Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(8): 5328–5351.
DOI: 10.3934/era.2026237
Received: 19 May 2026
Revised: 12 June 2026
Accepted: 17 June 2026
Published: 23 June 2026

Research article

Nonparametric modal regression based on orthogonal series estimation

Zhongzheng Wang and Xiaobing Zhao*

School of Data Sciences, Zhejiang University of Finance and Economics, Hangzhou 310018, China

* **Correspondence:** Email: maxbzhao@zufe.edu.cn.

Abstract: In this paper, we proposed an orthogonal series nonparametric modal regression (OS-NMR) method for estimating the conditional mode under a unimodal setting. The unknown modal regression function was approximated by a truncated expansion of orthogonal basis functions, which represents the nonparametric component through a finite-dimensional coefficient vector. These coefficients were estimated using a kernel-density-based modal regression criterion and computed by the modal expectation-maximization (MEM) algorithm. Compared with local polynomial modal regression, the proposed method provides a global approximation and avoids repeated pointwise estimation. The orthogonal structure further separates coefficient estimation error from truncation approximation error, which makes the theoretical arguments clearer and more tractable. We established the convergence rate and asymptotic normality of the proposed estimator, and evaluated its finite-sample performance through simulations and a real data application.

Keywords: orthogonal series expansion; modal regression; MEM algorithm; kernel density estimation; mode

1. Introduction

Modal regression [1–3] has become an important tool for studying conditional distributions through their most probable values. Unlike mean regression and quantile regression, it does not summarize the conditional distribution by an average level or by a prescribed probability level. Instead, it directly targets the point around which the conditional density is most concentrated. This feature is particularly useful for skewed, heavy-tailed, or multimodal data [4, 5], where summaries based on the conditional mean or conditional quantiles may fail to represent the main structure of the data, or may even lead to misleading conclusions. In this sense, the conditional mode provides a robust and interpretable way to describe the relationship between variables in complex data. Mode-based ideas have also found applications in many empirical areas. For example, modal regression and related methods have been applied to wage analysis [6], transportation and traffic-flow studies [7, 8], public health and epidemiological

forecasting [9, 10], environmental risk analysis such as forest-fire data [3, 11], and economic clustering or macroeconomic applications [12, 13].

Although modal regression has received increasing attention, its development is still less extensive than that of mean regression and quantile regression. Early studies mainly focused on linear or parametric modal regression models, including the works of Lee [1, 14], Kemp and Santos Silva [15], Yao and Li [3], Zhou and Huang [16], Chen et al. [4], Li and Huang [17], and Jing and Guo [10], among others. These methods provide useful tools for mode-based inference, but a linear modal regression model imposes a fixed parametric structure on the relationship between the conditional mode and the covariates. In practice, such a structure may be too restrictive, especially when the most likely response changes with covariates in a nonlinear way.

To relax the parametric restriction, several nonparametric and semiparametric modal regression methods have been proposed. For example, Yao et al. [2] and Xiang and Yao [18] developed local polynomial modal regression, which estimates the conditional mode through local polynomial approximation. Ota et al. [19] studied modal regression from the perspective of the derivative of the quantile function, while Feng et al. [20] formulated nonparametric modal regression through empirical risk minimization. Semiparametric extensions have also been considered. Kreief [21] investigated partial linear modal regression and established consistency and asymptotic normality of the proposed estimator. Ullah et al. [9, 22] applied modal regression to fixed-effects panel data and time series settings, respectively, and Ullah et al. [23] further developed a partially linear varying-coefficient modal regression model. More recently, Wang [6] proposed a parametrically guided nonparametric estimator, Han and Zhao [13] studied nonparametric modal regression with mixed discrete and continuous covariates, and Yuan et al. [11] introduced a generalized hyperbolic distribution-based bandwidth selection method. Wang and Yao [24] extended nonparametric mode-oriented regression to spatially dependent data and developed asymptotic theory under general mixing conditions. Wang and Yao [25] proposed an online kernel-based mode learning method for large scale data, aiming to improve robustness against outliers and heavy-tailed errors while reducing storage and computational costs. Xiang et al. [26] provided a systematic review of recent developments in modal regression, covering theoretical foundations, computational algorithms, and practical implementations.

