Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Estatística

Moda condicional: uma abordagem via regressão quantílica suavizada

Artur Mattia Ongaratto

Porto Alegre, Junho de 2021.

Dissertação submetida por Artur Mattia Ongaratto[1] como requisito parcial para a obtenção do título de Mestre em Estatística pelo Programa de Pós-Graduação em Estatística da Universidade Federal do Rio Grande do Sul.

Orientador:

Dr. Eduardo de Oliveira Horta (UFRGS)

Comissão Examinadora:

Dr. Francisco Cribari Neto (UFPE)

Dra. Nathalie Christine Gimenes (PUC-Rio)

Dra. Taiane Schaedler Prass (UFRGS)

Data de Apresentação: 30 de Junho de 2021

[1]Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES

Agradecimentos

Agradeço:

- primeiramente aos meus pais, Eneri e Gilberto, pelo apoio e suporte dado ao longos destes anos de vida acadêmica;
- ao meu namorado, Arthur, pelo companheirismo nas melhores (e compreensão nas piores) horas dessa caminhada;
- ao meu orientador, Eduardo, pelo enorme auxílio ao longo desses mais de dois anos de PPGEst;
- aos meus ex-colegas, hoje amigos, pela parceria em todos os momentos (alô Culto em Capão);
- aos demais professores e funcionários do PPGEst, pelo empenho e dedicação para tornar o Programa um sucesso;
- à Andiara, minha ex-colega de apartamento e melhor amiga para a vida, por sempre me apoiar;
- aos demais amigos e parentes fora da academia, me abstendo de citar nomes, para não correr o risco de deixar alguém de fora;
- ao finado Baltazar, por ser o melhor cachorro do mundo;
- ao Ensino público, que fez parte da minha vida desde a infância, e que apesar de suas falhas, cumpre papel vital na sociedade;
- à banca, pela disponibilidade;
- à CAPES, pela bolsa que me oportunizou escrever esta dissertação;
- à Helen e à Débora pela cessão do modelo de dissertação;
- à Cris, que me apoiou quando decidi entrar nessa jornada, e que infelizmente não está aqui hoje para ver o resultado.

CIP - Catalogação na Publicação

Ongaratto, Artur
Moda condicional: uma abordagem via regressão quantílica suavizada / Artur Ongaratto. -- 2021.
51 f.
Orientador: Eduardo Horta.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Matemática e Estatística, Programa de Pós-Graduação em Estatística, Porto Alegre, BR-RS, 2021.

1. Moda condicional. 2. Regressão quantílica suavizada. 3. Largura de banda baseada nos dados. 4. Simulação de Monte Carlo. I. Horta, Eduardo, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Resumo

Recentemente, Ota, Kato e Hara (2019) propuseram estimar a moda condicional de uma resposta, dado um vetor de covariáveis, por um estimador escalonável computacionalmente derivado do modelo de regressão quantílica linear proposto por Koenker e Bassett (1978). Alternativamente, propomos estimar a moda condicional maximizando o estimador de densidade condicional de Fernandes, Guerre e Horta (2021). Esta aborgadem tem pelo menos dois benefícios: eficiência computacional e bom comportamento assintótico, que, em particular, "contornam" a maldição da dimensionalidade.
Palavras-chave: regressão modal, regressão quantílica, largura de banda baseada nos dados, simulação de Monte Carlo

Abstract

Recently, Ota, Kato e Hara (2019) 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 proposed by Koenker e Bassett (1978). Alternatively, we propose to estimate the conditional mode by maximizing the conditional density estimator of Fernandes, Guerre e Horta (2021). This approach offers at least two benefits: computational efficiency and good asymptotic behavior which, in particular, "bypasses" the curse of dimensionality.
Keywords: modal regression, *quantile regression*, *data-driven bandwidth*, *Monte Carlo simulation*

Índice

Capítulo 1

Introdução

As vantagens de se utilizar a moda como medida de tendência central em um cenário univariado são bastante conhecidas. Além de ser insensível a *outliers*, quando lidamos com distribuições assimétricas ou com caudas pesadas, a moda nos permite capturar informações muitas vezes ignoradas pela média ou pela mediana. Contudo, na prática a configuração univariada é bastante limitada. Em vários cenários, o valor de uma variável de interesse (a variável resposta) é parcialmente determinado pelos valores de outras variáveis (as covariáveis ou regessores), estas conhecidas pelo pesquisador. A análise de regressão é possivelmente o método estatístico mais utilizado para modelar a estrutura de dependência entre uma resposta e covariáveis. Nesse contexto, a regressão modal especifica a forma funcional da moda condicional de uma variável resposta Y, dado um vetor de covariáveis X. Quando as variáveis aleatórias em questão são contínuas, a moda condicional, ou valor "condicionalmente mais provável", nada mais é do que o valor que maximiza a função densidade de probabilidade condicional. Esta forma de regressão tem sido utilizada em diversas áreas, tais como economia, meio ambiente Ullah, Wang e Yao (2021), astronomia Bamford et al. (2008), *machine learning* Feng, Fan e Suykens (2020), medicina Wang et al. (2017) e geologia Chacón (2020).

Na literatura, encontram-se diferentes abordagens para estimar a moda condicional. Dentre as mais conhecidas estão métodos não paramétricos, como em Chen et al. (2016), baseados em um estimador de densidade por kernel, e regressão modal linear, onde assume-se que a moda condicional é linear nas covariáveis, como em Lee (1989) e Yao e Li (2014). Contudo, a primeira alternativa apresenta baixa taxa de convergência (por exemplo, com complexidade $n^{-2/(d+7)}$ no caso de Chen et al. (2016), onde d representa o número de covariáveis no modelo), enquanto a segunda, apesar de contornar a maldição da dimensionalidade, envolve um problema de otimização multidimensional não-convexo. Com o objetivo de evitar os problemas dessas duas abordagens, recentemente Ota, Kato e Hara (2019) propuseram estimar a moda condicional através de um estimador da densidade quantílica condicional, obtido via diferenciação numérica do estimador de Koenker e Bassett (1978) para a função quantílica condicional. Com este método, o problema de otimização multidimensional torna-se um problema unidimensional que pode ser resolvido por uma busca em *grid* e, consequentemente, contorna a maldição da dimensionalidade. Ademais, a proposta dos autores permite maior flexibilidade na forma funcional da moda condicional, assumindo linearidade somente na função quantílica condicional.

Inspirados pela abordagem de Ota, Kato e Hara (2019), nesta dissertação propomos um novo estimador para a moda condicional, obtido via minimização de uma densidade quantílica condicional computada não por diferenciação numérica, mas sim através dos métodos de suavização propostos em Fernandes, Guerre e Horta (2021), os quais permitem que se obtenha um estimador diferenciável para a função quantílica condicional e, *a fortiori* da densidade quantílica condicional. O estimador para a função quantílica introduzido por esses autores apresenta boas propriedades assintóticas: ele é assintoticamente não-viesado e normalmente distribuído e possui menor variância e erro quadrático médio em comparação ao estimador canônico. Partimos da premissa de que estas boas propriedades devem ser herdadas pelo estimador que estamos propondo, e realizamos simulações de Monte Carlo para estudar seu comportamento assintótico em cinco diferentes cenários, comparando nossos resultados aos obtidos utilizando o estimador de Ota, Kato e Hara (2019). Evitando ao máximo o *spoiler*, podemos adiantar que, em nossas simulações e quanto ao erro quadrático médio, um dos estimadores foi sistematicamente superior ao outro na maioria dos cenários. Saindo do "mundo teórico", utilizamos nosso estimador para reproduzir a aplicação a dados reais de Ota, Kato e Hara (2019). Nesta aplicação, coletaram-se dados de uma usina, e primeiramente foram estimadas modas condicionais para a produção de energia elétrica. Em seguida, através do procedimento *split conformal prediction* de Lei et al. (2018), criamos intervalos de predição para a moda condicional, para os quais mais uma vez os resultados foram comparados com os de Ota, Kato e Hara (2019).

Objetivo

Nesta dissertação, nosso objetivo é propor um novo estimador para a moda condicional e estudar seu comportamento assintótico através de simulações, comparando seu desempenho com o do estimador de Ota, Kato e Hara (2019), cuja abordagem deu origem a este trabalho.

Novidades do artigo

Propomos um estimador que faz uso da suavização proposta por Fernandes, Guerre e Horta (2021) para uma densidade quantílica condicional. Este pode ser uma nova ferramenta para a estimação da moda condicional, destacando-se em termos de erro quadrático médio.

Suporte computacional

Realizamos estudos de Monte Carlo para avaliar nosso estimador em termos de viés e erro quadrático médio em diferentes cenários. Também repetimos uma aplicação a dados reais. Toda a parte computacional desta dissertação foi realizada utilizando o software `R`, versão 4.0.3 R Core Team (2021).

Compõem esta dissertação: uma pequena revisão bibliográfica, onde introduzimos conceitos básicos de regressão quantílica e apresentamos as bases do novo estimador proposto; a conclusão acerca do trabalho desenvolvido e alguns possíveis caminhos futuros; o artigo anexo, onde apresentamos nosso estimador proposto;

Capítulo 2

Artigo

O artigo no qual propomos nosso estimador para o moda condiconal, encontra-se no anexo A desta dissertação.

