A new bandwidth selection method for nonparametric modal regression based on generalized hyperbolic distributions

Hongpeng Yuan [*], Sijia Xiang [†], and Weixin Yao [‡]

Abstract

As a complement to standard mean and quantile regression, nonparametric modal regression has been broadly applied in various fields. By focusing on the most likely conditional value of Y given x, the nonparametric modal regression is shown to be resistant to outliers and some forms of measurement error, and the prediction intervals are shorter when data is skewed. However, the bandwidth selection is critical but very challenging, since the traditional cross-validation method cannot be applied. We propose to select the bandwidth by applying the asymptotic global optimal bandwidth and the flexible generalized hyperbolic (GH) distribution as the distribution of the error. Unlike the plug-in method, the new method does not require preliminary parameters to be chosen in advance, is easy to compute by any statistical software, and is computationally efficient compared to the existing kernel density estimator (KDE) based method. Numerical studies show that the GH based bandwidth performs better

[*] School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang 310017, P.R. China.

[†] Corresponding author. School of Data Sciences, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang 310018, P.R. China.

[‡] Department of Statistics, University of California, Riverside, CA 92887, U.S.A.

than existing bandwidth selector, in terms of higher coverage probabilities. Real data applications also illustrate the superior performance of the new bandwidth.

Key words: generalized hyperbolic distribution; nonparametric modal regression; kernel density estimator

1 Introduction

When we try to model the relationship between a response Y and covariates $\mathbf{X}$, the conditional mean $\mathrm{E}(Y|\mathbf{X})$, is modeled linearly or nonparametrically. However, when data is skewed, contaminated or contain outliers, such as wages, prices and expenditures in econometric applications, the mean cannot reveal much useful information, and thus, the mean regression is no longer appropriate. As a result, the quantile regression and modal regression, regression analysis built on the conditional quantiles and mode, were proposed as complements. Figure 1 illustrates the relationship between the three location measures, namely the mean, median and mode, show interesting stories from the perspective of confidence intervals. For skewed data, when the confidence intervals are of the same length, the CI for mode has the highest coverage probability. When the CIs are of the same coverage probability, the CI for mode is the shortest in width.

By assuming the mode of conditional density $f(y|\mathbf{x})$ to be a linear function of $\mathbf{x}$, Lee (1989, 1993); Lee and Kim (1998); Kemp and Santos Silva (2012); Yao and Li (2014) proposed modal linear regressions. Better prediction performance and robustness show the superior performance of the new regression methodology in the numerical studies. Ota et al. (2019) proposed to estimate the conditional mode based on a linear quantile regression model, and studied its asymptotic distribution. By minimizing the derivative of estimated conditional quantile function, Zhang et al. (2021) estimated the conditional mode, and further developed

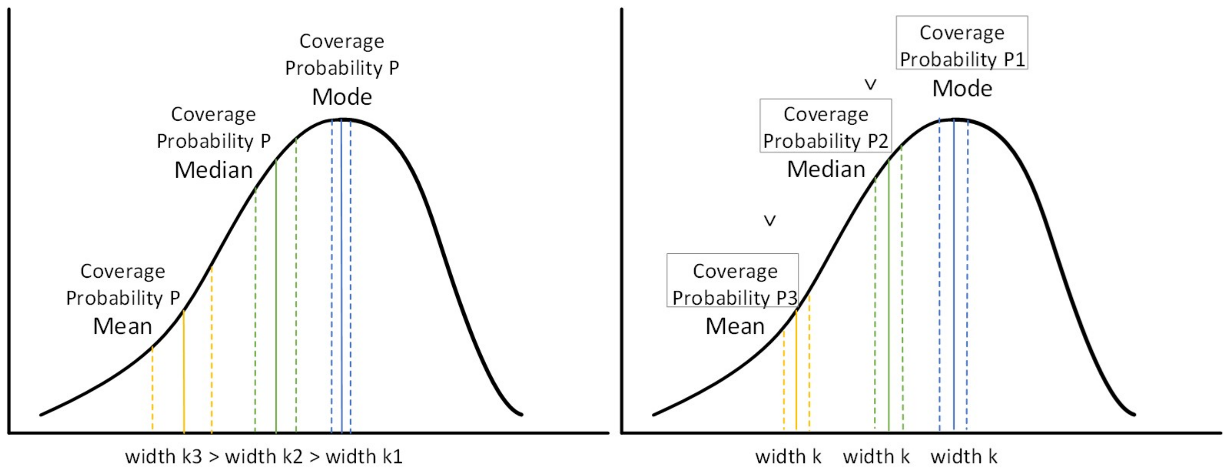

Figure 1: Comparison of mean, median and mode: (a) widths of intervals with the same coverage probabilities; (b) coverage probabilities of intervals with the same width.

bootstrap inference technique for the conditional mode estimator.

However, similar to any other parametric method, modal linear regression might be misleading and give unsatisfactory prediction performance when the strong parametric assumptions about linearity does not hold. Feng et al. (2020) approached the nonparametric modal regression problem from a classical empirical risk minimization point of view. Xiang and Yao (2022) proposed a local polynomial modal regression (LPMR), which estimate the mode of $f(y|\mathbf{x})$ through local polynomial regressions. Krief (2017) studied a partial linear modal regression problem, and proposed a consistent and asymptotic normal estimator. Ullah et al. (2021, 2022a) further applied modal regression to fixed effects panel data and time series sequences, respectively. Ullah et al. (2022b) proposed a semiparametric partially linear varying coefficient modal regression to expand the applicability of the modal regression.

Due to the "curse of dimensionality", the dimension of $\mathbf{X}$ is restricted to be $d = 1$, and assume $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ to be an iid sample from $f(x, y)$. Then, the LPMR assumes that

$$\text{Mode}(y|x) = \arg\max_y f(y|x) = m(x),$$

where $m(\cdot)$ is an unknown but smooth function. Xiang and Yao (2022) proposed to estimate $m(\cdot)$ at a fixed point x_0 by kernel density estimator (KDE). To be more specific, the LPMR maximizes the following log-likelihood:

$$\ell(\boldsymbol{\theta}) \equiv \frac{1}{n}\sum_{i=1}^{n} K_{h_1}(x_i - x_0)\phi_{h_2}\left(y_i - \sum_{j=0}^{p}\beta_j(x_i - x_0)^j\right), \tag{1.1}$$

where $\boldsymbol{\theta} = (\beta_0, \ldots, \beta_p)^\top$, $K_h(x) = h^{-1}K(x/h)$ and $\phi_h(t) = h^{-1}\phi(t/h)$ are the symmetric kernel functions and (h_1, h_2) are the bandwidths. An EM algorithm is proposed, and then the v-th derivative of $m(x)$ can be estimated by $\hat{m}_v(x_0) = v!\hat{\beta}_v,$ for $v = 0, \ldots, p$.

