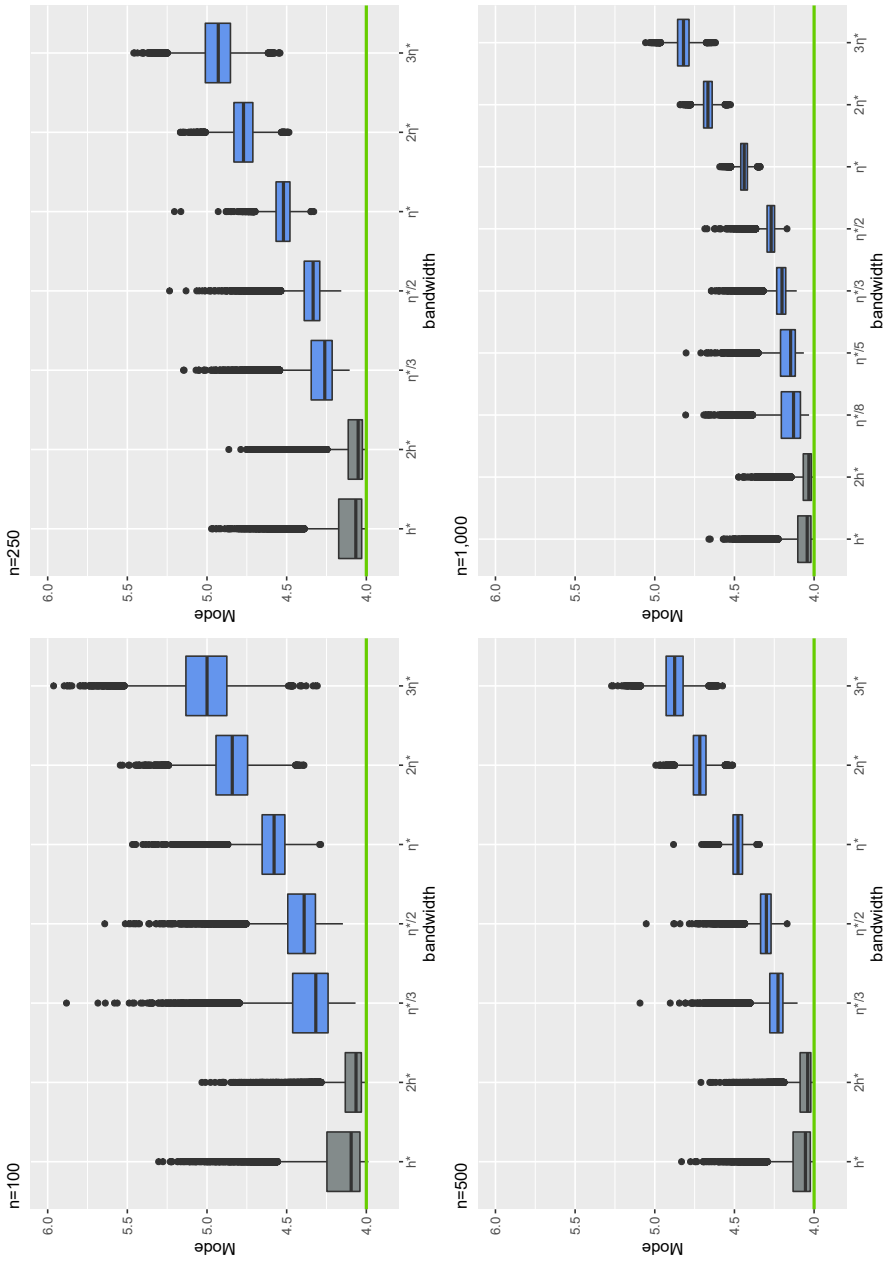

FIGURE 2: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^\circ(x)$ for $x = (1 \quad 3)'$ with $Z \sim$ Exponential.

TABLE 1: Bias and MSE of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ for $x = (1 \quad 1.5)'$ and $x = (1 \quad 3)'$ with $Z \sim$ Exponential.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
		$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^*}$	0.261	0.209	0.180	0.083	0.142	0.048	0.112	0.030
	$\widehat{\mathrm{m}}_{2h^*}$	0.170	0.072	0.129	0.038	0.103	0.025	0.080	0.015
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.890	1.786	0.543	0.923	0.364	0.433	0.268	0.183
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.767	1.342	0.483	0.585	0.366	0.270	0.300	0.134
	$\widehat{\mathrm{m}}_{\eta^*}$	0.690	0.760	0.559	0.441	0.495	0.296	0.443	0.202
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.871	0.845	0.780	0.643	0.721	0.527	0.669	0.458
	$\widehat{\mathrm{m}}_{3\eta^*}$	1.021	1.111	0.937	0.900	0.878	0.781	0.823	0.682
$\tilde{x} = 3$	$\widehat{\mathrm{m}}_{h^*}$	0.178	0.072	0.130	0.040	0.102	0.023	0.081	0.015
	$\widehat{\mathrm{m}}_{2h^*}$	0.106	0.036	0.088	0.017	0.069	0.010	0.054	0.006
	$\widehat{\mathrm{m}}_{\eta^*/8}$	–	–	–	–	–	–	0.159	0.035
	$\widehat{\mathrm{m}}_{\eta^*/5}$	–	–	–	–	–	–	0.177	0.038
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.380	0.185	0.302	0.109	0.253	0.072	0.218	0.051
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.433	0.217	0.357	0.138	0.311	0.101	0.276	0.078
	$\widehat{\mathrm{m}}_{\eta^*}$	0.592	0.364	0.526	0.281	0.481	0.233	0.440	0.194
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.851	0.746	0.774	0.607	0.719	0.521	0.667	0.447
	$\widehat{\mathrm{m}}_{3\eta^*}$	1.013	1.067	0.935	0.889	0.878	0.777	0.822	0.678

3.3.2. Gumbel Error Term

In Table 2 we see that, for the Gumbel error term, the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ presented less bias compared to $\widehat{\mathrm{m}}_\eta(x)$, for all the h and η levels explored, except in the scenario where $\tilde{x} = 3$ and $n = 1000$. However, with the proper bandwidth, the estimator $\widehat{\mathrm{m}}_\eta(x)$ performed better than $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$, in terms of $\mathrm{MSE}_{\mathrm{mc}}$, in all cases $\big(\tilde{x} = \{1.5, 3\}; n = \{100, 250, 500, 1000\}\big)$ explored. The superiority of $\widehat{\mathrm{m}}_\eta(x)$ in terms of $\mathrm{MSE}_{\mathrm{mc}}$ was even more evident in small samples. We highlight the scenario where $n = 100$. In this case, the mean squared error of $\widehat{\mathrm{m}}_{2\eta^*}$ was more than twice smaller than that of $\widehat{\mathrm{m}}_{h^*}$ when $\tilde{x} = 3$ and more than four times smaller when $\tilde{x} = 1.5$. Also, with $n = 1000$, $\tilde{x} = 3$, our estimator $\widehat{\mathrm{m}}_{\eta^*}$ outperformed $\widehat{\mathrm{m}}_{h^*}$ in both bias and mean squared error. In Figures 3 and 4, we see that, for $n = 1000$, the worst estimates produced by $\widehat{\mathrm{m}}_\eta(x)$, for the most part, overestimated $\mathrm{m}(x)$, while those produced by $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ underestimated $\mathrm{m}(x)$. We also had, for $\widehat{\mathrm{m}}_{2\eta^*}$ and $\widehat{\mathrm{m}}_{3\eta^*}$, a small dispersion compared to the others. However, at the cost of visible bias (clearly $\mathrm{m}(x)$ is overestimated).

3.3.3. Chi-Squared Error Term

In Table 3 we see that, for the χ^2 error term, the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ presented less bias than $\widehat{\mathrm{m}}_\eta(x)$ in all cases. However, our estimator $\widehat{\mathrm{m}}_\eta(x)$ presented a lower MSE than $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ with $\eta = \eta^*$; $\tilde{x} = 1.5$; $n = 1000$ and with $\eta = \eta^*/2$; $\tilde{x} = 3$; $n = \{250, 500, 1000\}$. This reinforces our idea presented in section 2, that the estimator $\widehat{\mathrm{m}}_\eta(x)$ must dominate $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ as to the asymptotic mean squared error. In Figures 5 and 6, the advantage of $\widehat{\mathrm{m}}_\eta(x)$ in relation to the $\mathrm{MSE}_{\mathrm{mc}}$ and of $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ in relation to $\mathrm{Bias}_{\mathrm{mc}}$ is clear. Looking specifically at the cases where $n = 1000$, it can be seen that, while the estimates by $\widehat{\mathrm{m}}_{h^*}$ were approximately symmetrical around the theoretical value $1+\tilde{x}$, the estimates by $\widehat{\mathrm{m}}_\eta(x)$ clearly showed a positive bias, but varied less.