Capítulo 3

Revisão Bibliográfica

Neste capítulo, apresentamos um breve resumo dos conceitos necessários para a compreensão do estudo, além de apresentar o estimador o estimador de regressão quantílica suavizado de Fernandes, Guerre e Horta (2021), utilizado como passo intermediário para construir o novo estimador no presente trabalho.

Seja Y uma variável aleatória escalar, $X \in \mathbb{R}^d$ um vetor de regressores e (Y_i, X_i) com $i = 1, \ldots, n$, uma amostra aleatória de $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$. Koenker e Bassett (1978) propuseram, para o modelo de regressão quantílica, a especificação linear

$$Q(\tau|x) = x'\beta(\tau), \quad \tau \in (0,1),\, x \in \text{suporte}(X), \tag{3.1}$$

onde $\beta\colon (0,1) \to \mathbb{R}^d$ é o parâmetro funcional de interesse. Assumindo que o modelo (3.1) é válido, é particularmente importante a relação

$$f\big(Q(\tau \mid x)\big|x\big) = \frac{1}{q(\tau|x)} \equiv \frac{1}{x'\beta^{(1)}(\tau)}, \quad \tau \in (0,1),\, x \in \text{suporte}(X) \tag{3.2}$$

onde $q(\tau|x) := \partial/\partial\tau Q(\tau|x)$ e $\beta^{(1)}(\tau) = \partial/\partial\tau\beta(\tau)$, válida desde que $F(\cdot|x)$ possua uma densidade contínua $f(\cdot|x)$. Nesse caso, definindo, para τ e x como acima,

$$\mathtt{f}(\tau|x) := \left(Q(\tau|x),\, \frac{1}{q(\tau|x)}\right) \equiv \left(x'\beta(\tau),\, \frac{1}{x'\beta^{(1)}(\tau)}\right), \tag{3.3}$$

vê-se que o *range* do caminho $\tau \mapsto \mathtt{f}(\tau|x)$ coincide com o gráfico de $\big(y, f(y|x)\big)$. Em particular, pode-se definir a **moda condicional de Y dado $X = x$** como sendo a função $\mathrm{m}(x) := Q(\tau_x|x)$, $x \in \text{suporte}(X)$, onde τ_x minimiza, com respeito a $\tau \in (0,1)$, a densidade quantílica $q(\tau|x)$.[1] Para cada $\tau \in (0,1)$ fixado, o vetor $\beta(\tau)$ em (3.1) soluciona um problema de minimização análogo ao encontrado no caso da esperança condicional em uma regressão linear, e sua função objetivo amostral, conforme proposto por Koenker e Bassett (1978), tem a forma

$$\widehat{R}(b;\tau) := \frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(Y_i - X_i'b\right), \quad \tau \in (0,1),\, b \in \mathbb{R}^d, \tag{3.4}$$

onde $\rho_\tau(u) := u \cdot (\tau - \mathbb{I}(u \leq 0))$, para $\tau \in (0,1)$ e $u \in \mathbb{R}$, é chamada de **check function** — ver, por exemplo, Angrist e Pischke (2009, equação 7.1.2). Para $\tau \in (0,1)$, o **estimador**

[1] Evidentemente, são necessárias algumas suposições de regularidade sobre $f(\cdot|x)$ para garantir que m esteja bem definida.

canônico de regressão quantílica é o vetor $\widehat{\beta}(\tau)$ dado por

$$\widehat{\beta}(\tau) = \arg\min_{b\in\mathbb{R}^d} \widehat{R}\,(b;\tau), \tag{3.5}$$

Portanto, é possível, em teoria, obter um estimador para a função densidade de probabilidade condicional de Y dado $X = x$, utilizando a relação em (3.2), através da diferenciação de um estimador do parâmetro $\beta(\tau)$. Contudo, conforme mostram Bassett e Koenker (1982, p.409), a função quantílica condicional empírica $\tau \mapsto x'\widehat{\beta}(\tau)$, estimada através de (3.5), possui saltos, o que dificulta que se explore a relação em (3.2) para estimação da densidade condicional e, *a fortiori*, da moda condicional. Para contornar este problema, Fernandes, Guerre e Horta (2021) sugerem estimar o parâmetro $\beta(\tau)$ em (3.1) de maneira diferente, substituindo a *check function* em (3.4) por uma versão suavizada através de um *kernel*, de maneira similiar a Nadaraya (1964). Como resultado, no lugar da função objetivo amostral $\widehat{R}(b;\tau)$, tem-se uma versão suavizada

$$\widehat{R}_h\,(b;\tau) := \frac{1}{n}\sum_{i=1}^{n} k_h * \rho_\tau(Y_i - X_i'b), \quad \tau \in (0,1),\, b \in \mathbb{R}^d, \tag{3.6}$$

em que $*$ denota a operação de convolução[2] e onde k_h é um *kernel* com parâmetro de suavização $h > 0$ (h tende a 0 à medida que o tamanho amostral aumenta). Analogamente a (3.5), para $\tau \in (0,1)$, o **estimador suavizado de regressão quantílica** é o vetor $\widehat{\beta}_h(\tau)$ dado por

$$\widehat{\beta}_h(\tau) = \arg\min_{b\in\mathbb{R}^d} \widehat{R}_h\,(b;\tau). \tag{3.7}$$

Conforme demonstrado em Fernandes, Guerre e Horta (2021), ao contrário da função objetivo em (3.4), aquela expressa em (3.6) é continuamente diferenciável, e isso traz pelo menos dois benefícios. O primeiro é que o estimador suavizado em (3.7) herda a regularidade da função objetivo suavizada em (3.6). O segundo é que a diferenciabilidade permite estimar a matriz de covariância assintótica dos coeficientes de inclinação de um modo canônico — por exemplo, ver Newey e McFadden (1994).

Incidentalmente, as propriedades citadas acima permitem que se obtenha, de maneira natural, um estimador para a função densidade de probabilidade condicional de Y dado $X = x$: a derivada segunda da função objetivo amostral suavizada (3.6) de Fernandes, Guerre e Horta (2021) é dada por

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

Ademais, visto que $\widehat{\beta}_h(\tau)$ satisfaz a condição $\widehat{R}_h^{(1)}\big(\widehat{\beta}_h(\tau);\tau\big) = 0$, onde $\widehat{R}_h(b;\tau)$ denota o gradiente de $\widehat{R}_h(b;\tau)$ com respeito a b, segue pelo teorema da função implícita em Marsden e Tromba (2012) que $\widehat{\beta}_h(\tau)$ é continuamente diferenciável com respeito a τ, com

$$\frac{\partial}{\partial\tau}\widehat{\beta}_h(\tau) =: \widehat{\beta}_h^{(1)}(\tau) = \left[\widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau);\tau\right)\right]^{-1}\bar{X}. \tag{3.8}$$

Os autores utilizam esse fato para estimar $\mathtt{f}(\tau|x)$, substituindo $\beta(\tau)$ por $\widehat{\beta}_h(\tau)$ em (3.3) para

[2] Dadas duas funções reais φ e ψ satisfazendo uma condição de integrabilidade, define-se a convolução de φ e ψ pela igualdade $\varphi * \psi(u) = \int_{\mathbb{R}} \varphi(v)\psi(u-v)\,\mathrm{d}v$.

obter

$$\widehat{\mathtt{f}}_h(\tau|x) := \left(x'\widehat{\beta}_h(\tau), \frac{1}{x'\widehat{\beta}_h^{(1)}(\tau)} \right), \quad \tau \in (0,1), x \in \mathrm{suporte}(X). \tag{3.9}$$

Pela Proposição 1 em Fernandes, Guerre e Horta (2021), esse estimador é uniformemente consistente, com respeito a τ e h, para $\mathtt{f}$.

Como mencionamos no capítulo 1, a ideia de utilizar a relação em (3.2) para estimar a função densidade de probabilidade condicional foi abordada por Ota, Kato e Hara (2019), que propuseram obter um estimador para a densidade condicional $\mathtt{f}(\tau|x)$ a partir do cômputo de uma diferenciação numérica, e a partir disso, criaram o estimador para moda condicional $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$. Esta ideia inspirou Fernandes, Guerre e Horta (2021), fazendo com que propusessem o estimador suavizado para a densidade condicional $\widehat{\mathtt{f}}_h(\tau|x)$, computado através de uma suavização na função objetivo do estimador canônico. Utilizamos uma adaptação de $\widehat{\mathtt{f}}_h(\tau|x)$ como passo intermediário para propor um novo estimador para a moda condicional, $\widehat{\mathrm{m}}_\eta(x)$, apresentado no artigo anexo.