Bandwidth selection has always been a crucial issue in nonparametric and semiparametric techniques. Classical bandwidth selection techniques include cross-validation (CV), generalized cross-validation (GCV), plug-in method, and information based criteria, such as AIC and BIC. The CV and GCV aim to minimize an unbiased estimator of the mean average squared error, while AIC focused on the expected Kullback-Leibler discrepancy. Being totally automatic, CV and GCV are popularly used tools, but have always been criticized for being time consuming. In addition, several new bandwidth selection methodologies have been proposed in the last few years, see for example, Shang (2013); Levine (2013); Sun and Li (2011); Xiang and Yao (2016, 2018).

Chen et al. (2018) studied a new method that is different from the traditional plug-in method. By minimizing a cross-validated criterion function, the method is fully data-driven, and do not need a good initial value. While proposing a new nonparametric density estimation procedure, Kirkby et al. (2021) suggested to select the bandwidth by an efficient cross-validation procedure, based on closed-form expressions in terms of the primal and dual B-spline basis.

In this article, we propose a new bandwidth selector for LLMR based on generalized hyperbolic distribution (GH) and the asymptotic global bandwidth of Xiang and Yao (2022).

Xiang and Yao (2022) proposed nonparametric modal regression and provide the asymptotic global bandwidth by minimizing the asymptotic mean integrated squared errors. However, their "optimal" bandwidth can't be used directly because of the unknown error densities. We propose to choose the bandwidth by assuming a generalized hyperbolic distribution for the error density and then plugging-in unknown quantities with their estimators. First introduced by Barndorff-Nielsen (1978), GH distributions have been widely applied in financial modelling McNeil et al. (2005), mainly because GH distributions are semi-heavy tailed, and have quite some special and limiting cases, such as variance-gamma, hyperbolic, normal-inverse Gaussian, t distribution, skew t, etc. Browne and McNicholas (2015) introduced a multivariate mixture of generalized hyperbolic distributions, as a complement to the traditional mixture of Gaussian distributions, and mixture of t-distributions and mixture of skew-t distributions. By fitting generalized hyperbolic mixtures on a reduced subspace, Morris and McNicholas (2016) systematically applied GH mixtures to dimension reduction in clustering, classification and discriminant analysis. See Choi et al. (2021); Gaunt and Merkle (2021) for more related work on GH.

Numerical study show that the new method performs better or comparable to existing method, in terms of prediction performance. Some real data applications are also provided to illustrate the effectiveness of the new method. The rest of the article is organized as follows. The derivation of the new method is given in Section 2. In Section 3 and 4, simulation studies and real data examples are shown. A discussion section ends the article.

2 Generalized hyperbolic based bandwidth selector

Introduced by Barndorff-Nielsen (1977), the name of generalized hyperbolic (GH) distribution is based on the fact that the log-density of its distribution has the shape of a hyperbola. Due to the five free parameters, the GH distributions can be very flexible and effective, and contain many special and limiting cases, such as the Gaussian, t, variance-gamma, inverse Gaussian, Laplace and skew-t distribution, and has been popularly used in in modelling extreme values, and thus popularly used in financial and risk management. Recently, Browne and McNicholas (2015); Morris and McNicholas (2016) have applied generalized hyperbolic distributions to clustering, classification, discriminant analysis and dimension reduction.

Consider the GH distribution with probability density function (pdf)

$$f(x;\lambda,\alpha,\beta,\delta,\mu)=\frac{(\gamma/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\gamma)}e^{\beta(x-\mu)}\frac{K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2})}{(\sqrt{\delta^2+(x-\mu)^2}/\alpha)^{1/2-\lambda}},\quad x\in\mathbb{R},$$

where $\gamma=\sqrt{\alpha^2-\beta^2}$ and $K_{\lambda}(\cdot)$ is a modified Bessel function of the third kind, μ is a location parameter, δ is a scale parameter, λ is a shape parameter, and α and β describes the kurtosis and skewness. Browne and McNicholas (2015) describes several limiting cases of the GH distribution. For example, when $\lambda=1$, it becomes the hyperbolic distribution. For $\lambda=-1/2$, we obtain the normal-inverse Gaussian (NIG) distribution.

Figure 2 shows the log-density of several GH distributions. By taking a log transformation, the log-density of Gaussian, t and variance-gamma appear very close to each other, while the other two GH distributions are quite different. Therefore, the presence of the index parameter λ makes GH extremely flexible, which is not found in its special and limiting cases.

For the LPMR (1.1), as well known, the choice of kernels is not very important, but the selection of the optimal smoothing parameters (h_1,h_2) is critical to the estimation of $m(x)$. After deriving the asymptotic variance and asymptotic bias, Xiang and Yao (2022) used a global optimal bandwidth, which minimize the asymptotic weighted mean integrated squared error

$$\int\left([\text{bias}\{\hat{m}(x_0)\mid X\}]^2+\text{var}\{\hat{m}(x_0)\mid X\}\right)w(x)dx=\frac{K}{nh_1h_2^3}+Mh_1^4+Nh_2^4+2Lh_1^2h_2^2,$$