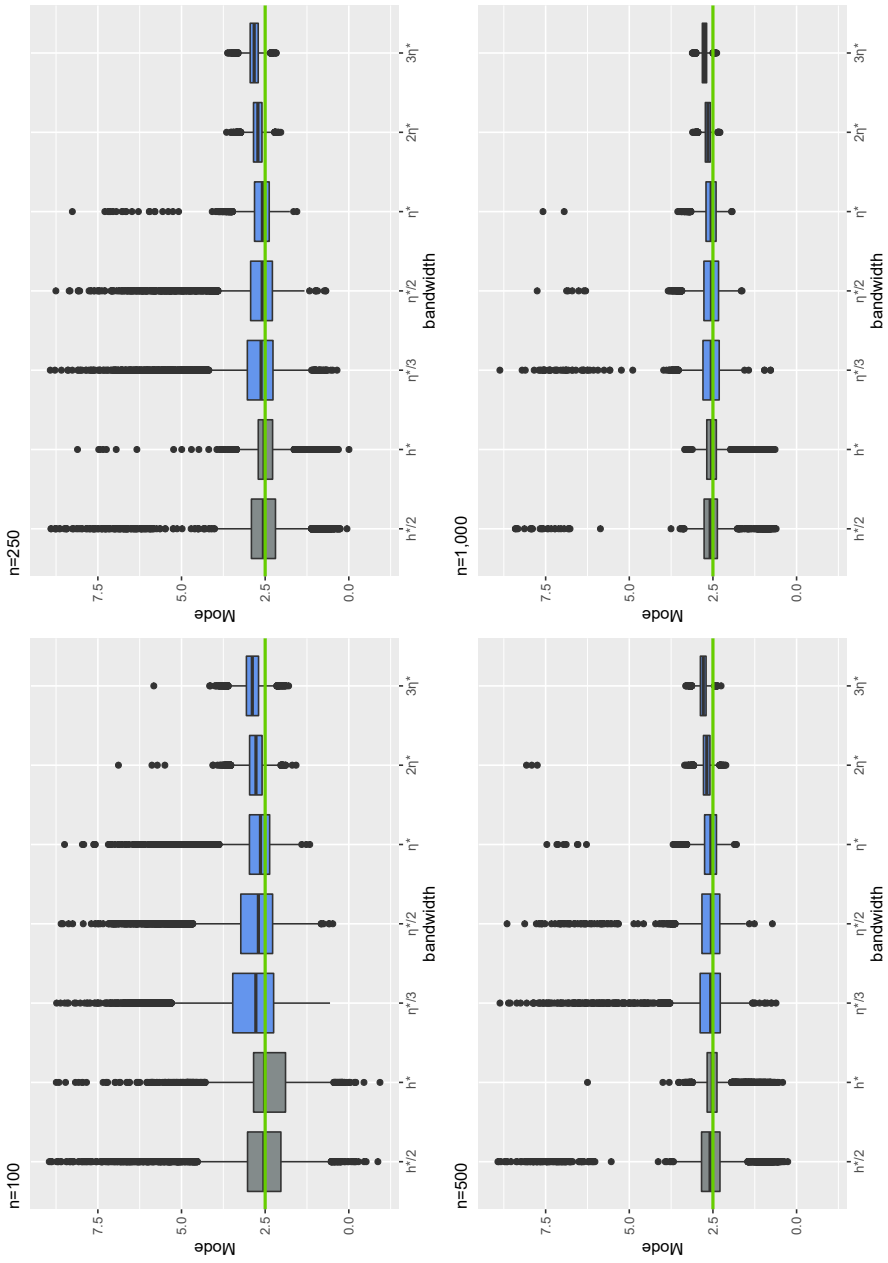

FIGURE 3: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^\circ(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim$ Gumbel

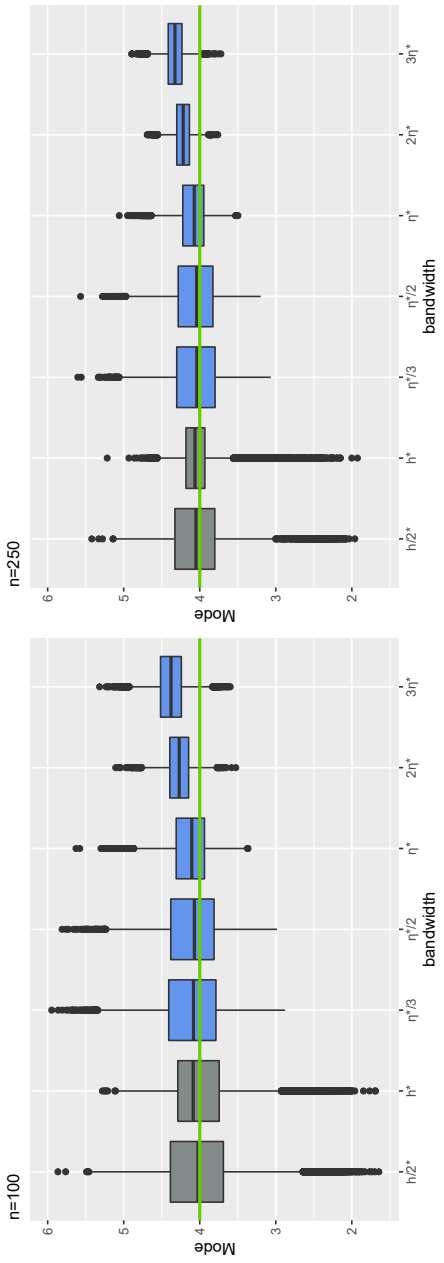

FIGURE 4: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^o(x)$ for $x = (1 \quad 3)'$ with $Z \sim$ Gumbel