Capítulo 4

Conclusões e trabalhos futuros

Nesta dissertação, propusemos o estimador para a moda condicional $\widehat{\mathrm{m}}_\eta(x)$, utilizando como passo intermediário, o estimador para a densidade condicional $\widehat{\mathtt{f}}_h(\tau|x)$ de Fernandes, Guerre e Horta (2021). Apresentamos também o estimador para a moda condicional $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ de Ota, Kato e Hara (2019), e comparamos os dois em um estudo de Monte Carlo, no qual medimos o viés absoluto e o erro quadrático médio dos estimadores em cenários de: assimetria "baixa", "intermediária" e "extrema", caudas pesadas e de heteroscedasticidade. Neste estudo, definimos o processo η como uma série de *bandwidth*'s proporcionais à *data-driven bandwidth* η^* de Silverman (1986), o que funcionou bem. Confirmando nossas suspeitas, em termos de EQM, $\widehat{\mathrm{m}}_\eta(x)$ superou sistematicamente $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ na maioria dos cenários explorados (assimetria "baixa", caudas pesadas e heteroscedasticidade), foi sistematicamente superado em apenas um (assimetria "extrema") e, no caso de assimetria "intermediária", foi sobrepujado por $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ para amostras menores e sobrepujou para amostras maiores. No geral, o menor EQM de $\widehat{\mathrm{m}}_\eta(x)$ teve como custo um viés um pouco maior em relação a $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ o que já era esperado. Ainda, reproduzimos a aplicação de Ota, Kato e Hara (2019) a dados reais, na qual estimamos densidades condicionais a partir de uma adaptação do estimador da densidade condicional de Fernandes, Guerre e Horta (2021), a saber, $\widehat{\mathtt{f}}_{\eta(\tau)}(\tau|x)$, e criamos intervalos de predição. Estes intervalos, apesar de terem tido comprimento maior do que os criados através de $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$, foram pouco influenciados pela variação de η.

Dentre os possíveis estudos futuros em relação ao estimador $\widehat{\mathrm{m}}_\eta(x)$, podemos considerar outras configurações para o processo η, que não a *bandwidth* η^* de Silverman (1986). Também é possível, como forma de mitigar o problema de viés excessivo em cenários de assimetria elevada, incorporar outros kernels na função objetivo suavizada $\widehat{R}_h\,(b;\tau)$ em (3.6). Concomitantemente ao presente artigo, Zhang, Kato e Ruppert (2021) vêm trabalhando em um aprimoramento para o estimador $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ de Ota, Kato e Hara (2019), utilizando uma suavização semelhante, embora distinta, à nossa. Almejamos comparar $\widehat{\mathrm{m}}_\eta(x)$ a esse novo estimador em um futuro próximo.

Por fim, nossa meta mais ambiciosa é desenvolver a teoria assintótica para o estimador $\widehat{\mathrm{m}}_\eta(x)$, uma vez que, na maioria dos cenários simulados via Monte Carlo, nosso estimador superou $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ em termos de erro quadrático médio.

Referências Bibliográficas

Angrist, J.; Pischke, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. 1. ed. Princeton, New Jersey: Princeton University Press, 2009.

Bamford, S. P. et al. Revealing components of the galaxy population through nonparametric techniques. *Mon. Not. Roy. Astron. Soc.*, v. 391, p. 607, 2008.

Bassett, G.; Koenker, R. An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association*, Taylor & Francis, v. 77, n. 378, p. 407–415, 1982.

Chacón, J. E. The modal age of statistics. *International Statistical Review*, v. 88, n. 1, p. 122–141, 2020. Disponível em: ⟨https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12340⟩.

Chen, Y.-C. et al. Nonparametric modal regression. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 44, n. 2, p. 489 – 514, 2016. Disponível em: ⟨https://doi.org/10.1214/15-AOS1373⟩.

Feng, Y.; Fan, J.; Suykens, J. A statistical learning approach to modal regression. *Journal of Machine Learning Research*, v. 21, n. 2, p. 1–35, 2020. Disponível em: ⟨http://jmlr.org/papers/v21/17-068.html⟩.

Fernandes, M.; Guerre, E.; Horta, E. Smoothing quantile regressions. *Journal of Business & Economic Statistics*, Taylor and Francis, v. 39, n. 1, p. 338–357, 2021. Disponível em: ⟨https://doi.org/10.1080/07350015.2019.1660177⟩.

Koenker, R.; Bassett, G. Regression quantiles. *Econometrica*, v. 46, n. 1, p. 33–50, 1978.

Lee, M. jae. Mode regression. *Journal of Econometrics*, v. 42, n. 3, p. 337–349, 1989. ISSN 0304-4076. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/0304407689900572⟩.

Lei, J. et al. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, Taylor & Francis, v. 113, n. 523, p. 1094–1111, 2018.

Marsden, J. E.; Tromba, A. *Vector Calculus*. 6. ed. New York, NY: W. H. Freeman, 2012.

Nadaraya, E. A. Some new estimates for distribution functions. *Theory of Probability & Its Applications*, v. 9, n. 3, p. 497–500, 1964.

Newey, W. K.; McFadden, D. Chapter 36: Large sample estimation and hypothesis testing. In: *Handbook of Econometrics*. [S.l.]: Elsevier, 1994. v. 4, p. 2111–2245.

Ota, H.; Kato, K.; Hara, S. Quantile regression approach to conditional mode estimation. *Electronic Journal of Statistics*, v. 13, p. 3120–3160, 2019. Disponível em: ⟨https://doi.org/10.1214/19-EJS1607⟩.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: ⟨https://www.R-project.org/⟩.

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

Ullah, A.; Wang, T.; Yao, W. Modal regression for fixed effects panel data. *Empirical Economics*, v. 60, n. 1, p. 261–308, January 2021. Disponível em: ⟨https://ideas.repec.org/a/spr/empeco/v60y2021i1d10.1007_s00181-020-01999-w.html⟩.

Wang, X. et al. Regularized modal regression with applications in cognitive impairment prediction. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: ⟨https://proceedings.neurips.cc/paper/2017/file/bea5955b308361a1b07bc55042e25e54-Paper.pdf⟩.

Yao, W.; Li, L. A new regression model: Modal linear regression. *Scandinavian Journal of Statistics*, v. 41, n. 3, p. 656–671, 2014. Disponível em: ⟨https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12054⟩.

Zhang, T.; Kato, K.; Ruppert, D. *Bootstrap inference for quantile-based modal regression*. 2021.

Anexo A

Artigo Mattia e Horta

Conditional mode: An approach via smoothed quantile regression

Artur Ongaratto[1] **, Eduardo Horta**[1]

[1]Universidade Federal do Rio Grande do Sul, Brazil

Abstract. *Recently, Ota, Kato and Hara (2019) 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 proposed by Koenker and Bassett (1978). Alternatively, we propose to estimate the conditional mode by maximizing the conditional density estimator of Fernandes, Guerre and Horta (2021). This approach offers at least two benefits: computational efficiency and good asymptotic behavior which, in particular, "bypasses" the curse of dimensionality.*

Keywords: mode regression, quantile regression, data-driven bandwidth, Monte Carlo study

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, Wang and Yao (2021), astronomy Bamford et al. (2008), machine learning Feng, Fan and Suykens (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 and 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, Kato and Hara (2019) proposed to estimate the conditional mode through a conditional quantile density estimator, obtained via numerical differentiation of the Koenker and 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 mean, assuming linearity only in the conditional quantile function.

Inspired by the approach of Ota, Kato and Hara (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, Guerre and Horta (2021), which allow one to obtain a differentiable estimator for the conditional quantile function and, *a fortiori*, of the conditional quantile density. 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, Kato and Hara (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, Kato and Hara (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, Kato and Hara (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, Guerre and Horta (2021), and we discuss the smoothing parameter for the kernel. We also present the estimator of Ota, Kato and Hara (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, Kato and Hara (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 and 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 $(y, f(y|x))$. 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

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

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 and Bassett (1978), has the form

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

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 and 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 explore the relationship in (3) to estimate the conditional density and, *a fortiori*, the conditional mode. To avoid this problem, Fernandes, Guerre and Horta (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, Guerre and Horta (2021), unlike the objective function in (6), the one expressed in (8) is continuously differentiable, and this has at least two benefits. The first

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

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

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 and 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, Guerre and Horta (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 and Tromba (2012) that $\widehat{\beta}_h(\tau)$ is continuously differentiable with respect to τ, with

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

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, Guerre and Horta (2021), this estimator is consistent for $\mathtt{f}$, uniformly with respect to τ and h.

2.1. Estimator for conditional mode

We propose to employ an adaptation of the estimator $\widehat{\mathtt{f}}_h(\tau|x)$ of Fernandes, Guerre and Horta (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 $\left[h_{[n]}, h^{[n]}\right]$ (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

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

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

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

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

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

and where h is a bandwidth that will be discussed, as well as the process η used for our estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$, in section 3. For more details about the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ in (13), see Ota, Kato and Hara (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 $h \in \left[h_{[n]}, h^{[n]}\right]$ — see Fernandes, Guerre and Horta (2021, Theorems 1, 3 and 5, and also Proposition 1). 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 Fernandes, Guerre and Horta (2021, Theorem 2). The properties mentioned above give us clues that the estimator $\widehat{\mathrm{m}}_\eta(x)$ must dominate $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ 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 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, Guerre and Horta (2021), we generate the data from the model $Y = X'\beta + Z$, where $X = (1 \quad \tilde{X})'$, with $\tilde{X} \sim \mathsf{Uniform}[1, 5]$ 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 ($\mathrm{skew}(Z) = 2$), where the conditional mode is a value at the boundary of the support of $f(\cdot|x)$. Here, the error term Z follows $\mathsf{Exponential}(1/\sqrt{2})$ distribution, independent of X.
2. **Gumbel error term.** The purpose of this specification is to evaluate the estimators under a "low" asymmetry scenario ($\mathrm{skew}(Z) \approx 1.44$). Here, the error term Z follows $\mathsf{Gumbel}(0, \sqrt{12}/\pi)$ distribution, independent of X.
3. χ^2 **error term.** In this specification, we want to evaluate the estimators in an "intermediate" asymmetry scenario ($\mathrm{skew}(Z) = 1.63$). Here, we have the error term $Z = W - 1$, where $W \sim \chi^2_3$ is independent of X.
4. **Student's $\boldsymbol{t}$ 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 $W \sim \mathsf{N}(0, 24/13)$ is independent of $\tilde{X}$.