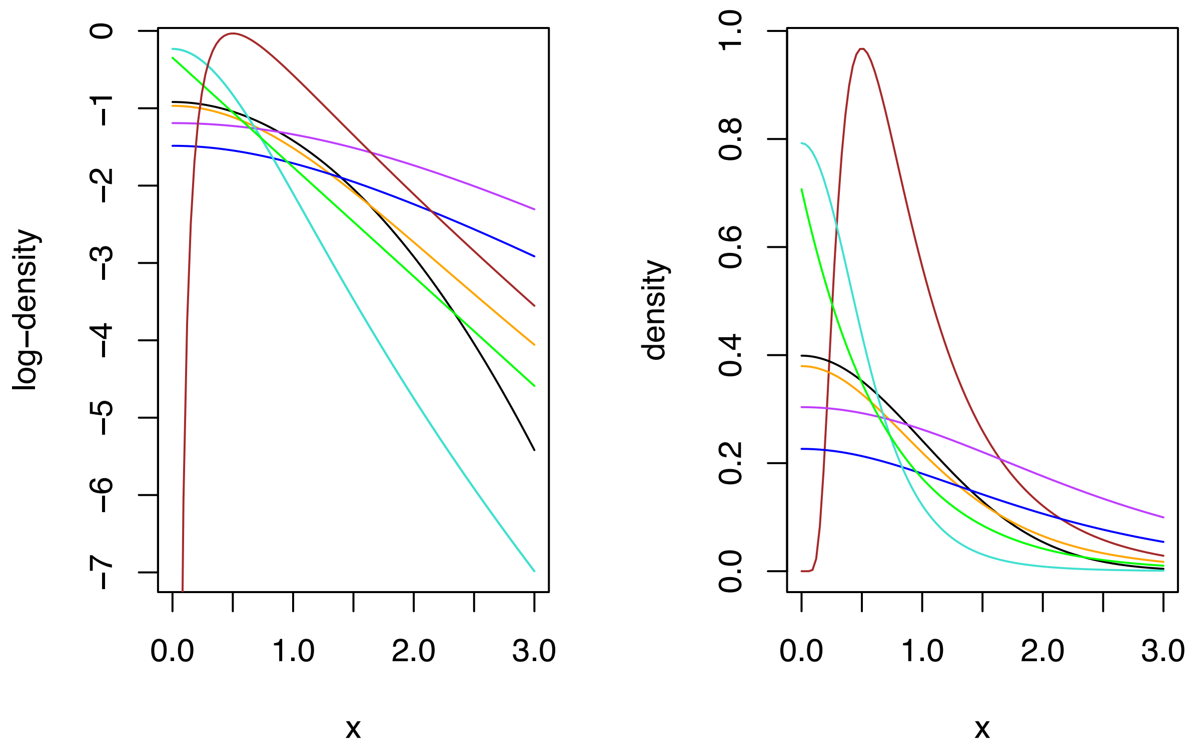

Figure 2: Log-density and density of the generalized hyperbolic distribution with $\lambda = 2$ (blue) and $\lambda = 2$ (turquoise), Gaussian distribution (black), t-distribution with five degrees of freedom (orange), variance-gamma distribution with Vgc=5(green), inverse Gaussian distribution with $\mu = 1, \lambda = 0.5$ (brown), skew t-distribution with five degrees of freedomskewing parameter $\gamma = 2$ (darkorchid).

where

$$K = \int \frac{g(0 \mid x)\tilde{\nu}\nu_0}{g''(0 \mid x)^2 f(x)} w(x)dx, \quad M = \int \left[\frac{1}{2}m''(x)\mu_2\right]^2 w(x)dx,$$
$$N = \int \left[-\frac{g'''(0 \mid x)}{2g''(0 \mid x)}\right]^2 w(x)dx, \quad L = \int \left[\frac{1}{2}m''(x)\mu_2\right]\left[-\frac{g'''(0 \mid x)}{2g''(0 \mid x)}\right] w(x)dx,$$

$w(x)$ is a weight function, such as 1 or the design density $f(x)$, and $g(\cdot|x)$ is the conditional distribution of the residual ϵ given x. When $p = 1$ and $v = 0$, the estimator is referred to as the local linear modal regression (LLMR), and the corresponding asymptotic global optimal bandwidth is

$$\begin{aligned} \hat{h}_1 &= \left[\frac{3K}{4n\delta^5(L+N\delta^2)}\right]^{1/8}, \\ \hat{h}_2 &= \delta\hat{h}_1, \end{aligned} \tag{2.1}$$

where $\delta^2 = (\sqrt{L^2+3MN}+L)/N$. However, since the quantities $g(0|x), g^{''}(0|x), g^{'''}(0|x)$ and $m^{''}(x)$ are unknown beforehand, the bandwidth (2.1) is not ready to use. In Xiang and Yao (2022), the authors applied a series of approximations to estimate the unknown quantities. First, a polynomial of order three is applied to approximate the unknown function as $m(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$, and the parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)^\top$ can be readily estimated by the modal linear regression (Yao and Li, 2014), as $\hat{\boldsymbol{\alpha}}$. Then, $\hat{m}^{''}(x) = 2\hat{\alpha}_2 + 6\hat{\alpha}_3 x$ is ready for use. Let $\hat{\epsilon}_i = y_i - \hat{m}(x_i)$, Xiang and Yao (2022) applied KDE to approximate $g(0|x), g^{''}(0|x), g^{'''}(0|x)$ as

$$\hat{g}^{(\nu)}(0|x) = \frac{1}{h^{\nu+1}}\sum_{i=1}^{n} K^{(\nu)}\left\{\frac{\hat{\epsilon}_i - \hat{m}(x_i)}{h}\right\}, \nu = 0, 2, 3.$$

In this article, we assume the error distribution $g(\cdot)$ to be GH. After the estimated errors are calculated based on the aforementioned polynomial approximation, we fit a GH to $\hat{\epsilon}_i$, and the maximum likelihood estimators of the parameters are calculated. Note that the GH is very general with 5 parameters and the generality makes the likelihood function very flat, causing problems with fitting. To solve this problem, in practice, we fit the data with different sub models, such as hyperbolic distribution (hyperb), generalized inverse Gaussian (gig), normal inverse Gaussian (NIG), variance gamma (VG) and skew hyperbolic (SH), and pick the one that has the largest log-likelihood. After the parameters are estimated, the quantities $g(0|x), g^{''}(0|x), g^{'''}(0|x)$ can then be calculated. The detailed formulations are deferred to the Appendix.

3 Numerical studies

In this section, we use Monte Carlo simulations and real data examples to investigate the finite sample performance of the newly proposed bandwidth selection method, and compare it with the existing KDE based bandwidth selector. All the simulations are conducted using R, and the fitting of the GH distribution is done through the `GeneralizedHyperbolic` package.