These studies have greatly enriched the methodology of modal regression. Even so, there is still room for improvement in how the unknown modal function is approximated and estimated. Local polynomial modal regression [2, 18] is flexible and widely used, but it is essentially a pointwise method. As a result, the regression function needs to be estimated repeatedly at different design points, which can be computationally demanding. Moreover, its performance is closely tied to bandwidth selection. Spline-based modal regression, such as the B-spline method proposed by Yang et al. [8], offers a more global approximation and can estimate the regression function in one step. However, spline methods usually depend on knot selection, and the resulting basis functions are not generally orthogonal, which may make the theoretical decomposition of estimation error less transparent.

Motivated by these considerations, we propose a nonparametric modal regression method based on orthogonal series estimation, which we refer to as orthogonal series nonparametric modal regression (OSNMR). The basic idea is to approximate the unknown modal regression function by a truncated expansion of orthogonal basis functions. In this way, the infinite-dimensional nonparametric problem is converted into the estimation of a finite-dimensional coefficient vector, while the remaining part is treated as a truncation error. This formulation has two main advantages. First, it provides a global estimator of

the modal regression function, avoiding repeated pointwise estimation as in local polynomial methods. Second, the orthogonality of the basis functions leads to a clearer separation between the coefficient estimation error and the approximation error caused by truncation. This makes the theoretical analysis more transparent and allows the true modal function to belong to a general smooth function class, rather than being treated as if it were exactly represented by a fixed finite-dimensional basis.

For numerical implementation, we develop an estimation procedure based on the modal expectation-maximization (MEM) algorithm. The proposed method combines the flexibility of nonparametric modal regression with the structural simplicity of orthogonal series approximation. Since the present study focuses on the unimodal setting, the conditional density is assumed to have a unique mode, so that the modal regression function can be treated as a well-defined single-valued function. Within this framework, our method retains the robustness of modal regression for skewed and heavy-tailed data, while offering better computational convenience and clearer theoretical analysis than purely local smoothing approaches. We further establish the asymptotic properties of the proposed estimator, including its convergence rate and asymptotic normality, and evaluate its finite-sample performance through simulation studies and a real data application.

The remainder of this paper is organized as follows. Section 2 introduces the nonparametric modal regression model and the orthogonal series approximation. Section 3 proposes an estimation procedure based on the MEM algorithm and discusses bandwidth selection. Section 4 establishes the main asymptotic properties of the proposed estimator, including the convergence rate and asymptotic normality. Sections 5 and 6 report simulation studies and a real data application, respectively. Section 7 concludes the paper with several remarks. All technical proofs are collected in Appendix A.

2. Modal regression and orthogonal series approximation

Suppose that, given a predictor $X = x$, the response variable Y has a conditional density $f(y \mid x)$. Throughout this paper, we focus on the unimodal case and assume that the conditional density has a unique mode. The conditional mode is then defined as $\text{Mode}(Y \mid X = x) = \arg\max_{y\in\mathbb{R}} f(y \mid x)$. We denote the corresponding modal regression function by $m(x) = \text{Mode}(Y \mid X = x)$. Given an independent sample $\{(y_i, x_i), i = 1, \ldots, n\}$, our goal is to estimate the unknown function $m(\cdot)$. For simplicity, we consider the case where X is a scalar predictor. The nonparametric modal regression model can be written as

$$y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where ϵ_i denotes the error term, which is independent of X. Since no parametric form is imposed on $m(\cdot)$, model (2.1) is a nonparametric regression model.

A key assumption in modal regression is that the error term has a conditional mode of zero. More precisely, for any x in the support of X, we assume that $\text{Mode}(\epsilon \mid X = x) = \arg\max_{t\in\mathbb{R}} g_{\epsilon|X}(t \mid x) = 0$, where $g_{\epsilon|X}(t \mid x)$ denotes the conditional density of ϵ given $X = x$. Under this zero-mode error assumption, the conditional mode of Y given $X = x$ is exactly $m(x)$. We further assume that $g_{\epsilon|X}(t \mid x)$ is continuous and bounded in $t \in \mathbb{R}$ for each x. It is worth noting that we do not require the error distribution to be homogeneous, symmetric, or normal. Thus, heteroscedastic errors and skewed error distributions are allowed in the present framework.