[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.

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 $\mathrm{Var}(Z) = 2$. For each of the five specifications described above, we sample $n \in \{100, 250, 500, 1{,}000\}$ 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 $j \in \{1, \ldots, n_{\text{repl}}\}$, we compute the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}[j]}(x)$ of Ota, Kato and Hara (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 $\mathrm{Bias}_{\text{mc}}(m, x)$ expressed by

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

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

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

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, Guerre and Horta (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, Guerre and Horta (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

[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.

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 $Z \sim \mathsf{Exponential}(1/\sqrt{2})$, $\tilde{x} = 3$ and $n = 1{,}000$, we noticed (in preliminary simulations) that smaller η's produced estimates with smaller MSE_{mc} and $\mathrm{Bias}_{\text{mc}}$. 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 $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ of Ota, Kato and Hara (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 $\mathsf{N}(0,1)$, 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, Kato and Hara (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 $\widehat{\mathrm{m}}_h^{\mathrm{o}}$ 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 $\mathrm{Bias}_{\mathrm{mc}}$ and $\mathrm{MSE}_{\mathrm{mc}}$. 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 $n = 1{,}000$ and $\tilde{x} = 3$, the idea of using $\widehat{\mathrm{m}}_{\eta^*/8}$ was rewarding (resulting in the smaller $\mathrm{Bias}_{\mathrm{mc}}$ and $\mathrm{MSE}_{\mathrm{mc}}$ among all $\widehat{\mathrm{m}}_\eta(x)$ estimators).

		$n=100$		$n=250$		$n=500$		$n=1{,}000$	
		$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$
$\tilde{x}=1.5$	$\widehat{\mathrm{m}}_{h^*}$	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

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

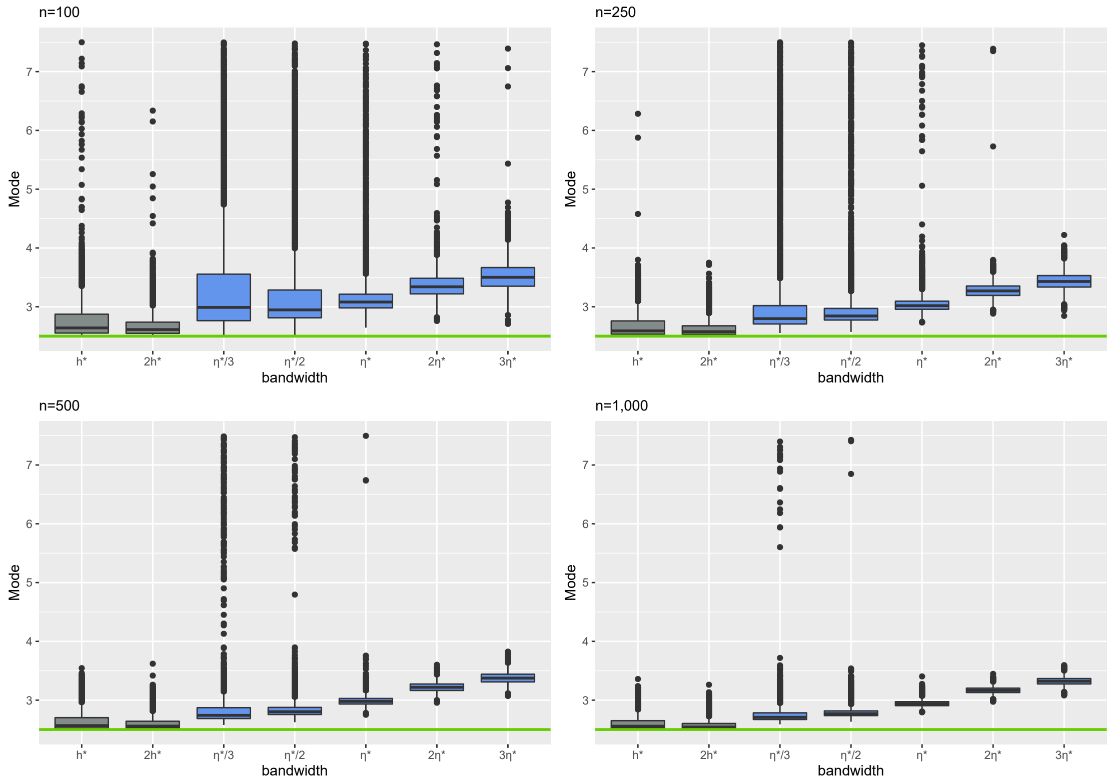

Figure 1. Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim$ Exponential

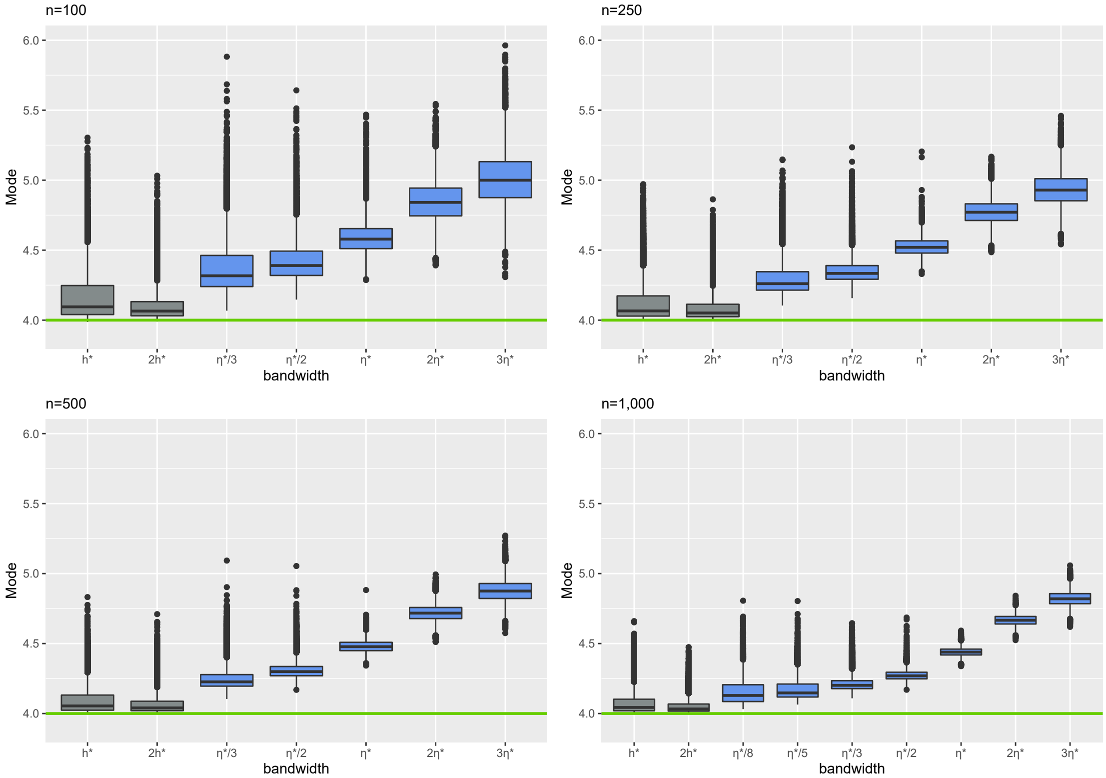

Figure 2. Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}^o_h(x)$ for $x = (1 \quad 3)'$ with $Z \sim$ Exponential

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 $n = 1{,}000$. 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 $n = 1{,}000$, $\tilde{x} = 3$, 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 $n = 1{,}000$, 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).