3.1 Simulation study

The model considered in this simulation study is as follows:

$$Y = 2\sin(\pi X) + \sigma(X)\epsilon,$$

where $X \sim U(0,1)$, $\sigma(X) = 1 + 2X$. The error is assumed to follow one of the following distributions:

(I.) $\epsilon \sim N(0,1)$;

(II.) $\epsilon \sim t_3$;

(III.) $\epsilon \sim 0.5N(-1, 2.5^2) + 0.5N(1, 0.5^2)$;

(IV.) $\epsilon \sim \chi^2(4)$;

(V.) $\epsilon \sim SkewLaplace(1,1,2)$;

(VI.) $\epsilon \sim 0.5N(-2, 1^2) + 0.5N(2, 2^2)$;

where the pdf of skew Laplace distribution is

$$f(x;\mu,\alpha,\beta) \begin{cases} \frac{1}{\alpha+\beta} e^{\frac{x-\mu}{\alpha}}, & \text{if } x \leq \mu; \\ \frac{1}{\alpha+\beta} e^{-\frac{x-\mu}{\beta}}, & \text{otherwise.} \end{cases}$$

Case I and II are symmetric, and Case II is a classic heavy-tailed distribution, Case III, IV and V are skewed distributions, and Case VI is a multi-modal distribution. Suppose that the error density ϵ has a mode at c. Then, the modal regression in our simulation study is

$$\text{Mode}(y|x) = \arg\max_y f(y|x) = m(x) = 2\sin(\pi X) + c\sigma(x).$$

Different shapes of error distributions have different c values and therefore produce different shapes of modal regression. Figure 3 shows the modal regression $m(x)$ of each case.

The performance of the LPMR is reported, where the bandwidth is selected by either Xiang and Yao (2022), referred to as LPMR(KDE), or the new GH based method, referred to as LPMR(GH). The sample sizes $n = 100, 200$ and 400 are conducted over 500 replications.

The prediction performance are compared in two aspects. Table 1 reports the average(std) of the coverage probabilities of prediction intervals of the same length (symmetric about each estimate), over the 500 replications. Three interval lengths are considered: $0.1\sigma, 0.2\sigma$ and 0.5σ, where $\sigma = 2$ approximates the standard deviation of ϵ. In addition, with 30%, 50% and 90% as the levels of confidence, Table **??** reports the average widths and percentage of coverage of the prediction intervals. The confidence interval is constructed based on the similar method suggested by Yao and Li (2014), which could make use of the skewness of the error distribution assumed by LPMR. From Table 1 and 2, we can see that when the error is symmetrically distributed (Case I & II), the new method performs comparable to the KDE based method. When the error distribution is skewed or bimodal (Case III-VI), for confidence intervals with the same length, the CI based on LPMR(GH) has higher coverage

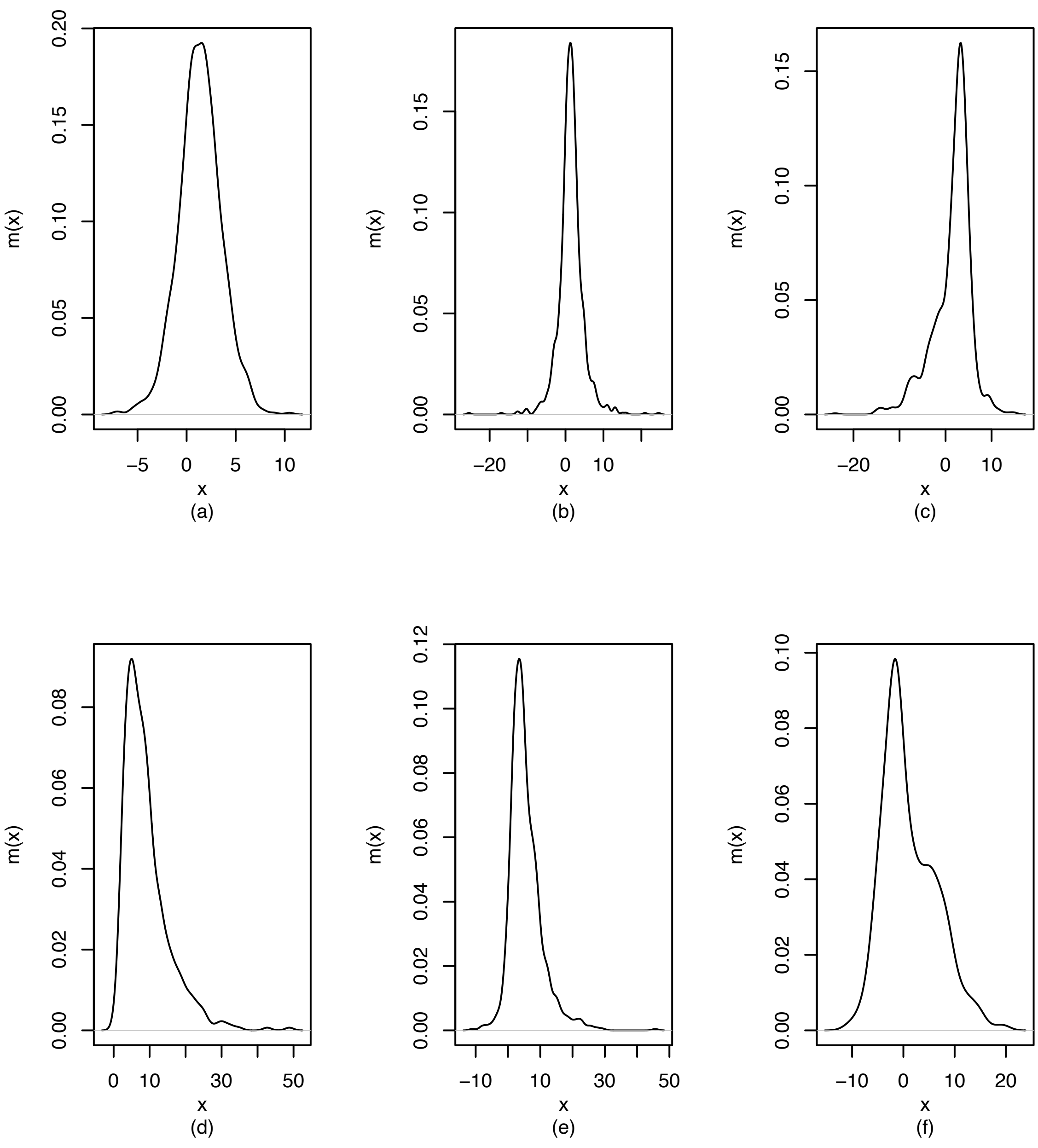

Figure 3: The mode regression $m(x)$ of Case I – VI.

probability, especially for wider CIs.

Table 1: Average (Std) of percentage of coverage with $\sigma = 2$.