Before introducing the series estimator, we first specify the function space in which the unknown modal regression function is approximated. The choice of this space is closely related to the support of

the covariate. For example, if the covariate is supported on a bounded interval, such as $[0, 1]$ or $[-1, 1]$, it is natural to work with the usual L^2 space on that interval. If the covariate has unbounded support, one may instead consider an L^2 space on the real line, or a weighted L^2 space.

In this paper, we present the method in a general Hilbert space framework. Suppose that the modal regression function $m(\cdot)$ belongs to $L^2(\mathcal{X})$, where $\mathcal{X}$ denotes the support of X. Let $\{\psi_j(x)\}_{j=0}^{\infty}$ be an orthonormal basis of $L^2(\mathcal{X})$. Then $m(\cdot)$ admits the orthogonal expansion

$$m(x) = \sum_{j=0}^{\infty} \beta_j \psi_j(x),$$

where $\beta_j = \int_{\mathcal{X}} m(x)\psi_j(x)\,dx$. Here the integral should be understood with respect to the underlying measure of the chosen L^2 space. For instance, in a weighted L^2 space, the corresponding weight function is included in the inner product.

Remark 1. *The basis $\{\psi_j(x)\}_{j=0}^{\infty}$ is kept general in the theoretical development [27]. This is because the proposed method does not depend on one particular basis, but rather on the orthogonal series representation itself. Different choices of orthonormal bases may be convenient for different supports of the covariate. For bounded intervals, Legendre polynomials, a trigonometric basis, or other orthogonal polynomial systems can be used after a suitable rescaling of the covariate. For unbounded domains, a weighted orthogonal basis, such as a Hermite-type basis, may be more appropriate. In the numerical studies of this paper, we use the normalized Legendre polynomial basis because the covariate is considered on a bounded interval and this basis provides a simple and stable implementation.*

For a positive integer k, we approximate $m(\cdot)$ by the truncated series $m_k(x) = \sum_{j=0}^{k-1} \beta_j \psi_j(x)$. The truncation error is defined as $\gamma_k(x) = \sum_{j=k}^{\infty} \beta_j \psi_j(x)$, so that $m(x) = m_k(x) + \gamma_k(x)$. The parameter k controls the complexity of the series approximation. A smaller k gives a smoother approximation but may lead to larger approximation bias, whereas a larger k allows more flexibility but may increase estimation variability.

Substituting this decomposition into model (2.1), we obtain

$$y_i = m_k(x_i) + \gamma_k(x_i) + \epsilon_i, \qquad i = 1, \ldots, n. \tag{2.2}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{k-1})^\top$, $\boldsymbol{\gamma} = (\gamma_k(x_1), \ldots, \gamma_k(x_n))^\top$. Define $\Psi_k(x) = (\psi_0(x), \ldots, \psi_{k-1}(x))^\top$ and $\boldsymbol{\Psi} = (\Psi_k(x_1), \ldots, \Psi_k(x_n))^\top$, where $\boldsymbol{\Psi}$ is an $n \times k$ design matrix. Then the truncated modal regression function can be written as $m_k(x) = \Psi_k(x)^\top \boldsymbol{\beta}$.

In practice, the truncation error $\gamma_k(\cdot)$ is not directly estimated and is treated as an approximation error. The proposed orthogonal series nonparametric modal regression estimator is obtained by maximizing the kernel-density-based modal objective function

$$Q(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \phi_h \left(y_i - \Psi_k(x_i)^\top \boldsymbol{\beta} \right), \tag{2.3}$$

where $\phi_h(a) = \frac{1}{h}\phi\left(\frac{a}{h}\right)$, $h > 0$ is a bandwidth parameter, and $\phi(\cdot)$ is a kernel density function symmetric about zero. For computational convenience, we take $\phi(\cdot)$ to be the standard normal density in the implementation.

Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_{k-1})^\top$ be the maximizer of $Q(\boldsymbol{\beta})$. The proposed estimator of the modal regression function is then defined as $\widehat{m}_k(x) = \Psi_k(x)^\top \widehat{\boldsymbol{\beta}}$.