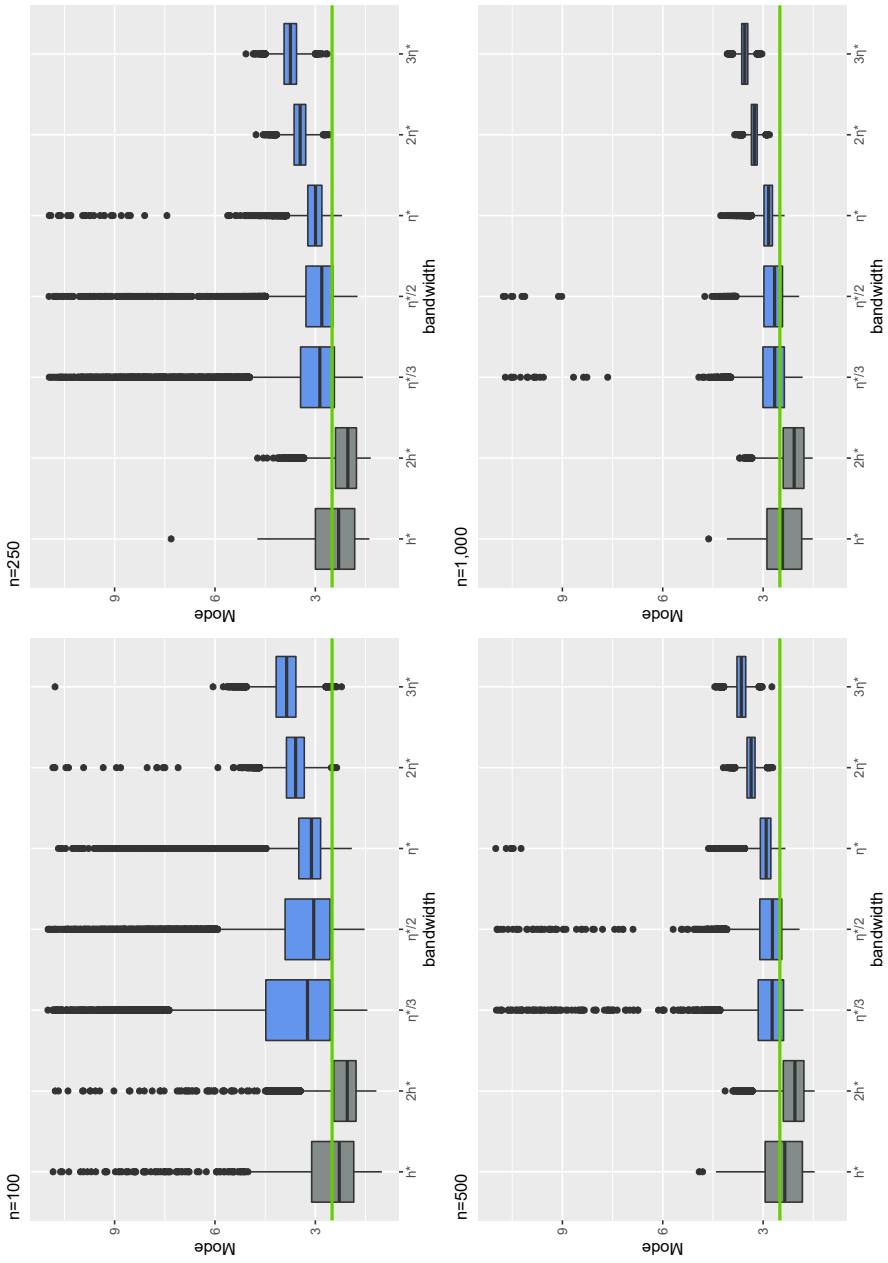

FIGURE 5: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^\circ(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim \boldsymbol{\chi^2_3}$

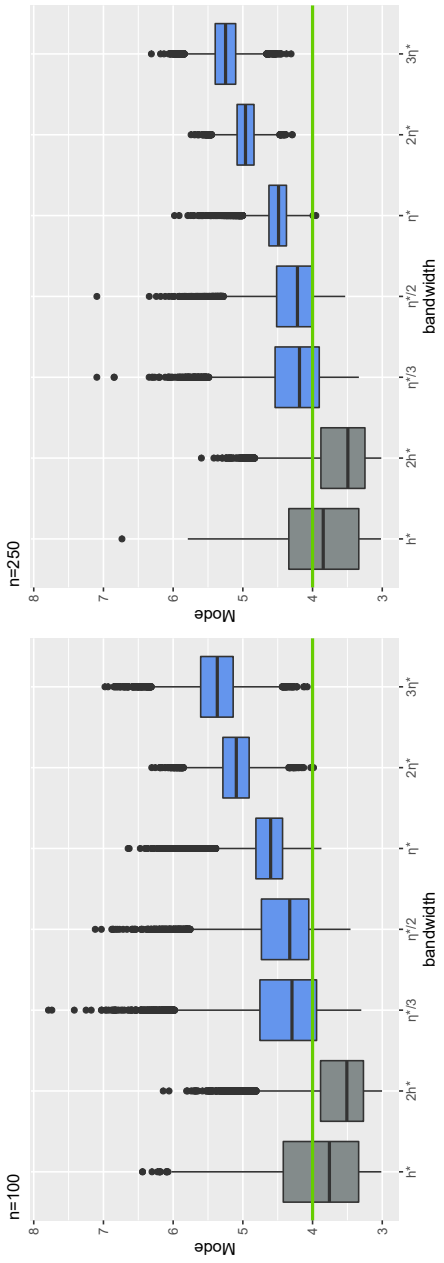

FIGURE 6: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^o(x)$ for $x = (1 \quad 3)'$ with $Z \sim \boldsymbol{\chi}^{\mathbf{2}}_{\mathbf{3}}$

TABLE 2: Bias and MSE of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ for $x = (1 \quad 1.5)'$ and $x = (1 \quad 3)'$ with $Z \sim$ Gumbel.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
		$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^{*/2}}$	0.089	1.210	0.030	0.716	0.075	0.524	0.063	0.176
	$\widehat{\mathrm{m}}_{h^*}$	0.096	0.676	0.086	0.280	0.020	0.130	0.034	0.053
	$\widehat{\mathrm{m}}_{\eta^{*/3}}$	0.478	1.379	0.262	0.795	0.142	0.420	0.083	0.225
	$\widehat{\mathrm{m}}_{\eta^{*/2}}$	0.379	1.010	0.175	0.474	0.096	0.256	0.063	0.118
	$\widehat{\mathrm{m}}_{\eta^*}$	0.228	0.398	0.122	0.155	0.087	0.093	0.068	0.058
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.286	0.165	0.224	0.086	0.190	0.065	0.156	0.036
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.385	0.228	0.330	0.142	0.291	0.101	0.253	0.073
$\tilde{x} = 3$	$\widehat{\mathrm{m}}_{h^{*/2}}$	0.012	0.596	0.056	0.394	0.071	0.265	0.049	0.049
	$\widehat{\mathrm{m}}_{h^*}$	0.052	0.269	0.016	0.096	0.041	0.037	0.062	0.024
	$\widehat{\mathrm{m}}_{\eta^{*/3}}$	0.110	0.211	0.062	0.132	0.047	0.097	0.038	0.072
	$\widehat{\mathrm{m}}_{\eta^{*/2}}$	0.109	0.179	0.067	0.109	0.049	0.080	0.039	0.058
	$\widehat{\mathrm{m}}_{\eta^*}$	0.137	0.098	0.094	0.053	0.071	0.034	0.055	0.023
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.274	0.110	0.221	0.064	0.187	0.044	0.153	0.028
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.380	0.189	0.326	0.125	0.290	0.094	0.252	0.069

TABLE 3: Bias and MSE of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ for $x = (1 \quad 1.5)'$ and $x = (1 \quad 3)'$ with $Z \sim \boldsymbol{\chi^2_3}$.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
		$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^*}$	0.061	1.020	0.058	0.486	0.092	0.385	0.006	0.319
	$\widehat{\mathrm{m}}_{2h^*}$	0.283	0.680	0.354	0.352	0.362	0.323	0.369	0.304
	$\widehat{\mathrm{m}}_{\eta^{*/3}}$	1.356	5.352	0.756	2.799	0.398	1.097	0.251	0.553
	$\widehat{\mathrm{m}}_{\eta^{*/2}}$	1.072	3.820	0.556	1.515	0.369	0.794	0.245	0.368
	$\widehat{\mathrm{m}}_{\eta^*}$	0.831	1.605	0.596	0.823	0.465	0.362	0.369	0.200
	$\widehat{\mathrm{m}}_{2\eta^*}$	1.120	1.467	0.971	1.052	0.867	0.805	0.766	0.603
	$\widehat{\mathrm{m}}_{3\eta^*}$	1.385	2.124	1.252	1.647	1.153	1.369	1.050	1.121
$\tilde{x} = 3$	$\widehat{\mathrm{m}}_{h^*}$	0.074	0.467	0.110	0.362	0.108	0.300	0.082	0.236
	$\widehat{\mathrm{m}}_{2h^*}$	0.348	0.374	0.352	0.353	0.378	0.333	0.358	0.295
	$\widehat{\mathrm{m}}_{\eta^{*/3}}$	0.405	0.532	0.260	0.281	0.192	0.189	0.140	0.132
	$\widehat{\mathrm{m}}_{\eta^{*/2}}$	0.444	0.473	0.297	0.243	0.221	0.155	0.165	0.106
	$\widehat{\mathrm{m}}_{\eta^*}$	0.653	0.535	0.514	0.308	0.425	0.204	0.346	0.134
	$\widehat{\mathrm{m}}_{2\eta^*}$	1.104	1.302	0.965	0.963	0.860	0.755	0.763	0.589
	$\widehat{\mathrm{m}}_{3\eta^*}$	1.383	2.045	1.252	1.620	1.150	1.348	1.049	1.112

3.3.4. Student's *t* error term

In Table 4 we see that, for the Student's $\boldsymbol{t}$ error term, the $\mathrm{Bias}_{\mathrm{mc}}$ of both estimators was negligible in all evaluated scenarios, as expected because it is a symmetric distribution. Therefore, we will only analyze the $\mathrm{MSE}_{\mathrm{mc}}$. It is noteworthy that $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ presented much higher $\mathrm{MSE}_{\mathrm{mc}}$ than $\widehat{\mathrm{m}}_\eta(x)$, when $\tilde{x} = 1.5$ and $n = 100$. We also see that in practically all scenarios, the best performance occurred with $\widehat{\mathrm{m}}_{3\eta^*}$, the maximum smoothing explored in this study. The only scenario where $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ performed similarly to $\widehat{\mathrm{m}}_\eta(x)$ was when $n = 1000$. Figures 7 and 8, in turn, allow us to see that in fact, using $\widehat{\mathrm{m}}_{3\eta^*}$, the estimated modes are more concentrated around of the theoretical value $1 + \tilde{x}$ compared to those estimated by $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$.