Case	Width	$n = 100$		$n = 200$		$n = 400$	
		LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)
I	0.1σ	0.082(0.002)	0.081(0.003)	0.083 (0.001)	0.082 (0.002)	0.084 (0.001)	0.083 (0.001)
	0.2σ	0.164(0.003)	0.160(0.005)	0.165 (0.002)	0.164 (0.003)	0.166 (0.001)	0.165 (0.002)
	0.5σ	0.392(0.007)	0.385(0.012)	0.395(0.004)	0.392 (0.007)	0.398 (0.002)	0.395 (0.005)
II	0.1σ	0.074(0.003)	0.073(0.003)	0.076 (0.001)	0.075 (0.002)	0.077 (0.001)	0.075 (0.002)
	0.2σ	0.147(0.005)	0.145(0.006)	0.150 (0.003)	0.148 (0.004)	0.152 (0.002)	0.150 (0.003)
	0.5σ	0.350(0.011)	0.344(0.013)	0.356 (0.005)	0.352 (0.009)	0.360 (0.003)	0.355 (0.007)
III	0.1σ	0.073(0.015)	0.078(0.015)	0.078 (0.012)	0.080 (0.011)	0.087 (0.012)	0.085 (0.012)
	0.2σ	0.144(0.029)	0.154(0.028)	0.154 (0.024)	0.158 (0.020)	0.172 (0.022)	0.168 (0.023)
	0.5σ	0.336(0.053)	0.355(0.048)	0.354 (0.040)	0.359 (0.033)	0.388 (0.037)	0.381 (0.038)
IV	0.1σ	0.033(0.003)	0.036(0.002)	0.035 (0.002)	0.037 (0.002)	0.036 (0.002)	0.037 (0.001)
	0.2σ	0.067(0.005)	0.072(0.004))	0.069 (0.004)	0.074 (0.004)	0.071 (0.003)	0.075 (0.003)
	0.5σ	0.166(0.013)	0.178(0.011)	0.172 (0.010)	0.183 (0.009)	0.178 (0.008)	0.185 (0.006)
V	0.1σ	0.051(0.006)	0.057(0.006)	0.054 (0.005)	0.059 (0.004)	0.056 (0.005)	0.061 (0.003)
	0.2σ	0.103(0.011)	0.113(0.011)	0.108 (0.010)	0.118 (0.007)	0.111 (0.009)	0.121 (0.005)
	0.5σ	0.253(0.022)	0.271(0.020)	0.265 (0.018)	0.281 (0.011)	0.273 (0.016)	0.286 (0.006)
VI	0.1σ	0.031(0.008)	0.039(0.009)	0.029 (0.005)	0.035 (0.005)	0.033 (0.005)	0.037 (0.003)
	0.2σ	0.061(0.016)	0.078(0.018)	0.059 (0.011)	0.069 (0.010)	0.067 (0.009)	0.074 (0.006)
	0.5σ	0.154(0.038)	0.192(0.041)	0.147 (0.024)	0.170 (0.021)	0.165 (0.020)	0.180 (0.013)

3.2 Real data examples

Example 1 (*Air quality data*) The Air Quality data contains the hourly air pollutant data of the air quality monitoring stations of 12 stations in Beijing, and is from the Beijing Municipal Environmental Monitoring Center. The data set includes the hourly data of six major air pollutants and six associated meteorological variables at each location.

In this study, we are interested in how PM2.5 is related to dew point temperature, and we apply the data from 0am to 5am of 2017. Figure 4 shows the fitted modal regression and 95% prediction bounds of the air quality data based on different bandwidths. It can be seen that the curve fitted by the GH based bandwidth is smoother than the KDE based bandwidth. In addition, we use Monte-Carlo cross-validation (MCCV) and d-fold cross-validation (CV) to compare the prediction performance. In MCCV, the data points are randomly partitioned into disjoint training subset of size $n - d$, and testing subsets of size d, and the procedure is repeated for 200 times. The average (std) of coverage probability of CIs with different widths are summarized in Table 3, where $\sigma = 2$. It can be seen that the GH based bandwidth selector could offer higher coverage probability than the KDE based method.

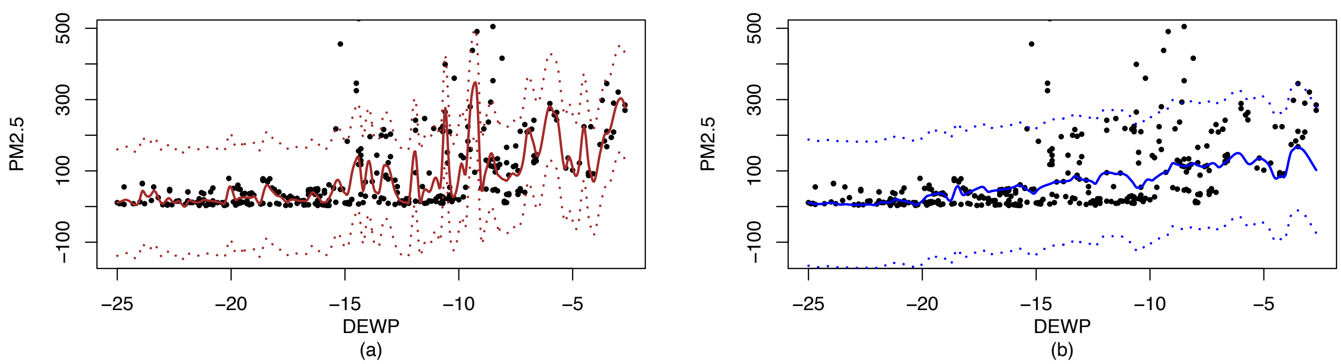

Figure 4: Scatter plot, fitted modal regression and 95% prediction bounds of the air quality data: (a) KDE based bandwidth; (b) GH based bandwidth.

Example 2 (*Forest Fire Data*)

The forest fire data contains 517 observations, collected between January 2000 and December 2003, from the Montesinho Natural Park in the northeastern region of Portugal. We are interested in how temperature (temp) affects relative humidity (RH). Figure 5 shows the fitted modal regression and 95% prediction bounds of the data. We have similar findings to Example 1, i.e, GH based bandwidth provides smoother curves.

Again, with CV and MCCV, we calculate the average (std) of percentage of coverage of CIs with the same widths with $\sigma = 2$, and the results are summarized in Table 4. We can also see that the coverage probability improved a lot by LPMR(GH).