3. Estimation algorithm and hyperparameter selection

In this section, we describe the numerical procedure for estimating the coefficient vector in the proposed orthogonal series nonparametric modal regression model. We first present the MEM algorithm for maximizing the kernel-density-based modal objective function. We then discuss the selection of tuning parameters, including the bandwidth parameter h and the truncation parameter k.

3.1. The MEM algorithm

In this subsection, we use the MEM algorithm to compute the proposed estimator. The MEM algorithm and its convergence properties have been discussed in Yao and Li [3] and Li et al. [28]. Starting from an initial value $\boldsymbol{\beta}^{(0)} = \left(\beta_0^{(0)}, \beta_1^{(0)}, \cdots, \beta_{k-1}^{(0)}\right)^\top$, the algorithm alternates between the following E step and M step until convergence.

E step: Given the current value $\boldsymbol{\beta}^{(r)}$, compute the weight assigned to the ith observation, for $i = 1, 2, \cdots, n$, as

$$\pi\left(i \mid \boldsymbol{\beta}^{(r)}\right) = \frac{\phi_h\left(y_i - \Psi_k(x_i)^\top \boldsymbol{\beta}^{(r)}\right)}{\sum_{j=1}^n \phi_h\left(y_j - \Psi_k(x_j)^\top \boldsymbol{\beta}^{(r)}\right)} \propto \phi_h\left(y_i - \Psi_k(x_i)^\top \boldsymbol{\beta}^{(r)}\right).$$

M step: Update $\boldsymbol{\beta}^{(r+1)}$ by

$$\boldsymbol{\beta}^{(r+1)} = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^n \pi\left(i \mid \boldsymbol{\beta}^{(r)}\right) \ln \phi_h\left(y_i - \Psi_k(x_i)^\top \boldsymbol{\beta}\right).$$

When $\phi(\cdot)$ is taken to be the Gaussian kernel, the M step has the closed-form update

$$\boldsymbol{\beta}^{(r+1)} = \left(\boldsymbol{\Psi}^\top \boldsymbol{W}_r \boldsymbol{\Psi}\right)^{-1} \boldsymbol{\Psi}^\top \boldsymbol{W}_r \boldsymbol{y},$$

where $\boldsymbol{W}_r = \operatorname{diag}\left\{\pi\left(1 \mid \boldsymbol{\beta}^{(r)}\right), \pi\left(2 \mid \boldsymbol{\beta}^{(r)}\right), \ldots, \pi\left(n \mid \boldsymbol{\beta}^{(r)}\right)\right\}$.

3.2. Selection of hyperparameters

The proposed OSNMR estimator involves two tuning parameters: the bandwidth h in the kernel-density-based modal objective function and the truncation parameter k in the orthogonal series approximation. These two parameters play different roles, but both have a direct effect on the finite-sample performance of the estimator.

The bandwidth h controls the amount of smoothing in the kernel-density-based modal objective function. A smaller bandwidth makes the objective function more sensitive to local features of the residual distribution, but it may also increase estimation variance and cause numerical instability. In contrast, a larger bandwidth gives a smoother objective function, but it may oversmooth the residual density and lead to a larger estimation bias. Therefore, the choice of h is important for the practical performance of the proposed estimator. The truncation parameter k determines the number of basis functions used to approximate the unknown modal regression function. A smaller k gives a smoother and more stable approximation, but may lead to a larger approximation bias if the true modal function has a complex shape. A larger k allows more flexibility, but may increase estimation variability and computational cost. Thus, the selection of h and k reflects the usual trade-off between bias and variance.

It is worth noting that least-squares leave-one-out cross-validation, which is commonly used in mean regression, may not be well-suited to modal regression [29]. The reason is that modal regression targets the most probable value of the conditional distribution, rather than the conditional mean. Following this idea, Yang et al. [8] introduced a cross-validation criterion designed for modal regression. The basic principle is simple: a good modal regression estimator should produce a fixed-width prediction interval that contains as many observed responses as possible.