TABLE 4: Bias and MSE of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ for $x = (1 \quad 1.5)'$ and $x = (1 \quad 3)'$ with $Z \sim \boldsymbol{t}$.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
		Bias_{mc}	MSE_{mc}	Bias_{mc}	MSE_{mc}	Bias_{mc}	MSE_{mc}	Bias_{mc}	MSE_{mc}
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^*}$	0.005	0.454	0.003	0.073	0.002	0.028	0.000	0.006
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.004	0.711	0.008	0.466	0.000	0.253	0.000	0.141
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.001	0.543	0.003	0.301	0.002	0.166	0.003	0.084
	$\widehat{\mathrm{m}}_{\eta^*}$	0.002	0.275	0.001	0.138	0.001	0.063	0.002	0.035
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.001	0.083	0.001	0.051	0.000	0.019	0.000	0.006
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.003	0.057	0.000	0.023	0.001	0.009	0.000	0.005
$\tilde{x} = 3$	$\widehat{\mathrm{m}}_{h^*}$	0.002	0.054	0.000	0.031	0.000	0.015	0.000	0.002
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.004	0.087	0.005	0.056	0.002	0.041	0.002	0.031
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.001	0.074	0.004	0.046	0.002	0.033	0.001	0.024
	$\widehat{\mathrm{m}}_{\eta^*}$	0.000	0.035	0.003	0.019	0.002	0.012	0.001	0.008
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.000	0.020	0.001	0.009	0.001	0.005	0.001	0.003
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.001	0.026	0.001	0.010	0.001	0.005	0.000	0.002

3.3.5. Heteroskedastic Error Term

In Table 5 we see that, for the heteroskedastic error term, as in the previous case, the Bias_{mc} of both estimators was very small in all evaluated scenarios. So again we will only evaluate the MSE_{mc}. Also as in the case of the $\boldsymbol{t}$ error, in practically all scenarios, the best results were obtained using $\widehat{\mathrm{m}}_{3\eta^*}$. As expected due to the composition of the error term (with heteroskedasticity and proportional to $\tilde{x}$), the MSE_{mc} was higher when we condition on $x = (1 \quad 4.5)'$, compared to $x = (1 \quad 1.5)'$ or $x = (1 \quad 3)'$. Especially for small n, $\widehat{\mathrm{m}}_{3\eta^*}$ was vastly superior to $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$. Figures 9, 10 and 11 visually reinforce what we saw above. Notably in the cases where the bandwidths $2\eta^*$ and $3\eta^*$ were used, our estimator performed very well.

TABLE 5: Bias and MSE of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ for $x = (1 \quad 1.5)'$, $x = (1 \quad 3)'$ and $x = (1 \quad 4.5)'$ with Z heteroskedastic.

		$n = 100$		$n = 250$		$n = 500$		$n = 1000$	
		Bias_{mc}	MSE_{mc}	Bias_{mc}	MSE_{mc}	Bias_{mc}	MSE_{mc}	Bias_{mc}	MSE_{mc}
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^*/2}$	0.002	0.523	0.003	0.193	0.004	0.067	0.002	0.033
	$\widehat{\mathrm{m}}_{h^*}$	0.002	0.461	0.003	0.163	0.004	0.054	0.001	0.014
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.003	0.536	0.005	0.377	0.006	0.198	0.002	0.099
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.008	0.431	0.000	0.221	0.004	0.116	0.002	0.068
	$\widehat{\mathrm{m}}_{\eta^*}$	0.004	0.163	0.003	0.073	0.002	0.048	0.001	0.033
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.003	0.056	0.001	0.026	0.000	0.015	0.000	0.009
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.005	0.065	0.002	0.029	0.001	0.016	0.000	0.008
$\tilde{x} = 3$	$\widehat{\mathrm{m}}_{h^*/2}$	0.003	0.431	0.006	0.115	0.000	0.061	0.001	0.039
	$\widehat{\mathrm{m}}_{h^*}$	0.005	0.354	0.003	0.115	0.001	0.056	0.000	0.011
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.006	0.292	0.005	0.194	0.006	0.144	0.004	0.109
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.004	0.258	0.004	0.172	0.002	0.126	0.005	0.095
	$\widehat{\mathrm{m}}_{\eta^*}$	0.000	0.171	0.003	0.109	0.001	0.078	0.003	0.059
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.003	0.069	0.003	0.035	0.001	0.023	0.002	0.015
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.002	0.066	0.002	0.031	0.000	0.018	0.002	0.010
$\tilde{x} = 4.5$	$\widehat{\mathrm{m}}_{h^*/2}$	0.008	2.053	0.012	0.612	0.002	0.222	0.004	0.121
	$\widehat{\mathrm{m}}_{h^*}$	0.009	1.594	0.065	0.495	0.005	0.181	0.004	0.045
	$\widehat{\mathrm{m}}_{\eta^*/3}$	0.002	1.450	0.005	0.753	0.001	0.459	0.007	0.321
	$\widehat{\mathrm{m}}_{\eta^*/2}$	0.001	1.032	0.007	0.570	0.001	0.391	0.006	0.283
	$\widehat{\mathrm{m}}_{\eta^*}$	0.006	0.619	0.005	0.377	0.002	0.270	0.006	0.201
	$\widehat{\mathrm{m}}_{2\eta^*}$	0.003	0.269	0.003	0.141	0.002	0.096	0.003	0.068
	$\widehat{\mathrm{m}}_{3\eta^*}$	0.000	0.177	0.003	0.083	0.002	0.051	0.002	0.031

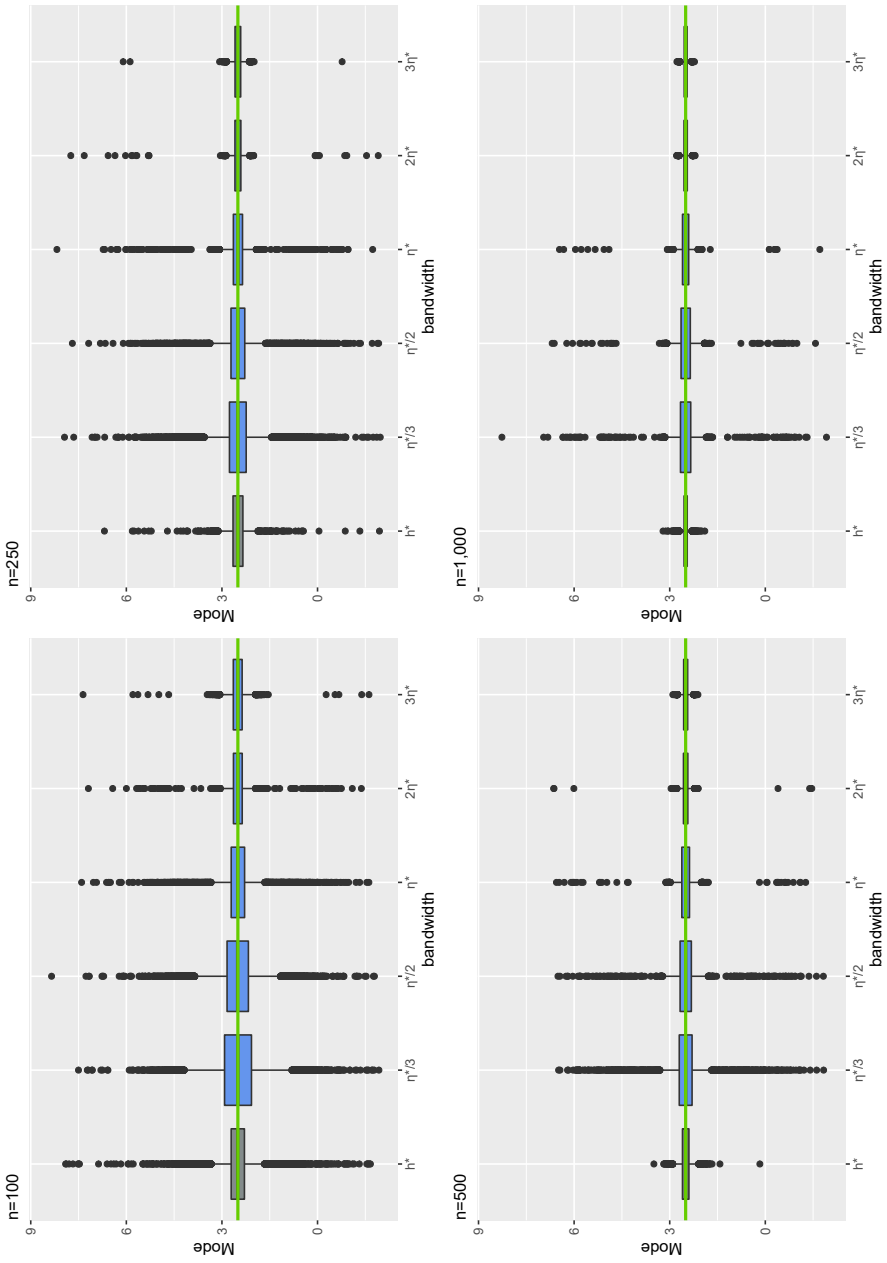

FIGURE 7: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim \boldsymbol{t}$.