		$n = 100$		$n = 250$		$n = 500$		$n = 1{,}000$	
		$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^*/2}$	**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 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

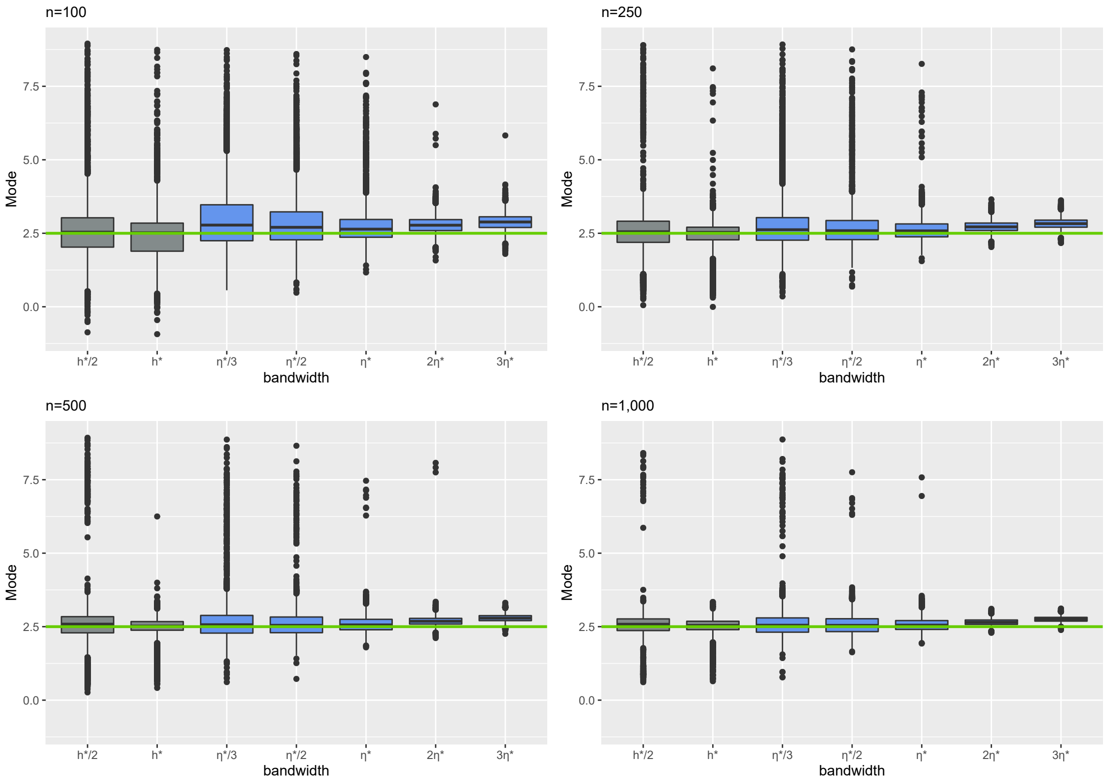

Figure 3. Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}^o_h(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim$ Gumbel

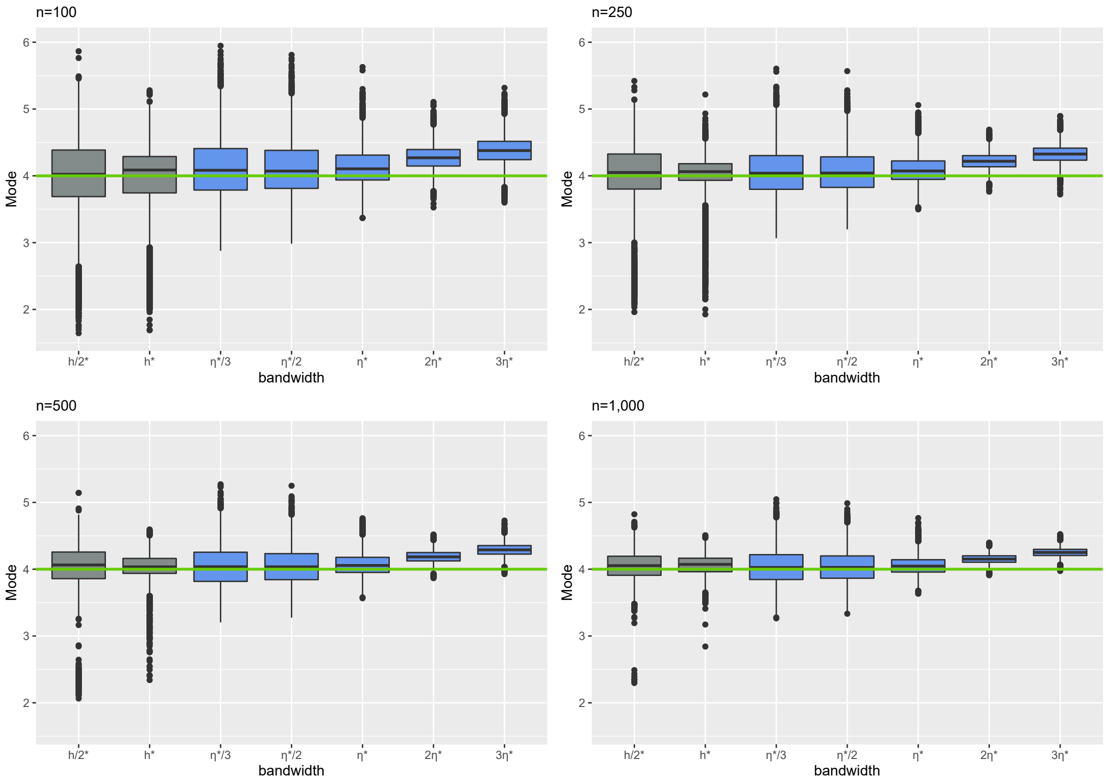

Figure 4. Boxplots of $\widehat{\mathrm{m}}_{\eta}(x)$ and $\widehat{\mathrm{m}}_{h}^{o}(x)$ for $x = (1 \quad 3)'$ with $Z \sim$ Gumbel

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 $\eta = \eta^*$; $\tilde{x} = 1.5$; $n = 1{,}000$ and with $\eta = {}^{\eta^*}\!/2$; $\tilde{x} = 3$; $n = \{250, 500, 1{,}000\}$. 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 $n = 1{,}000$, 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.

		$n = 100$		$n = 250$		$n = 500$		$n = 1{,}000$	
		$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$	$\mathrm{Bias}_{\mathrm{mc}}$	$\mathrm{MSE}_{\mathrm{mc}}$
$\tilde{x} = 1.5$	$\widehat{\mathrm{m}}_{h^*}$	**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

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 $Z \sim \chi^2_3$

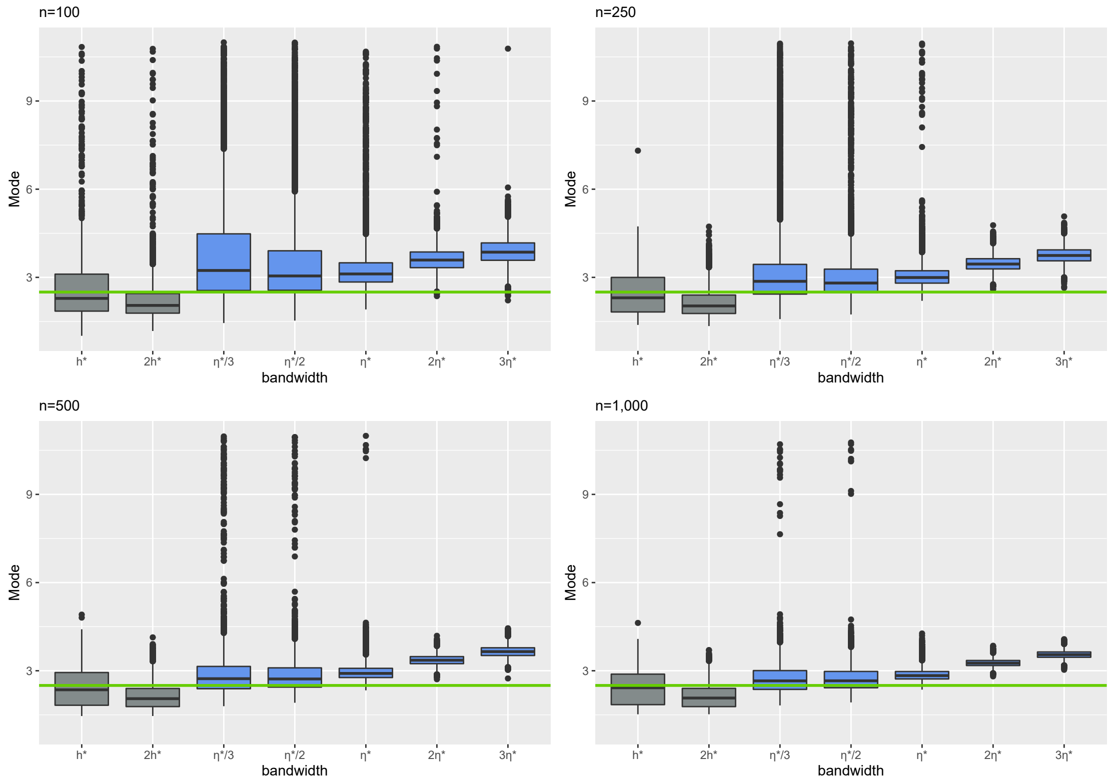

Figure 5. Boxplots of $\widehat{\mathrm{m}}_{\eta}(x)$ and $\widehat{\mathrm{m}}^{o}_{h}(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim \chi^2_3$