Let $\mathrm{I}(\cdot)$ denote the indicator function and let $c > 0$ be a fixed constant controlling the width of the prediction interval. For each candidate pair (h, k), we compute the corresponding leave-one-out estimator $\hat{m}_{k,h}^{(-i)}(\cdot)$, where $\hat{m}_{k,h}^{(-i)}(\cdot)$ denotes the estimator obtained after removing the ith observation. The hyperparameters are then selected by maximizing

$$\mathrm{CV}(h,k) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\left(\frac{\left|y_i - \hat{m}_{k,h}^{(-i)}(x_i)\right|}{c} \le 1\right).$$

Equivalently, this criterion counts the proportion of observations falling inside the interval $\left[\hat{m}_{k,h}^{(-i)}(x_i) - c,\ \hat{m}_{k,h}^{(-i)}(x_i) + c\right]$. The selected tuning parameters are given by $(\hat{h}, \hat{k}) = \arg\max_{(h,k)\in\mathcal{H}\times\mathcal{K}} \mathrm{CV}(h,k)$, where $\mathcal{H}$ and $\mathcal{K}$ are finite candidate sets for h and k, respectively.

This criterion is more consistent with the objective of modal regression than a squared-error criterion, because it focuses on concentration around the fitted modal curve rather than average prediction error. It is also relatively robust to outlying observations. In the empirical implementation, the constant c is taken as a fixed percentage of the empirical range of the response variable, that is, $c = C(\max(Y) - \min(Y))$, where C is a prespecified percentage, and a grid search over $\mathcal{H} \times \mathcal{K}$ is used to identify the selected pair $(\hat{h}, \hat{k})$.

4. Asymptotic properties

To establish the asymptotic properties of the proposed modal estimator, we impose the following regularity conditions. These conditions are standard in the literature on modal regression and nonparametric series estimation, and are generally satisfied in many common settings; see, for example, Kemp and Santos Silva [15], Yao and Li [3], Ullah et al. [9, 22, 23],Wang [6], and Dong and Gao [27].

Condition 1. *The observations $\{(Y_i, X_i), i = 1, \ldots, n\}$ are independent and identically distributed (i.i.d.), and X has a compact support $\mathcal{X}$.*

Condition 2. *Let $\epsilon = Y - m(X)$ and let $g(t \mid x)$ be the conditional density of ϵ given $X = x$. For every $x \in \mathcal{X}$, $g(t \mid x)$ has a unique global maximum at $t = 0$. Moreover, $g(t \mid x)$ is four times continuously differentiable in a neighborhood of zero, $g'(0 \mid x) = 0$, and $g''(0 \mid x) < 0$ uniformly in x.*

Condition 3. *The kernel function $\phi(\cdot)$ is a symmetric density with bounded derivatives up to order three, and satisfies $\int u\phi(u)du = 0$, $\int u^2\phi(u)du < \infty$, and $\int\{\phi'(u)\}^2 du < \infty$.*

Condition 4. *The modal regression function admits the orthogonal expansion $m(x) = \sum_{j=0}^{\infty}\beta_j\psi_j(x)$. For $m_k(x) = \sum_{j=0}^{k-1}\beta_j\psi_j(x)$ and $\gamma_k(x) = m(x) - m_k(x)$, we have $\|\gamma_k\|_{L^2} = O(k^{-s})$ for some $s > 0$. For pointwise results, $\sup_{x\in\mathcal{X}}|\gamma_k(x)| = O(a_k)$ with $a_k \to 0$.*

Condition 5. *Let $\Psi_k(x) = (\psi_0(x), \ldots, \psi_{k-1}(x))^\top$ and $\zeta_k = \sup_{x\in\mathcal{X}}\|\Psi_k(x)\|$. The eigenvalues of $E\{\Psi_k(X)\Psi_k(X)^\top\}$ are bounded away from zero and infinity.*

Condition 6. *As* $n \to \infty$, $h \to 0$, $k \to \infty$, *and* $\frac{k}{nh^3} \to 0$, $\frac{\zeta_k^2 \log n}{nh^3} \to 0$.