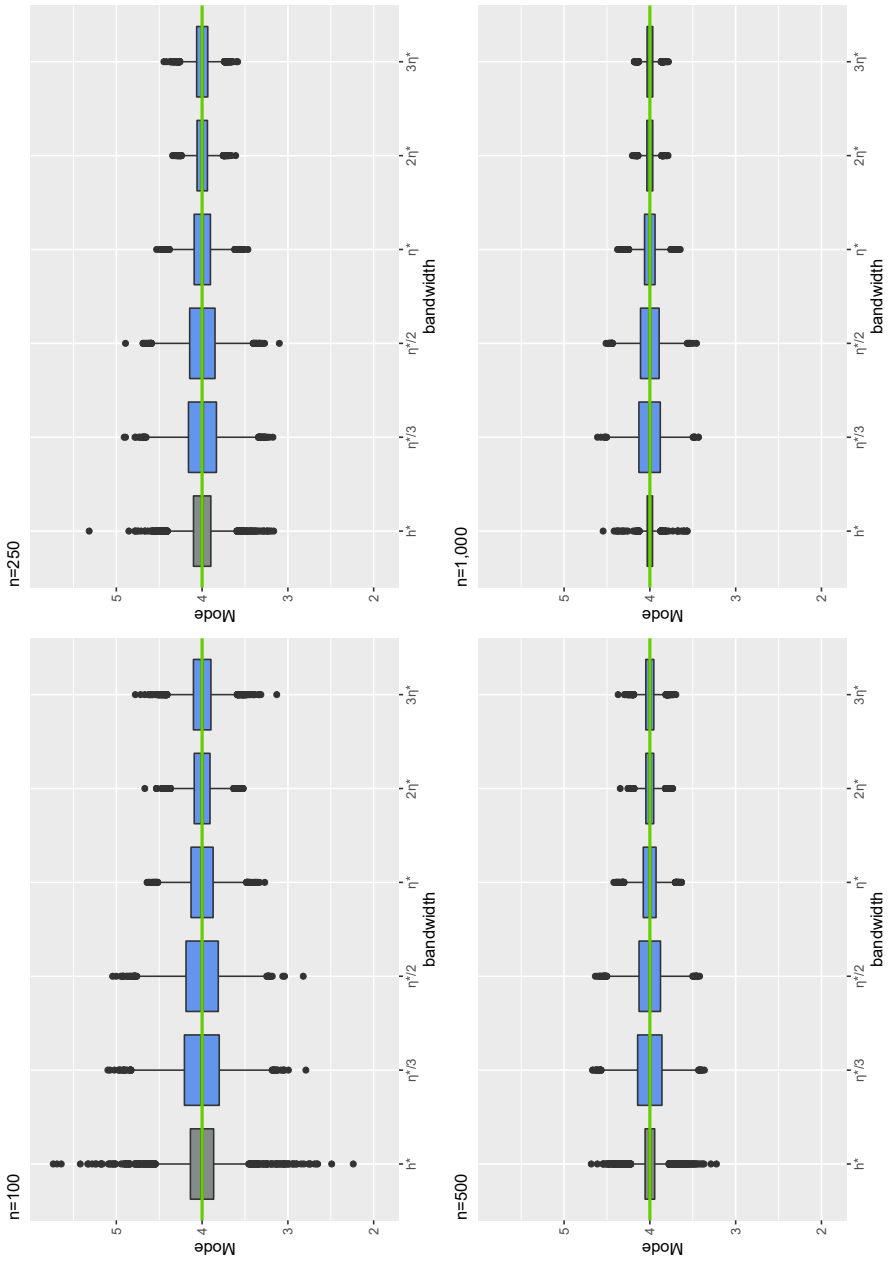

FIGURE 8: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^\circ(x)$ for $x = (1 \quad 3)'$ with $Z \sim \boldsymbol{t}$.

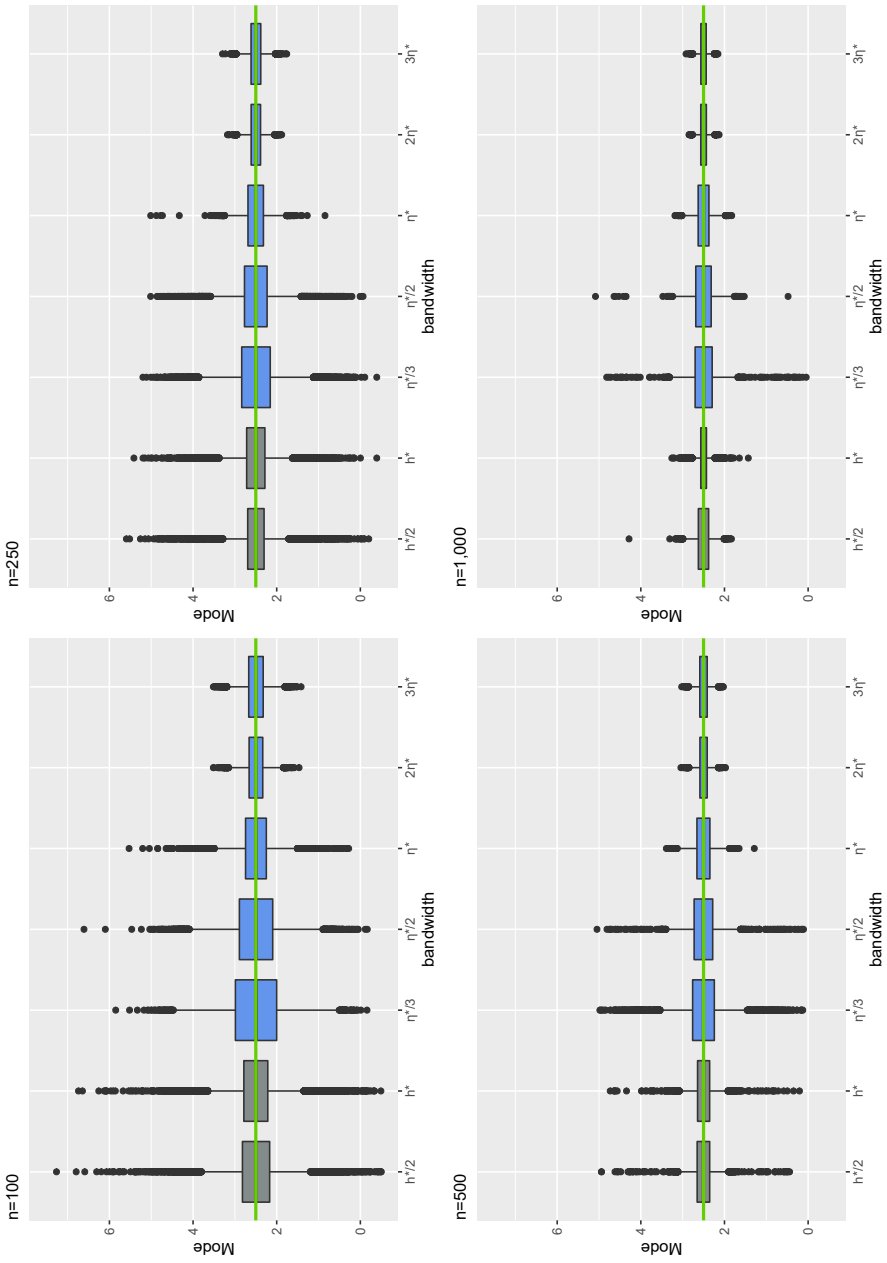

FIGURE 9: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^o(x)$ for $x = (1 \quad 1.5)'$ with Z heteroskedastic.