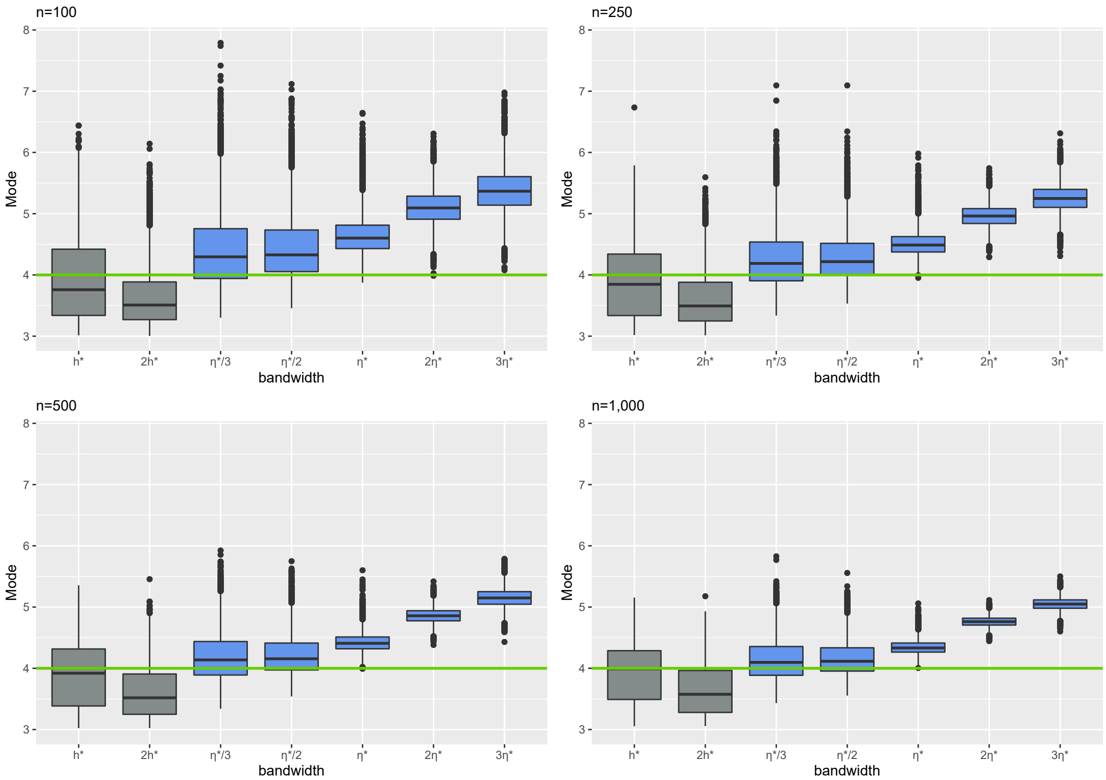

Figure 6. Boxplots of $\widehat{\mathrm{m}}_{\eta}(x)$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_{h}(x)$ for $x = (1 \quad 3)'$ with $Z \sim \chi^2_3$

3.3.4. Student's *t* error term

In Table 4 we see that, for the Student's $\boldsymbol{t}$ error term, the Bias_{mc} of both estimators was negligible in all evaluated scenarios, as expected because it is a symmetric distribution. Therefore, we will only analyze the MSE_{mc}. It is noteworthy that $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ presented much higher MSE_{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 $n = 1{,}000$. 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 $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$.

		$n = 100$		$n = 250$		$n = 500$		$n = 1{,}000$	
		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**

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 $Z \sim \boldsymbol{t}$

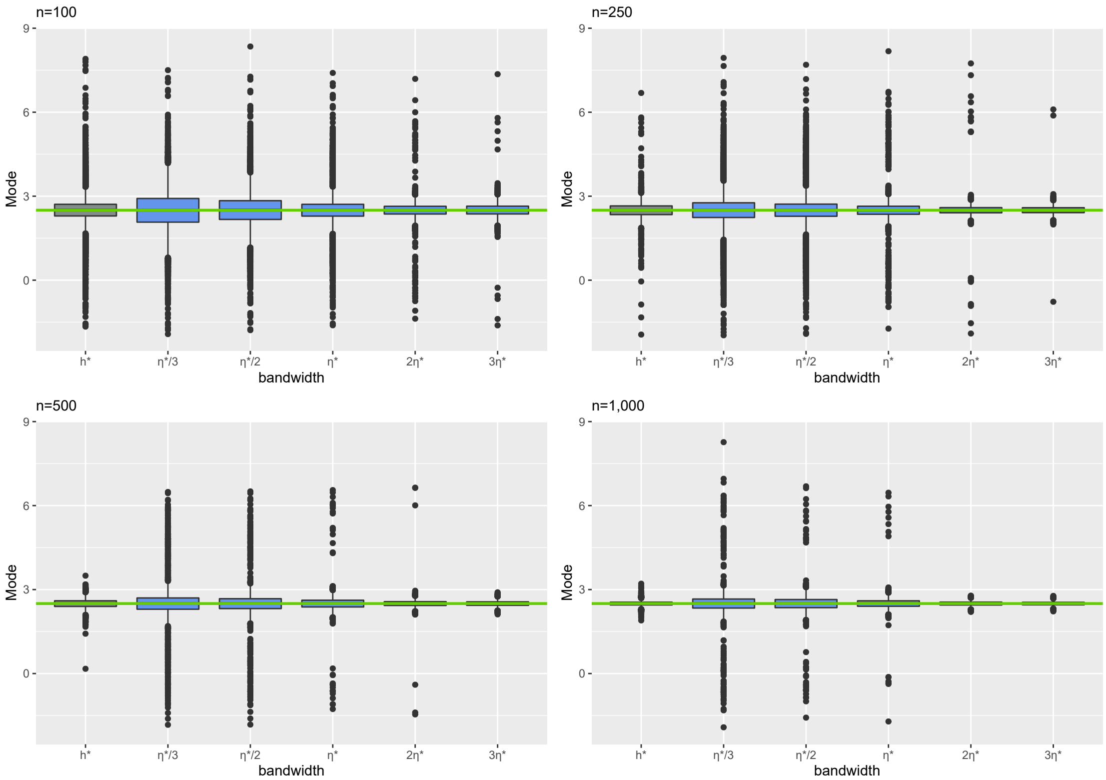

Figure 7. Boxplots of $\widehat{\mathrm{m}}_{\eta}(x)$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_{h}(x)$ for $x = (1 \quad 1.5)'$ with $Z \sim \boldsymbol{t}$

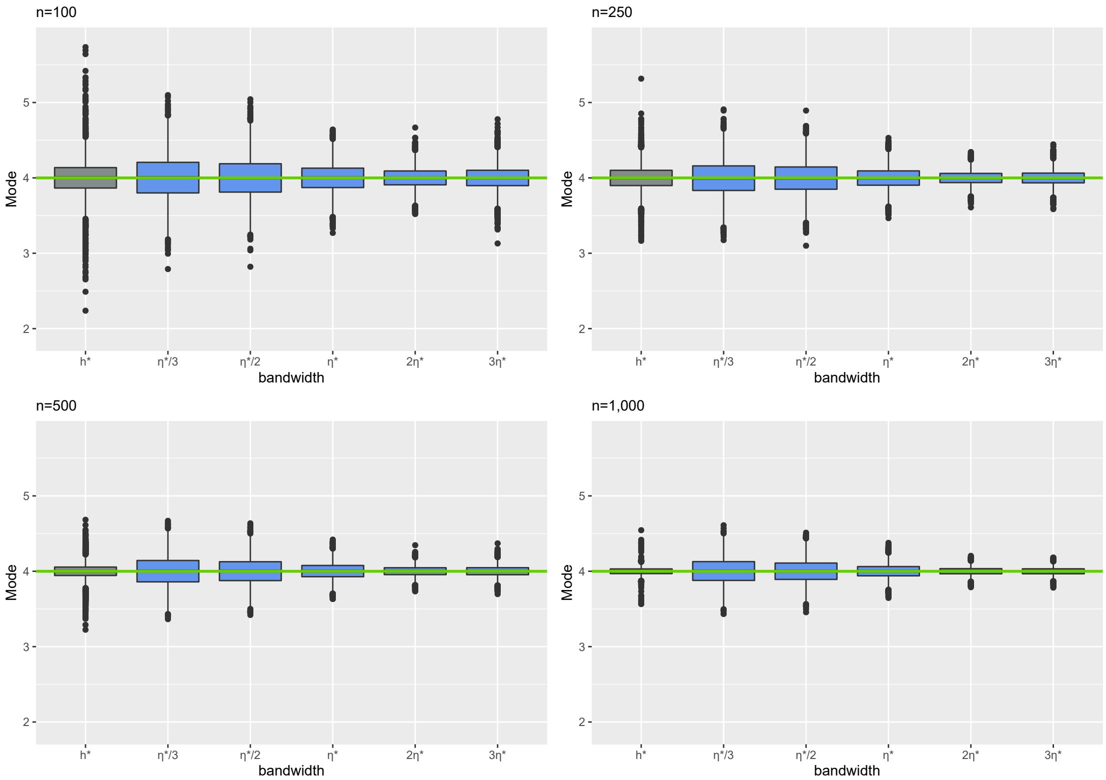

Figure 8. Boxplots of $\widehat{\mathrm{m}}_{\eta}(x)$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_{h}(x)$ for $x = (1 \quad 3)'$ with $Z \sim \boldsymbol{t}$

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 $\widehat{\mathrm{m}}^{\mathrm{o}}_{h}(x)$. 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.