A few comments on these conditions are useful. Condition 1 is a standard sampling condition. The compact support assumption is also natural here, since the proposed series estimator is developed on a bounded interval. Condition 2 is the key identification condition in modal regression. It states that the conditional error density has a unique mode at zero, so that the conditional mode of Y given $X = x$ is indeed $m(x)$. The smoothness and curvature requirements in this condition allow the kernel-based objective function to be expanded around the mode. Condition 3 imposes regularity on the kernel function. These requirements are mild and are satisfied by many commonly used kernels, including the Gaussian kernel used in our numerical studies. Condition 4 is the approximation condition for the orthogonal series expansion. It means that the unknown modal regression function can be well-approximated by the truncated series with k basis functions, while the remaining tail part $\gamma_k(x) = m(x) - m_k(x)$ becomes small as k increases. From a practical point of view, k controls the complexity of the fitted curve. A small k gives a smoother but possibly biased approximation, whereas a large k allows more flexibility but may also introduce additional variability. The term k^{-s} in Theorem 1 corresponds to this truncation error. Condition 5 is a stability condition for the chosen basis system. The matrix $E\{\Psi_k(X)\Psi_k(X)^\top\}$ plays the role of a population design matrix for the truncated series with k basis functions. Requiring its eigenvalues to be bounded away from zero and infinity avoids severe collinearity among the basis functions and ensures that the finite-dimensional coefficient vector is identifiable. The quantity $\zeta_k = \sup_{x\in\mathcal{X}} \|\Psi_k(x)\|$ measures the maximal size of the basis vector. For commonly used orthonormal bases on compact intervals, such as the normalized Legendre basis used in our simulations, this quantity grows at a controlled rate. Condition 6 describes how the smoothing bandwidth h and the truncation parameter k should be balanced as the sample size increases. The requirement $h \to 0$ makes the kernel-based modal criterion increasingly local around the conditional mode. The requirement $k \to \infty$ allows the series approximation space to become richer, so that the truncation bias can vanish. At the same time, k cannot grow too fast relative to the effective sample size in the modal objective. The condition $k/(nh^3) \to 0$ controls the estimation error of the k series coefficients, while $\zeta_k^2 \log n/(nh^3) \to 0$ strengthens this requirement uniformly over the covariate space.

Theorem 1 (Consistency of the OSNMR estimator)**.** *Suppose that Conditions 1–6 hold. Then the proposed OSNMR estimator satisfies*

$$\|\hat{m}_k - m\|_{L^2} = O_p\left(\sqrt{\frac{k}{nh^3}} + h^2 + k^{-s}\right).$$

Theorem 2 (Asymptotic normality)**.** *Suppose that Conditions 1–6 hold. For any fixed interior point* $x \in \mathcal{X}$,

$$\frac{\sqrt{nh^3}\,\{\hat{m}_k(x) - m(x) - B_{k,h}(x)\}}{\left[\Psi_k(x)^\top \boldsymbol{H}_k^{-1}\boldsymbol{V}_k\boldsymbol{H}_k^{-1}\Psi_k(x)\right]^{1/2}} \xrightarrow{d} N(0,1).$$

If, in addition, the undersmoothing condition $\sqrt{nh^3}\{h^2 + a_k\} \to 0$ *holds, then the bias term is asymptotically negligible and*

$$\frac{\sqrt{nh^3}\,\{\hat{m}_k(x) - m(x)\}}{\left[\Psi_k(x)^\top \boldsymbol{H}_k^{-1}\boldsymbol{V}_k\boldsymbol{H}_k^{-1}\Psi_k(x)\right]^{1/2}} \xrightarrow{d} N(0,1).$$

All details of the notation and proofs are provided in Appendix A.

5. Simulation

5.1. Numerical study design

In this section, we conduct simulation studies to evaluate the finite-sample performance of the proposed orthogonal series nonparametric modal regression estimator (OSNMR). We compare OSNMR with the B-spline modal regression estimator (BMR) [8] and the local polynomial modal regression estimator (LPM) [2]. The comparison focuses mainly on estimation accuracy and computational efficiency.

For the proposed OSNMR method, we use the normalized Legendre orthonormal basis in the series approximation. The covariate is generated independently from $X \sim U[-1, 1]$, so the Legendre basis is a natural choice for this bounded support. In order to focus on the effect of different modal regression functions, we fix the error distribution throughout this simulation study as $\epsilon \sim 0.1N(-1, 2.5^2) + 0.9N(1, 0.5^2)$. Let $b = \text{Mode}(\epsilon)$ denote the mode of this mixture distribution.