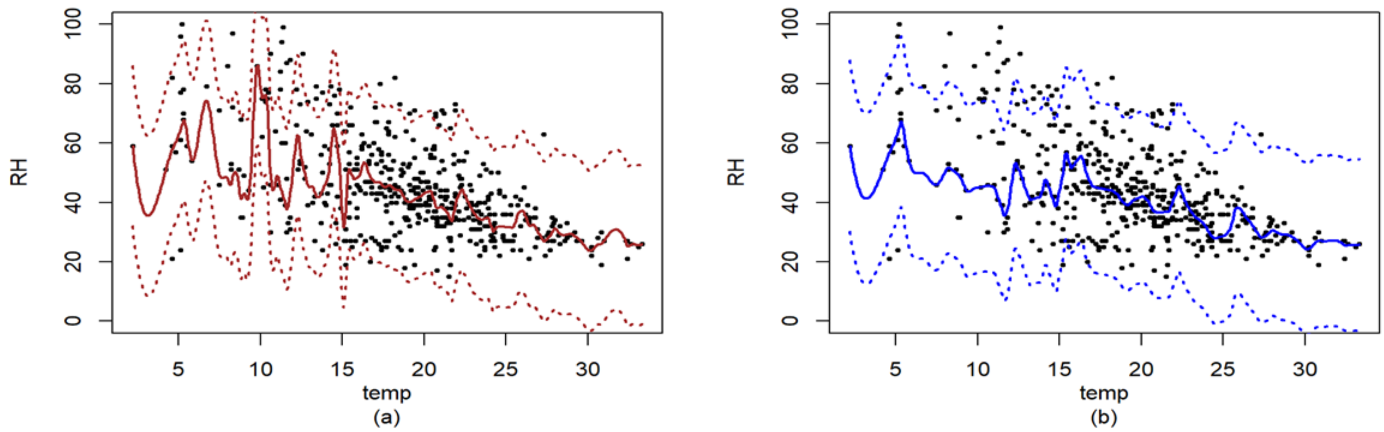

Figure 5: Scatter plot, fitted modal regression and 95% prediction bounds of the forest fire data: (a) KDE based bandwidth; (b) GH based bandwidth.

Example 3 (*Income distribution data*) In this study, we consider the 2017 China General Social Survey (CGSS) data, which is the first national, comprehensive, and continuous large-scale social survey project in China. We are intersted in how age, years of education and years of work are related to income (divided by 1000). After deleting the observations with missing values, the final sample consisted of 2,149 urban residents, including females aged 18-55 and males aged 18-60.

Due to the "curse of dimensionality", the nonparametric modal regression considered in this article is not appropriate for multivariate predictors. As a result, we apply the linear modal regression (Yao and Li, 2014). Similarly, we apply both the KDE and GH methods to select the bandwidth, and the results are denoted by LMR(KDE) and LMR(GH), respectively. The results are compared to mean regression (LSE) and median regression (MEDREG). 90% of the data is considered as training set, and the remaining 10% are used for prediction. The process is repeated over 500 repetitions. Table 5 and Table 6 report the average (std) of the coverage probabilities of prediction interals (of the same length), and average (std) of widths of the prediction intervals (at the same level of confidence). We can see that when the CI is of the same length, LMR(GH) could provide higher coverage probability. In the meanwhile, at the same desired confidence lever, the CI of LMR(GH) is the shortest, especially for higher confidence levels.

4 Concluding remarks

In this article, we propose a new bandwidth selection method for nonparametric modal regression based on the generalized hyperbolic distribution. With 5 free parameters, the GH is very general and have many special and limiting cases, like variance-gamma, hyperbolic, normal-verse Gaussian, t, and so on. The method can be easily implemented using any statistical software and is intuitively appealing. Also, unlike the plug-in type method, the new method does not require preliminary parameters to be chosen, and is desirable in real data applications. Simulation studies and real data examples show that, compared to the existing KDE based bandwidth selector, the new method can offer higher coverage probability for confidence intervals with the same width.

In this article, we only investigate the bandwidth selector for nonparametric modal regression analysis. It is also of great interest to extend our work to other nonparametric or semiparametric modal regression analysis tools.

Acknowledgement

Xiang's research is supported by Xiang's research is supported by the National Social Science Foundation of China (21CTJ010). Yao's research is supported by supported by an NSF grant DMS-2210272.

Appendix

Calculation details.

Hyperbolic distribution

The pdf is:

$$f(x) = \frac{1}{2\delta\sqrt{1+\pi^2}G_1(\xi)} e^{-\xi\left[\sqrt{1+\pi^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}-\pi\frac{x-\mu}{\delta}\right]},$$

where $G_\nu()$ is a modified Bessel function of the third kind. Then,

$$\begin{aligned}
f'(x) &= f(x)\left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2}\right) \\
&= \frac{1}{2\delta\sqrt{1+\pi^2}G_1(\xi)} e^{-\xi[\sqrt{1+\pi^2}} \sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2 - \pi\frac{x-\mu}{\delta}\Bigg]} \left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2}\right) \\
f''(x) &= f'(x)\left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2}\right) \\
&\quad + f(x)\left(\xi\sqrt{1+\pi^2}(x-\mu)^2\delta^{-4}\left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{3}{2}} - \xi\sqrt{1+\pi^2}\left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{1}{2}}\delta^{-2}\right) \\
&= \frac{1}{2\delta\sqrt{1+\pi^2}G_1(\xi)} e^{-\xi[\sqrt{1+\pi^2}} \sqrt{\sqrt{1+\frac{x-\mu}{\delta}\Bigg)^2} - \pi\frac{x-\mu}{\delta}\Bigg]} \left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2}\right) \\
&\quad \times \left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2}\right) \\
&\quad + \frac{1}{2\delta\sqrt{1+\pi^2}G_1(\xi)} e^{-\xi\left[\sqrt{1+\pi^2}\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}-\pi\frac{x-\mu}{\delta}\right]} \\
&\quad \times \left(\xi\sqrt{1+\pi^2}(x-\mu)^2\delta^{-4}\left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{3}{2}} - \xi\sqrt{1+\pi^2}\left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{1}{2}}\delta^{-2}\right)
\end{aligned}$$

$$I_1 \stackrel{\text{def}}{=} \left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2} \right)$$

$$I_2 \stackrel{\text{def}}{=} f(x) \left(\xi\sqrt{1+\pi^2}(x-\mu)^2\delta^{-4} \left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{3}{2}} - \xi\sqrt{1+\pi^2} \left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{1}{2}} \delta^{-2} \right)$$

$$A \stackrel{\text{def}}{=} \xi\sqrt{1+\pi^2}(x-\mu)^2\delta^{-4} \left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{3}{2}} - \xi\sqrt{1+\pi^2} \left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{1}{2}} \delta^{-2}$$

$$A' = 3\xi\sqrt{1+\pi^2}(x-\mu)^2\delta^{-4} \left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{3}{2}} - 3\xi\sqrt{1+\pi^2}(x-\mu)^3\delta^{-6} \left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{5}{2}}$$

$$I_1' = f''(x) \left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2} \right) + f'(x)A$$

$$I_2' = f'(x)A + f(x)A'$$

$$f'''(x) = I_1' + I_2' = f''(x) \left(\frac{\xi\pi}{\delta} - \frac{\delta\sqrt{1+\pi^2}(x-\mu)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\delta^2} \right) + 2f'(x)A + f(x)A'$$