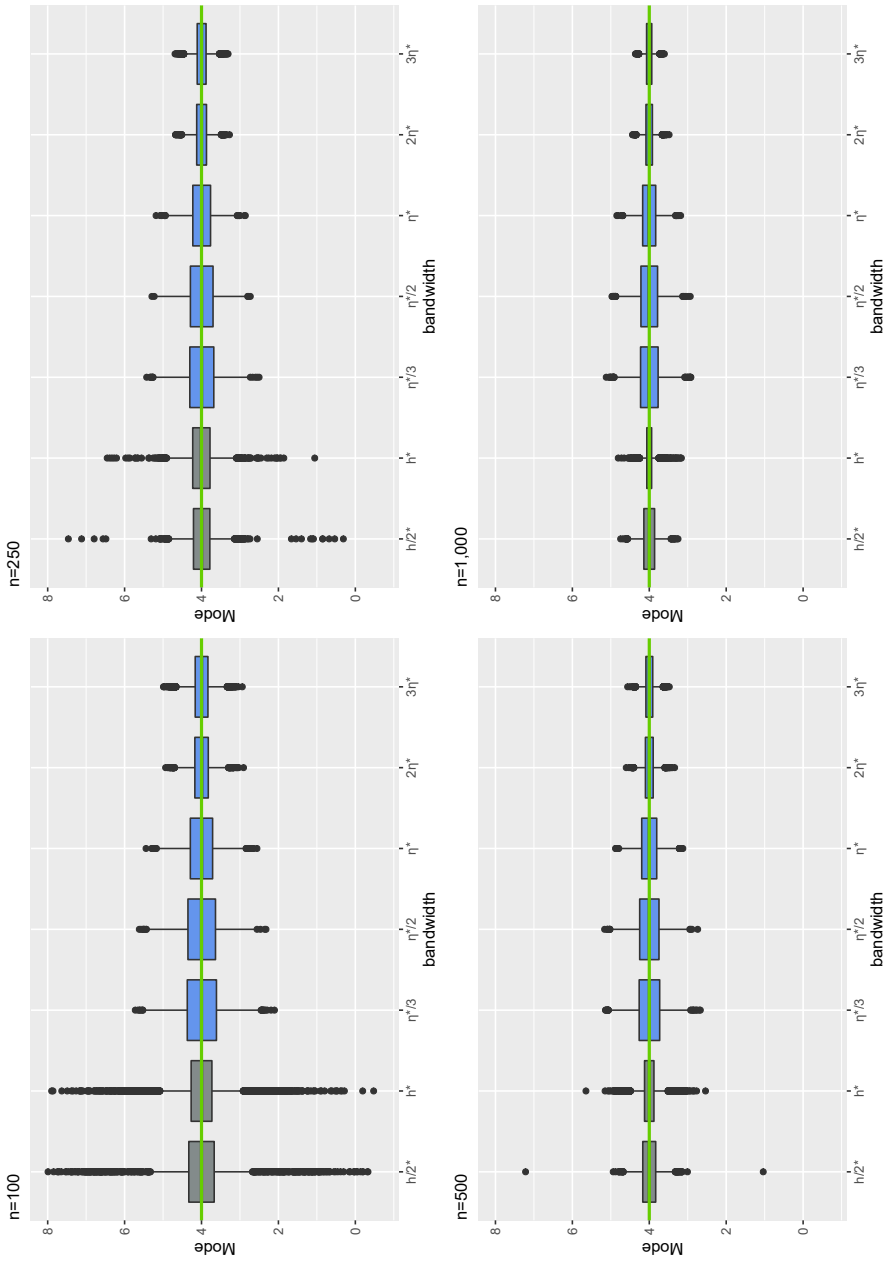

FIGURE 10: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\circ}(x)$ for $x = (1 \quad 3)'$ with Z heteroskedastic.

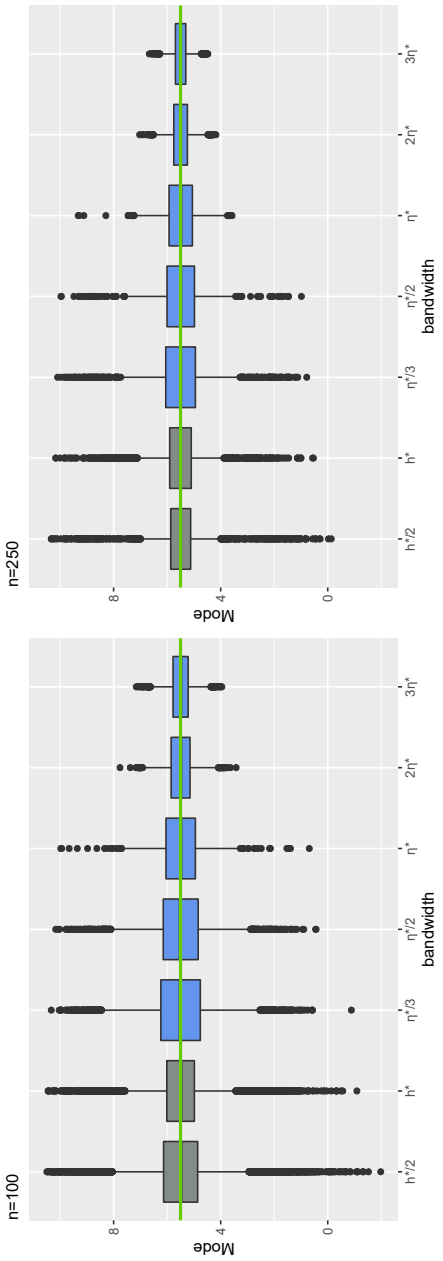

FIGURE 11: Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^\circ(x)$ for $x = (1 \quad 4.5)'$ with Z heteroskedastic.

3.3.6. Relative Mean Squared Errors: A Summary

Table 6 summarizes the performance difference between the estimators $\widehat{\mathrm{m}}_\eta$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}$ in terms of their relative mean squared errors (RMSE). It can be seen that, in the scenarios with Exponential, Gumbel, $\boldsymbol{t}$ and heteroskedastic error term, there was systematic overperformance of one estimator in relation to the other ($\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ dominated in the first and $\widehat{\mathrm{m}}_{\eta(x)}$ in the last three). Only in the case where the error term follows a χ^2, the best performance varied according to the scenario, notably with $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ better with $n = 100$ and $\widehat{\mathrm{m}}_\eta(x)$ better with $n = 1000$. It should be noted that we use all available decimal places to avoid a tie in the case where $Z \sim \boldsymbol{t}$, $\tilde{x} = 1.5$ and $n = 1000$.

TABLE 6: RMSE $\widehat{\mathrm{m}}_\eta \,/\, \widehat{\mathrm{m}}_h^{\mathrm{o}}$ with η and h corresponding to the smallest $\mathrm{MSE}_{\mathrm{mc}}$ in each scenario.

		Exp	Gumb	χ^2	$\boldsymbol{t}$	Het
$\tilde{x} = 1.5$	$n = 100$	10.556	0.244	2.360	0.126	0.141
	$n = 250$	11.605	0.307	2.338	0.315	0.160
	$n = 500$	10.800	0.500	1.121	0.321	0.278
	$n = 1000$	8.933	0.679	0.658	0.833	0.571
$\tilde{x} = 3$	$n = 100$	5.139	0.364	1.265	0.370	0.186
	$n = 250$	6.412	0.552	0.688	0.290	0.270
	$n = 500$	7.200	0.919	0.517	0.333	0.321
	$n = 1000$	5.833	0.958	0.449	0.976	0.909
$\tilde{x} = 4.5$	$n = 100$	–	–	–	–	0.111
	$n = 250$	–	–	–	–	0.168
	$n = 500$	–	–	–	–	0.282
	$n = 1000$	–	–	–	–	0.689

4. Application

In this section, we describe the results of applying our proposed estimator $\widehat{\mathrm{m}}_\eta$ and its comparison the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}$, proposed by Ota et al. (2019), using the same dataset analyzed in their study, which contains hourly measurements of a Combined Cycle Power Plant, available in the UCI Machine Learning Repository `https://archive.ics.uci.edu/ml/datasets/combined+cycle+power+plant`. For more details on the dataset, see Tufekci (2014) and Kaya & Tufekci (2012).