For each simulated dataset, the response is generated from

$$y_i = m(x_i) + \epsilon_i, \qquad i = 1, 2, \ldots, n.$$

Since the error distribution has mode b, the true conditional mode of Y given $X = x$ is $\text{Mode}(Y \mid X = x) = m(x) + b$. Thus, the target modal regression function in the simulation is denoted by $m_0(x) = m(x) + b$.

We consider three choices of $m(x)$, ranging from a relatively simple smooth function to a more oscillatory nonlinear function:

- Case 1: $m(x) = 2\sin(\pi x)$, where the true modal regression function is $m_0(x) = 2\sin(\pi x) + b$.
- Case 2: $m(x) = x^2 \sin(x)$, where the true modal regression function is $m_0(x) = x^2 \sin(x) + b$.
- Case 3: $m(x) = 2x^2 \sin(2\pi x)$, where the true modal regression function is $m_0(x) = 2x^2 \sin(2\pi x) + b$.

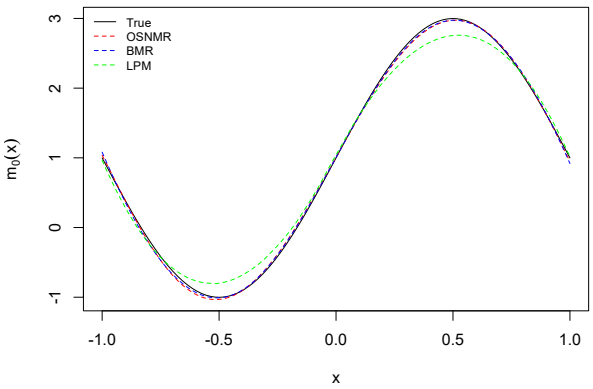
For all cases, we use Gaussian kernel function, defined by $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$. The hyperparameters are selected by the cross-validation criterion described in Section 3.2. For OSNMR, both the bandwidth h and the truncation parameter k are selected from finite candidate sets. For BMR and LPM, the corresponding smoothing parameters are also selected using the same modal cross-validation idea, so that the comparison is based on a common tuning strategy. The constant c in the cross-validation criterion is defined as $c = C(\max(Y) - \min(Y))$, where C is taken as 5% in each simulated dataset.

5.2. Performance measures and simulation results

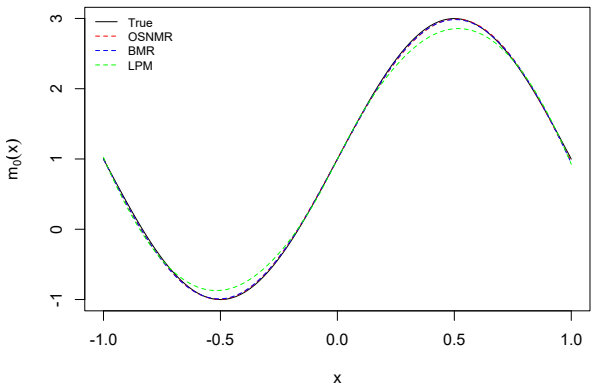
We consider sample sizes $n \in \{100, 200, 400\}$, and each experiment is repeated $R = 200$ times. Let $\widehat{m}^{(r)}(x)$ denote the estimator obtained in the rth replication. To evaluate the estimation accuracy of the regression function, we compute the mean squared error and standard deviation of the estimated function at a grid of points.