		$n = 100$		$n = 250$		$n = 500$		$n = 1{,}000$	
		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**

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

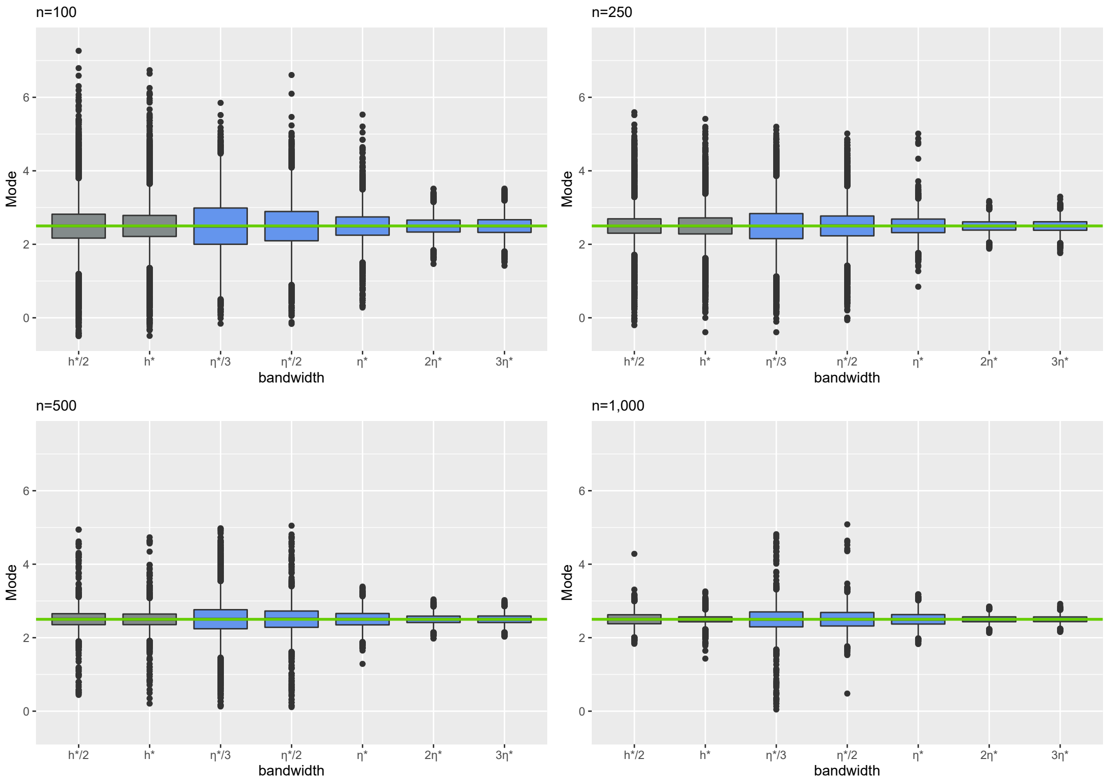

Figure 9. Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ for $x = (1 \quad 1.5)'$ with Z heteroskedastic

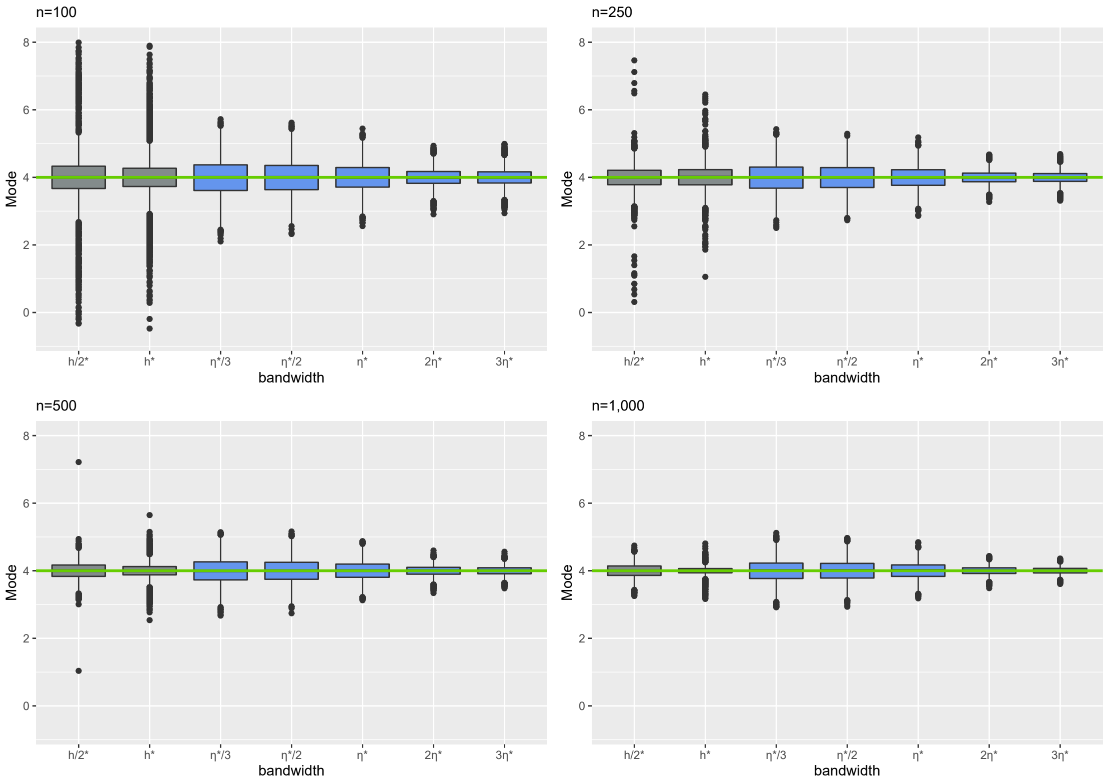

Figure 10. Boxplots of $\widehat{\mathrm{m}}_\eta(x)$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$ for $x = (1 \quad 3)'$ with Z heteroskedastic

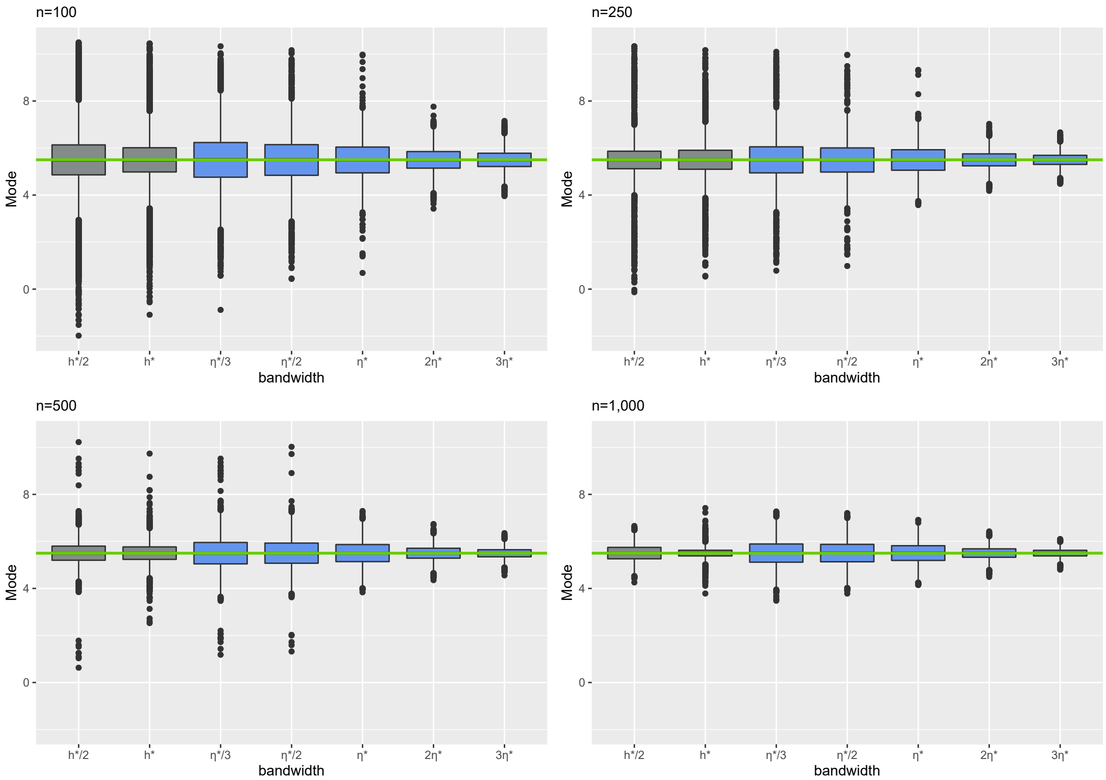

Figure 11. Boxplots of $\widehat{\mathrm{m}}_{\eta}(x)$ and $\widehat{\mathrm{m}}_{h}^{\mathrm{o}}(x)$ for $x = (1 \quad 4.5)'$ with Z heteroskedastic

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 $n = 1{,}000$. 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 $n = 1{,}000$.

		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
	$n = 1{,}000$	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
	$n = 1{,}000$	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
	$n = 1{,}000$	–	–	–	–	0.689

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

4. Application

In this section, we describe the result of the repetition of an application performed by Ota, Kato and Hara (2019), using and comparing the estimators $\widehat{\mathrm{m}}_\eta$ and $\widehat{\mathrm{m}}^{\mathrm{o}}_h$ described in section 2. The application was made using the Combined Cycle Power Plant Data, 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 and Tufekci (2012). The bank 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 Ota, Kato and Hara (2019, Figure 3). In Table 7, we see the comparison between the estimates for the conditional mode using $\widehat{\mathrm{m}}_{\eta^*}$ and $\widehat{\mathrm{m}}_{h^*}$.