The choice of this dataset is motivated by two main reasons. First, it allows a direct and fair comparison between our estimator and that of Ota et al. (2019), under identical conditions. Second, this dataset provides an ideal empirical environment to illustrate the strengths of our approach: the response variable exhibits noticeable asymmetry and heteroskedasticity, features that challenge traditional estimators but are naturally handled by the smoothed quantile regression framework. Moreover, this application has practical relevance, since accurately modeling energy output is essential for assessing plant efficiency and operational performance.

The dataset has $n = 9{,}568$ observations, collected between 2006 and 2011. There is a dependent variable $Y :=$ 'Net Hourly Electrical Energy Output (EP)', measured in Megawatt (MW), varying in the sample between 420.26 and 495.76. Among the independent variables, we worked with $\tilde{X} :=$ 'Exhaust Vacuum (V)', measured in centimeters of mercury (cmHg), ranging in the sample between 25.36 and 81.56. In what follows, $X := (1 \quad \tilde{X})'$.

Initially, we estimated the conditional densities for 15 values of $x = (1 \quad \tilde{x})'$, with the values of $\tilde{x}$ equally spaced between the minimum and maximum of $\tilde{X}$ in the sample, i.e. $\tilde{x} \in \{25.36, \ldots, 81.56\}$, using the estimator $\widehat{\mathtt{f}}_{\eta(\tau)}(\tau|x)$, $\tau \in \mathscr{T}$. Having the estimated densities, we estimate the conditional mode using the estimator $\widehat{\mathrm{m}}_{\eta^*}$. The conditional density estimates can be seen in Figure 12. A similar figure can be seen in Figure 3 in Ota et al. (2019). In Table 7, we see the comparison between the estimates for the conditional mode using $\widehat{\mathrm{m}}_{\eta^*}$ and $\widehat{\mathrm{m}}_{h^*}$.

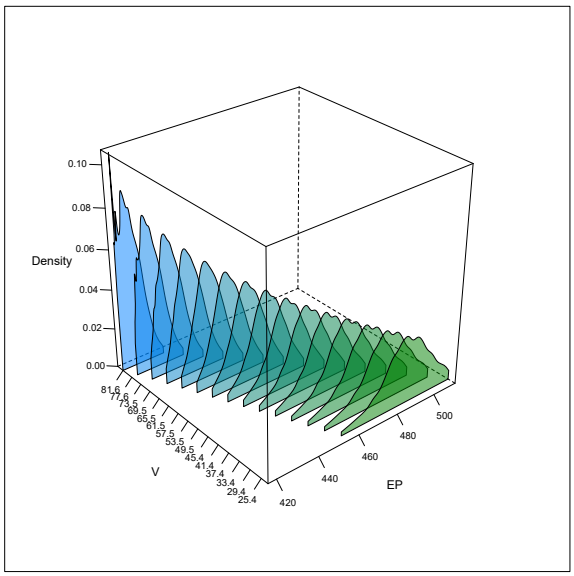

FIGURE 12: Estimated conditional densities of net hourly electrical energy output given exhaust vacuum.

Figure 12 shows that there is asymmetry in the estimated conditional densities, notably for higher values of $\tilde{x}$. In Table 7, we see that, for the 15 values to which $\tilde{x}$ was conditioned, it occurred that $\widehat{\mathrm{m}}_{\eta^*}(\tilde{x}) < \widehat{\mathrm{m}}_{h^*}(\tilde{x})$.

In addition to estimating the conditional densities, Ota et al. (2019) created prediction intervals (PI) with theoretical coverage of 95%, using their estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_{h}(x)$, with $h = h^*$ via the split conformal prediction procedure of Lei et al. (2018). Following the authors' idea, we build prediction intervals using our estimator $\widehat{\mathrm{m}}_{\eta}(x)$, with $\eta = \{\eta^*/2, \eta^*, 2\eta^*\}$, using the same procedure, described in the following steps:

TABLE 7: Comparison of conditional modes estimated by $\widehat{\mathrm{m}}_{\eta^*}$ and $\widehat{\mathrm{m}}_{h^*}$.

$\tilde{X} = \tilde{x}$	$\widehat{\mathrm{m}}_{\eta^*}$	$\widehat{\mathrm{m}}_{h^*}$	$(\widehat{\mathrm{m}}_{\eta^*} - \widehat{\mathrm{m}}_{h^*})$
25.4	486.98	487.61	−0.64
29.4	482.36	482.97	−0.61
33.4	477.75	478.33	−0.57
37.4	473.13	473.67	−0.54
41.4	468.52	469.02	−0.51
45.4	463.67	464.57	−0.90
49.4	459.07	459.92	−0.85
53.5	454.47	455.26	−0.80
57.5	449.67	450.61	−0.95
61.5	444.89	445.96	−1.07
65.5	439.98	441.31	−1.33
69.5	435.27	436.66	−1.39
73.5	430.44	431.31	−0.86
77.5	425.55	426.11	−0.56
81.6	417.08	420.46	−3.38

1. The sample units $i \in \{1, \ldots, n\}$ are randomly divided into three groups, G_1, G_2 and G_3, of sizes, respectively, $n_1 = 7{,}272$, $n_2 = 1818$ and $n_3 = 478$;

2. Data $(Y_i, \tilde{X}_i : i \in \mathrm{G}_1)$ are used to construct an estimator $\widehat{m}(\cdot)$ for $\mathrm{m}(\cdot)$;

3. The quantiles $\alpha := 0.025$ and $\gamma := 1 - \alpha = 0.975$ of $\{Y_i - \widehat{m}(X_i) : i \in \mathrm{G}_2\}$ are computed, denoted respectively by $\widehat{z}_\alpha$ and $\widehat{z}_\gamma$;

4. The empirical prediction interval $C(x) = [\widehat{m}(x) + \widehat{z}_\alpha;\ \widehat{m}(x) + \widehat{z}_\gamma]$ is constructed;

5. Finally, the empirical coverage $(1/n_3) \sum_{i \in \mathrm{G}_3} \mathbb{I}\{Y_i \in C(X_i)\}$ is calculated.

We replicate the procedure 250 times. Table 8, shows the average and median lengths of the intervals, for $\widehat{m} \in \left\{\widehat{\mathrm{m}}_{\eta^*/2}, \widehat{\mathrm{m}}_{\eta^*}, \widehat{\mathrm{m}}_{2\eta^*}, \widehat{\mathrm{m}}_{h^*}\right\}$, and also the empirical coverage of $\mathrm{PI}(1 - 2\alpha) \equiv \mathrm{PI}(95\%)$. As we can see, the mean and median lengths of the intervals are greater for our estimator $\widehat{\mathrm{m}}_{\eta}$ compared to $\widehat{\mathrm{m}}_{h^*}$. In a way, this result was expected, since, as shown by Figures 3-7 in Fernandes et al. (2021), in their simulations, $\widehat{\beta}_h(\tau)$ stood out by the MSE, not by the empirical coverage of the confidence intervals. Still, the empirical coverages reported in Table 8 are satisfactory and appear to have some robustness to variations in the scale of the smoothing parameter.

TABLE 8: Average and median lengths and empirical coverage of PI(95%).

Estimator	Average length	Median length	Coverage probability
$\widehat{\mathrm{m}}_{\eta^*/2}$	34.63	34.64	0.948
$\widehat{\mathrm{m}}_{\eta^*}$	34.25	34.27	0.944
$\widehat{\mathrm{m}}_{2\eta^*}$	35.70	35.04	0.943
$\widehat{\mathrm{m}}_{h^*}$	19.01	19.02	0.950

5. Final Considerations

In this paper, we proposed an estimator for the conditional mode $\widehat{\mathrm{m}}_\eta(x)$, using as an intermediate step the conditional density estimator $\widehat{\mathtt{f}}_h(\tau|x)$ of Fernandes et al. (2021). We also present the estimator for the conditional mode $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ of Ota et al. (2019) and compare them in a Monte Carlo study, in which we measure the absolute bias and the mean squared error of the estimators in scenarios of: "low", "intermediate" and "extreme" asymmetry, heavy tails and heteroskedasticity. In this study, we defined the process η as a series of bandwidths proportional to the data-driven bandwidth η^* of Silverman (1986), which worked well. Confirming our suspicions, in terms of MSE, $\widehat{\mathrm{m}}_\eta(x)$ systematically outperformed $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ in most of the explored scenarios ("low" asymmetry, heavy tails, and heteroskedasticity), was systematically overcome in only one ("extreme" asymmetry) and, in the case of "intermediate" asymmetry, each of the two estimators $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ won in four of the eight scenarios. Overall, the price to pay for $\widehat{\mathrm{m}}_\eta(x)$ having a smaller MSE when compared to $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$, was a slightly higher bias, which was expected.