Normal inverse Gaussian (NIG)

$$
\begin{aligned}
f(x) &= e^{\delta\sqrt{\alpha^2-\beta^2}} \frac{\alpha\delta}{\pi\sqrt{\delta^2+(x-\mu)^2}} G_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
f'(x) &= \alpha\beta\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-\frac{1}{2}} G_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- \alpha^2\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
f''(x) &= \alpha\beta^2\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-\frac{1}{2}} G_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- \alpha^2(\beta+1)\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&+ \alpha^3\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)^2\left(\delta^2+(x-\mu)^2\right)^{-\frac{3}{2}} G_3\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- \alpha^2\beta\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
f''(x) &= \alpha\beta^2\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-\frac{1}{2}} G_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- \alpha^2(\beta+1)\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&+ \alpha^3\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)^2\left(\delta^2+(x-\mu)^2\right)^{-\frac{3}{2}} G_3\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- \alpha^2\beta\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
f'''(x) &= \alpha\beta^3\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-\frac{1}{2}} G_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- 2\alpha^2\beta^2\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- 2\alpha^2\beta\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-1} G_2\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&+ \alpha^3(\beta+1)\delta\pi^{-1}e^{\sigma\sqrt{\alpha^2-\beta^2}}\left(\delta^2+(x-\mu)^2\right)^{-\frac{3}{2}} G_3\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&+ 2\alpha^3\beta\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)^2\left(\delta^2+(x-\mu)^2\right)^{-\frac{3}{2}} G_3\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&+ 2\alpha^3\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)\left(\delta^2+(x-\mu)^2\right)^{-\frac{3}{2}} G_3\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right) e^{\beta(x-\mu)} \\
&- \alpha^4\delta\pi^{-1}e^{\delta\sqrt{\alpha^2-\beta^2}}(x-\mu)^3\left(\delta^2+(x-\mu)^2\right)^{-2} G_4
\end{aligned}
$$

References

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logrithm of particle size. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Eng. Sci.*, 353:401–419.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, 5:151–157.

Browne, R. and McNicholas, P. (2015). A mixture of generalized hyperbolic distributions. *The Canadian Journal of Statistics*, 43:176–198.

Chen, X., Gao, W., and Li, Z. (2018). A data-driven bandwidth selection method for the smoothed maximum score estimator. *Economics Letters*, 170:24–26.

Choi, J., Du, Y., and Song, Q. (2021). Inverse gaussian quadrature and finite normal-mixture approximation of the generalized hyperbolic distribution. 388:113–302.

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

Gaunt, R. and Merkle, M. (2021). On bounds for the mode and median of the generalized hyperbolic and related distributions. 493:124–508.

Kemp, G. and Santos Silva, J. (2012). Regression towards the mode. *Journal of Economics*, 170:92–101.

Kirkby, J., Leitao, A., and Nguyen, D. (2021). Nonparametric density estimation and bandwidth selection with b-spline bases: A novel galerkin method. *Computational Statistics and Data Analysis*, 159:1–28.

Krief, J. M. (2017). Semi-linear mode regression. *Econometrics Journal*, 20:149–167.

Lee, M. (1989). Mode regression. *Journal of Econometrics*, 42:337–349.

Lee, M. (1993). Quadratic mode regression. *Journal of Econometrics*, 57:1–19.

Lee, M. and Kim, H. (1998). Semiparametric econometric estimators for a truncated regression model: a review with an extension. *Statistica Neerlandica*, 52:200–225.

Levine, M. (2013). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: a possible approach. *Computational Statistics and Data Analysis*, 67:1985–198.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press, Princeton.

Morris, K. and McNicholas, P. (2016). Clustering, classification, discriminant analysis, and dimension reducton via generalized hyperbolic mixtures. 97:133–150.

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

Shang, H. (2013). Bayestian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Computational Statistics and Data Analysis*, 67:185–198.

Sun, Y. and Li, Q. (2011). Data-driven bandwidth selection for nonstationary semiparametric models. *Journal of Business and Economic Statistics*, 29:541–551.

Ullah, A., Wang, T., and Yao, W. (2021). Modal regression for fixed dffects panel data. *Empirical Economics*, 60:261–308.

Ullah, A., Wang, T., and Yao, W. (2022a). Nonlinear modal regression for dependent data with application for predicting covid-19. *Journal of the Royal Statistical Society, Series A*.

Ullah, A., Wang, T., and Yao, W. (2022b). Semiparametric partially linear varying coefficient modal regression. *Journal of Econometrics*.

Xiang, S. and Yao, W. (2016). A new information criterion-based bandwidth selection method for non-parametric regressions. *Journal of Statistical Computation and Simulation*, 87:3446–3455.

Xiang, S. and Yao, W. (2018). Semiparametric mixtures of nonparametric regressions. *Annals of the Institute of Statistical Mathematics*, 70:131–154.

Xiang, S. and Yao, W. (2022). *Journal of Computational and Applied Mathematics*, 409:114–130.

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

Zhang, T., Kato, K., and Ruppert, D. (2021). Bootstrap inference for quantile based modal regression. *Journal of the American Statistical Association*.

Table 2: Average widths (percentage of coverage) of the prediction intervals.