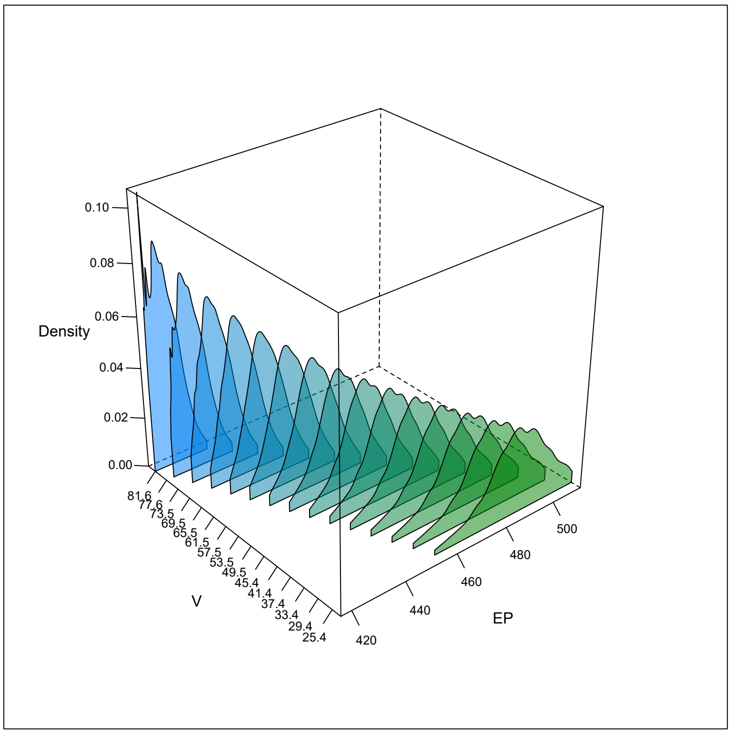

Figure 12. Estimated conditional densities of net hourly electrical energy output given exhaust vacuum

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, Kato and Hara (2019) created prediction intervals (PI) with theoretical coverage of 95%, using their estimator $\widehat{\mathrm{m}}^{\mathrm{o}}_h(x)$, with

$\tilde{X} = \tilde{x}$	$\widehat{\mathrm{m}}_{\eta^*}$	$\widehat{\mathrm{m}}_{h^*}$	$(\widehat{\mathrm{m}}_{\eta^*} - \widehat{\mathrm{m}}_{h^*})$
25.4	486.975	487.614	−0.64
29.4	482.361	482.966	−0.61
33.4	477.747	478.319	−0.57
37.4	473.132	473.671	−0.54
41.4	468.518	469.024	−0.51
45.4	463.671	464.569	−0.90
49.4	459.069	459.917	−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

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

$h = h^*$ 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:

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, $n_1 = 7{,}272$, $n_2 = 1{,}818$ 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 $\widehat{m} \in \{\widehat{\mathrm{m}}_{\eta^*/2}, \widehat{\mathrm{m}}_{\eta^*}, \widehat{\mathrm{m}}_{2\eta^*}, \widehat{\mathrm{m}}_{h^*}\}$, and also the empirical coverage of PI$(1-2\alpha) \equiv$ 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 Fernandes, Guerre and Horta (2021, Figures 3-7), 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.

Estimador	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

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

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, Guerre and Horta (2021). We also present the estimator for the conditional mode $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ of Ota, Kato and Hara (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, Kato and Hara (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, Kato and Ruppert (2021) have been working on an improvement for the estimator $\widehat{\mathrm{m}}_h^{\mathrm{o}}(x)$ of Ota, Kato and Hara (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, Guerre and Horta (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 $\{\eta(\tau)\colon \tau \in \mathscr{T}\}$ take 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

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

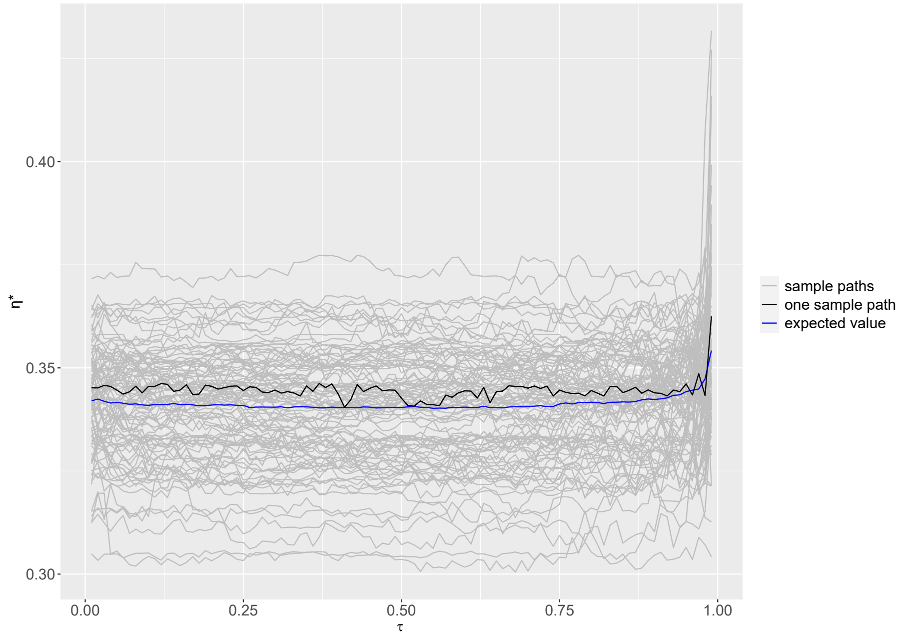

Figure 13. Sample paths $\tau \mapsto \eta^*(\tau)$, and their corresponding expected value, in 100 Monte Carlo replications. $Z \sim$ Gumbel

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, Kato and Hara (2019) estimator.

References

Bamford, S. P. et al. Revealing components of the galaxy population through nonparametric techniques. *Mon. Not. Roy. Astron. Soc.*, v. 391, p. 607, 2008. 1

Bassett, G.; Koenker, R. An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association*, Taylor & Francis, v. 77, n. 378, p. 407–415, 1982. 3

Chacón, J. E. The modal age of statistics. *International Statistical Review*, v. 88, n. 1, p. 122–141, 2020. Available on: ⟨https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12340⟩. 1

Chen, Y.-C. et al. Nonparametric modal regression. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 44, n. 2, p. 489 – 514, 2016. Available on: ⟨https://doi.org/10.1214/15-AOS1373⟩. 1

Feng, Y.; Fan, J.; Suykens, J. A statistical learning approach to modal regression. *Journal of Machine Learning Research*, v. 21, n. 2, p. 1–35, 2020. Available on: ⟨http://jmlr.org/papers/v21/17-068.html⟩. 1

Fernandes, M.; Guerre, E.; Horta, E. Smoothing quantile regressions. *Journal of Business & Economic Statistics*, Taylor and Francis, v. 39, n. 1, p. 338–357, 2021. Available on: ⟨https://doi.org/10.1080/07350015.2019.1660177⟩. 1, 2, 3, 4, 5, 6, 26, 27

He, X. et al. *Convolution-Type Smoothed Quantile Regression*. 2020. R package version 1.0.2. 5

Kaya, H.; Tufekci, P. Local and global learning methods for predicting power of a combined gas & steam turbine. In: *Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering ICETCEE 2012*. 2012. p. 13–18. 25

Koenker, R.; Bassett, G. Regression quantiles. *Econometrica*, v. 46, n. 1, p. 33–50, 1978. 1, 2, 3

Lee, M. jae. Mode regression. *Journal of Econometrics*, v. 42, n. 3, p. 337–349, 1989. ISSN 0304-4076. Available on: ⟨https://www.sciencedirect.com/science/article/pii/0304407689900572⟩. 1

Lei, J. et al. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, Taylor & Francis, v. 113, n. 523, p. 1094–1111, 2018. 2, 26

Marsden, J. E.; Tromba, A. *Vector Calculus*. 6. ed. New York, NY: W. H. Freeman, 2012. 4

Nadaraya, E. A. Some new estimates for distribution functions. *Theory of Probability & Its Applications*, v. 9, n. 3, p. 497–500, 1964. 3

Newey, W. K.; McFadden, D. Chapter 36: Large sample estimation and hypothesis testing. In: *Handbook of Econometrics*. : Elsevier, 1994. v. 4, p. 2111–2245. 4

Ota, H.; Kato, K.; Hara, S. Quantile regression approach to conditional mode estimation. *Electronic Journal of Statistics*, v. 13, p. 3120–3160, 2019. Available on: ⟨https://doi.org/10.1214/19-EJS1607⟩. 1, 2, 4, 6, 7, 25, 27, 28

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Available on: ⟨https://www.R-project.org/⟩. 5

Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall, 1986. 6, 27

Tufekci, P. 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 & Energy Systems*, v. 60, p. 126–140, 2014. ISSN 0142-0615. 25

Ullah, A.; Wang, T.; Yao, W. Modal regression for fixed effects panel data. *Empirical Economics*, v. 60, n. 1, p. 261–308, January 2021. Available on: ⟨https://ideas.repec.org/a/spr/empeco/v60y2021i1d10.1007_s00181-020-01999-w.html⟩. 1

Wang, X. et al. Regularized modal regression with applications in cognitive impairment prediction. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Available on: ⟨https://proceedings.neurips.cc/paper/2017/file/bea5955b308361a1b07bc55042e25e54-Paper.pdf⟩. 1

Yao, W.; Li, L. A new regression model: Modal linear regression. *Scandinavian Journal of Statistics*, v. 41, n. 3, p. 656–671, 2014. Available on: ⟨https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12054⟩. 1

Zhang, T.; Kato, K.; Ruppert, D. *Bootstrap inference for quantile-based modal regression*. 2021. 27