In Section 4, following Ota et al. (2019), we used real data from a power plant, to estimate conditional densities of net hourly electrical energy output given exhaust vacuum, using $\widehat{\mathtt{f}}_{\eta(\tau)}(\tau|x)$, as we saw in Figure 12 and, later, conditional modes using $\widehat{\mathrm{m}}_\eta(x)$. We also reproduce the creation of prediction intervals using $\widehat{\mathrm{m}}_\eta(x)$, with bandwidths proportional to η^*. These intervals were longer than those created using $\widehat{\mathrm{m}}_{h^*}$, but were robust to the variation of η.

Among the possible future studies regarding the estimator $\widehat{\mathrm{m}}_\eta(x)$, for the process η, we can consider other configurations than the bandwidth η^* of Silverman (1986). It is also possible, as a way to mitigate the problem of excessive bias in high asymmetry scenarios, to incorporate other kernels in the smoothed objective function $\widehat{R}_h(b;\tau)$ in (8) (in the `conquer` package, the uniform, parabolic and triangular kernels are also implemented). Concomitantly to this article, Zhang et al. (2021) have been working on an improvement for the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ of Ota et al. (2019), using a similar smoothing, although different, to ours. We aim to compare $\widehat{\mathrm{m}}_\eta(x)$ to this new estimator in the near future.

Finally, our most ambitious goal is to develop the asymptotic theory for the estimator $\widehat{\mathrm{m}}_\eta(x)$, since, in most scenarios simulated via Monte Carlo, our estimator has surpassed $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ in terms of mean squared error. In this sense, Proposition 1 in Fernandes et al. (2021) seems to indicate the path to be followed, at least as far as the consistency of $\widehat{\mathrm{m}}_\eta$ is concerned: according to this proposition, it is valid that

$$\left\|\widehat{\mathtt{f}}_h(\tau|x) - \mathtt{f}(\tau|x)\right\| = o(h^s) + O_{\mathbf{P}}\left(\sqrt{\log(n)/(nh)}\right)$$

uniformly for $\tau \in \mathscr{T}$, $h \in \left[h_{[n]}, h^{[n]}\right]$ and $x \in \mathrm{support}(X)$, where $s \geq 1$ controls the smoothness of the derivatives of $f(\cdot|x)$. This result tells us, in particular, that it is possible to identify, with a high probability, the "peak"

$$\big(\mathrm{m}(x), f\big(\mathrm{m}(x)|x\big)\big) \equiv \left(Q(\tau_x|x), \frac{1}{q(\tau_x|x)}\right)$$

using the estimator $\widehat{\mathtt{f}}_{\eta(\tau)}(\tau|x)$, since[5] the stochastic process $\{\eta(\tau)\colon \tau \in \mathscr{T}\}$ takes values in $\left[h_{[n]}, h^{[n]}\right]$ and have almost certainly continuous sample paths. In this sense, it is interesting to inspect Figure 13, where we see the sample paths $\tau \mapsto \eta^*(\tau)$ in 100 Monte Carlo replications with $Z \sim$ Gumbel. About the asymptotic distribution of the estimator proposed here, we conjecture that $\widehat{\mathrm{m}}_\eta(x)$ follows a Chernoff distribution, similarly to the Ota et al. (2019) estimator. See also Finn & Horta (2024) for the asymptotic theory of an estimator similar to ours.

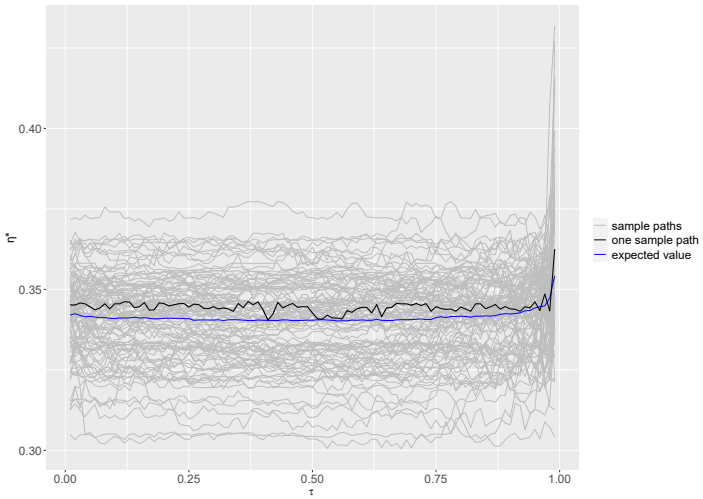

FIGURE 13: Sample paths $\tau \mapsto \eta^*(\tau)$, and their corresponding expected value, in 100 Monte Carlo replications. $Z \sim$ Gumbel.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Artur Mattia gratefully acknowledges the scholarship provided by the CAPES Social Demand (*CAPES DS*) program during his Master's studies.

[Received: September 2025 — Accepted: November 2025]

References

Bamford, S. P., Rojas, A. L., Nichol, R. C., Miller, C. J., Wasserman, L., Genovese, C. R. & Freeman, P. E. (2008), 'Revealing components of the galaxy population through nonparametric techniques', *Monthly Notices of the Royal Astronomical Society* **391**, 607.

[5] Of course, some additional regularity assumptions must be imposed.

Bassett, G. & Koenker, R. (1982), 'An empirical quantile function for linear models with iid errors', *Journal of the American Statistical Association* **77**(378), 407–415.

Chacón, J. E. (2020), 'The modal age of statistics', *International Statistical Review* **88**(1), 122–141.

Chen, Y.-C., Genovese, C. R., Tibshirani, R. J. & Wasserman, L. (2016), 'Non-parametric modal regression', *The Annals of Statistics* **44**(2), 489–514.

Feng, Y., Fan, J. & Suykens, J. (2020), 'A statistical learning approach to modal regression', *Journal of Machine Learning Research* **21**(2), 1–35.

Fernandes, M., Guerre, E. & Horta, E. (2021), 'Smoothing quantile regressions', *Journal of Business and Economic Statistics* **39**(1), 338–357.

Finn, E. S. & Horta, E. (2024), Convolution mode regression, arXiv preprint 2412.05736, arXiv.

He, X., Pan, X., Tan, K. M. & Zhou, W.-X. (2020), *Convolution-type Smoothed Quantile Regression*. R package version 1.0.2.

Kaya, H. & Tufekci, P. (2012), Local and global learning methods for predicting power of a combined gas and steam turbine, *in* 'Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE 2012)', pp. 13–18.

Koenker, R. & Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Lee, M.-J. (1989), 'Mode regression', *Journal of Econometrics* **42**(3), 337–349.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J. & Wasserman, L. (2018), 'Distribution-free predictive inference for regression', *Journal of the American Statistical Association* **113**(523), 1094–1111.

Marsden, J. E. & Tromba, A. (2012), *Vector Calculus*, 6 edn, W. H. Freeman, New York, NY.

Nadaraya, E. A. (1964), 'Some new estimates for distribution functions', *Theory of Probability and Its Applications* **9**(3), 497–500.

Newey, W. K. & McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle & D. L. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, pp. 2111–2245.

Ota, H., Kato, K. & Hara, S. (2019), 'Quantile regression approach to conditional mode estimation', *Electronic Journal of Statistics* **13**, 3120–3160.

R Core Team (2021), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Tufekci, P. (2014), 'Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods', *International Journal of Electrical Power and Energy Systems* **60**, 126–140.

Ullah, A., Wang, T. & Yao, W. (2021), 'Modal regression for fixed effects panel data', *Empirical Economics* **60**(1), 261–308.

Wang, X., Chen, H., Cai, W., Shen, D. & Huang, H. (2017), Regularized modal regression with applications in cognitive impairment prediction, *in* 'Advances in Neural Information Processing Systems', Vol. 30, Curran Associates.

Yao, W. & Li, L. (2014), 'A new regression model: Modal linear regression', *Scandinavian Journal of Statistics* **41**(3), 656–671.

Zhang, T., Kato, K. & Ruppert, D. (2021), 'Bootstrap inference for quantile-based modal regression'.