Case	confidence	$n = 100$		$n = 200$		$n = 400$	
		LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)
I	30%	1.477(0.298)	1.491(0.296)	1.464(0.295)	1.474(0.296)	1.456(0.2975)	1.460(0.297)
	50%	2.624(0.497)	2.647(0.496)	2.606(0.497)	2.627(0.497)	2.587(0.499)	2.603(0.498)
	90%	6.829(0.897)	6.908(0.897)	6.793(0.899)	6.835(0.899)	6.745(0.899)	6.776(0.899)
II	30%	1.649(0.298)	1.670(0.297)	1.631(0.296)	1.650(0.296)	1.619(0.298)	1.629(0.298)
	50%	3.026(0.497)	3.053(0.498)	2.984(0.495)	3.016(0.495)	2.970(0.499)	2.996(0.499)
	90%	9.682(0.8987)	9.710(0.898)	9.653(0.896)	9.696(0.896)	9.603(0.897)	9.640(0.898)
III	30%	1.494(0.297)	1.431(0.297)	1.466(0.297)	1.441(0.296)	1.347(0.297)	1.344(0.297)
	50%	3.069(0.496)	2.940(0.497)	3.043(0.498)	3.005(0.498)	2.729(0.497)	2.726(0.496)
	90%	13.312(0.896)	13.158(0.896)	13.449(0.896)	13.413(0.896)	12.871(0.897)	12.862(0.897)
IV	30%	3.262(0.295)	3.265(0.296)	3.214(0.297)	3.235(0.296)	3.189(0.295)	3.239(0.297)
	50%	5.891(0.496)	5.922(0.497)	5.783(0.497)	5.821(0.498)	5.724(0.497)	5.833(0.497)
	90%	16.696(0.899)	16.904(0.898)	16.306(0.898)	16.511(0.898)	16.167(0.896)	16.657(0.896)
V	30%	2.185(0.297)	2.165(0.297)	2.128(0.299)	2.112(0.299)	2.074(0.299)	2.073(0.298)
	50%	4.191(0.497)	4.163(0.497)	4.085(0.497)	4.079(0.498)	4.061(0.498)	4.057(0.497)
	90%	14.100(0.897)	14.083(0.897)	13.981(0.897)	13.998(0.897)	13.907(0.897)	13.930(0.897)
VI	30%	2.810(0.297)	2.776(0.296)	3.531(0.299)	3.486(0.298)	3.452(0.297)	3.441(0.297)
	50%	5.500(0.497)	5.440(0.497)	7.058(0.498)	7.008(0.498)	6.891(0.497)	6.909(0.498)
	90%	16.283(0.896)	16.282(0.896)	16.717(0.899)	16.583(0.899)	16.538(0.897)	16.501(0.896)

Table 3: Average (Std) of percentage of coverage for the *air pollution data*

	MCCV $d = 30$		MCCV $d = 70$	
Width	LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)
5σ	0.148 (0.056)	0.155 (0.058)	0.142 (0.038)	0.161 (0.042)
6σ	0.165 (0.059)	0.179 (0.061)	0.166 (0.042)	0.186 (0.042)
7σ	0.188 (0.062)	0.204 (0.064)	0.186 (0.044)	0.210 (0.045)
8σ	0.209 (0.067)	0.224 (0.066)	0.208 (0.047)	0.231 (0.047)
9σ	0.229 (0.068)	0.244 (0.068)	0.230 (0.049)	0.250 (0.050)
10σ	0.250 (0.067)	0.264 (0.067)	0.251 (0.051)	0.268 (0.052)
	10-fold CV		5-fold CV	
Width	LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)
5σ	0.077 (0.078)	0.097 (0.082)	0.109 (0.042)	0.097 (0.041)
6σ	0.101 (0.080)	0.106 (0.086)	0.120 (0.055)	0.119 (0.044)
7σ	0.106 (0.082)	0.124 (0.101)	0.128 (0.046)	0.132 (0.031)
8σ	0.115 (0.082)	0.139 (0.104)	0.136 (0.039)	0.146 (0.043)
9σ	0.121 (0.082)	0.147 (0.106)	0.148 (0.034)	0.166 (0.048)
10σ	0.130 (0.090)	0.171 (0.119)	0.166 (0.042)	0.196 (0.083)

Table 4: Average (Std) of percentage of coverage for the *forest fire data*

	MCCV $d = 50$		MCCV $d = 100$	
Width	LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)
6σ	0.309 (0.039)	0.336 (0.042)	0.341 (0.035)	0.368 (0.038)
8σ	0.393 (0.044)	0.425 (0.049)	0.432 (0.038)	0.463 (0.041)
10σ	0.469 (0.046)	0.501 (0.050)	0.514 (0.039)	0.542 (0.041)
12σ	0.535 (0.047)	0.566 (0.051)	0.582 (0.038)	0.609 (0.039)
16σ	0.644 (0.047)	0.671 (0.047)	0.689 (0.034)	0.710 (0.030)
20σ	0.725 (0.044)	0.752 (0.044)	0.771 (0.031)	0.789 (0.027)
	10-fold CV		5-fold CV	
Width	LPMR(KDE)	LPMR(GH)	LPMR(KDE)	LPMR(GH)
6σ	0.381 (0.083)	0.412 (0.103)	0.359 (0.088)	0.390 (0.114)
8σ	0.506 (0.098)	0.500 (0.110)	0.468 (0.095)	0.477 (0.134)
10σ	0.586 (0.097)	0.586 (0.100)	0.545 (0.113)	0.560 (0.150)
12σ	0.632 (0.093)	0.646 (0.097)	0.596 (0.120)	0.624 (0.122)
16σ	0.722 (0.090)	0.732 (0.087)	0.699 (0.089)	0.718 (0.115)
20σ	0.792 (0.076)	0.812 (0.083)	0.787 (0.073)	0.793 (0.080)

Table 5: Average (Std) of percentage of coverage for the *income distribution data*

Width	LSE	MEDREG	LMR(KDE)	LMR(GH)
4σ	0.071(0.017)	0.116(0.022)	0.107(0.024)	0.133(0.022)
6σ	0.109(0.019)	0.158(0.024)	0.171(0.030)	0.194(0.026)
8σ	0.145(0.022)	0.200(0.027)	0.207(0.033)	0.250(0.031)
10σ	0.177(0.023)	0.263(0.029)	0.257(0.034)	0.299(0.029)
12σ	0.219(0.024)	0.313(0.029)	0.326(0.034)	0.349(0.030)
14σ	0.259(0.026)	0.362(0.030)	0.365(0.040)	0.416(0.029)

Table 6: Average(std) widths of the prediction intervals for the *income distribution data*.

Confidence	LSE	MEDREG	LMR(KDE)	LMR(GH)
10%	8.419(1.621)	7.048(1.661)	6.042(1.940)	5.403(1.365)
20%	18.027(2.417)	15.595(2.557)	13.002(2.249)	12.087(1.951)
30%	27.484(3.279)	23.155(2.741)	19.796(2.524)	18.441(2.211)
40%	37.414(3.521)	32.205(3.107)	28.256(3.235)	26.565(3.058)
50%	48.177(3.830)	43.224(4.607)	37.693(3.766)	35.229(3.371)
60%	59.807(4.408)	55.809(4.631)	48.797(4.906)	46.106(4.872)