For a fixed evaluation point x_l, the pointwise mean squared error is defined as $\text{MSE}_l = \frac{1}{R}\sum_{r=1}^{R}\left[\widehat{m}^{(r)}(x_l) - m_0(x_l)\right]^2$, where $m_0(x)$ denotes the true modal regression function. The Monte Carlo average of the estimator is $\overline{\widehat{m}}(x_l) = \frac{1}{R}\sum_{r=1}^{R}\widehat{m}^{(r)}(x_l)$, and the corresponding standard deviation is $SD_l = \left[\frac{1}{R}\sum_{r=1}^{R}\left\{\widehat{m}^{(r)}(x_l) - \overline{\widehat{m}}(x_l)\right\}^2\right]^{1/2}$. After that, we compute the average mean squared error, $\text{MSE} = \frac{1}{L}\sum_{l=1}^{L}\text{MSE}_l$, and the average standard deviation, $\text{SD} = \frac{1}{L}\sum_{l=1}^{L} SD_l$. Here, L denotes

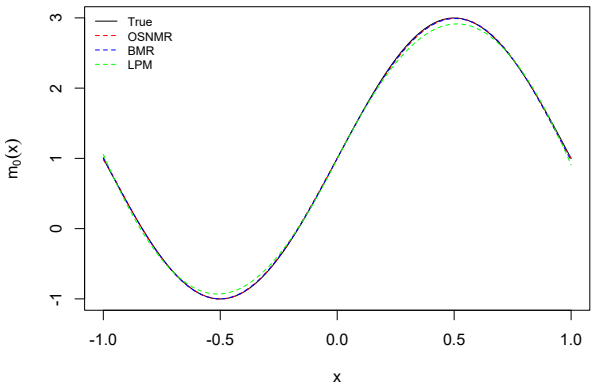
the number of evaluation points selected over the empirical range of the covariate, and in our simulations, we take $L = 100$. We also report empirical coverage probabilities at selected evaluation points. For each fixed point x_l, an empirical interval is constructed using the Monte Carlo standard deviation as $\widehat{m}^{(r)}(x_l) \pm 1.96SD_l$. The empirical coverage probability is then computed as $\mathrm{CP}_l = \frac{1}{R}\sum_{r=1}^{R} I\left[m_0(x_l) \in \left(\widehat{m}^{(r)}(x_l) - 1.96SD_l,\ \widehat{m}^{(r)}(x_l) + 1.96SD_l\right)\right]$, where $I(\cdot)$ is the indicator function. This coverage probability is used as a finite-sample empirical measure based on Monte Carlo replications.



(a) $n = 100$



(b) $n = 200$

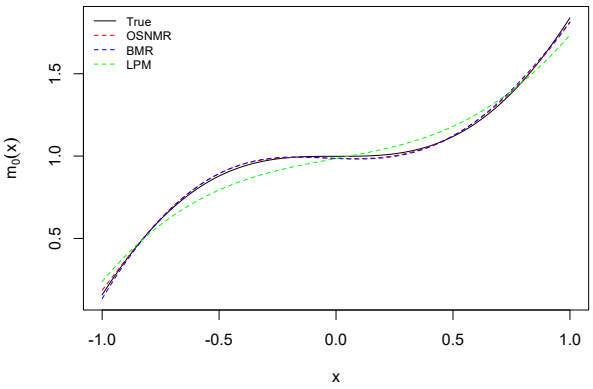


(c) $n = 400$

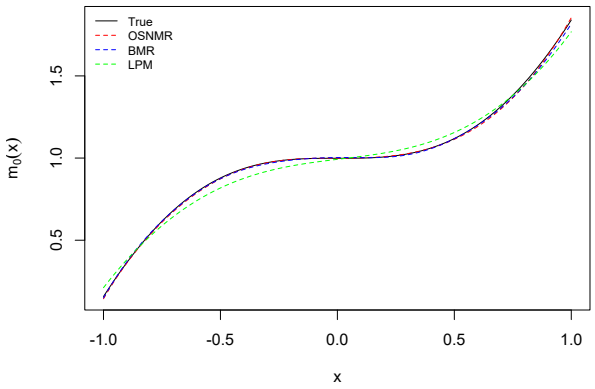
Figure 1. Estimated modal regression functions for Case 1 under different sample sizes.

The detailed simulation results are summarized in Tables 1 and 2, and the fitted curves are displayed in Figures 1–3. Overall, the three methods are able to recover the main shape of the true modal regression function. As the sample size increases from $n = 100$ to $n = 400$, the fitted curves become closer to the true curve, and the MSE values in Table 1 decrease steadily in all three cases. This pattern gives empirical support to the consistency of the proposed estimator.

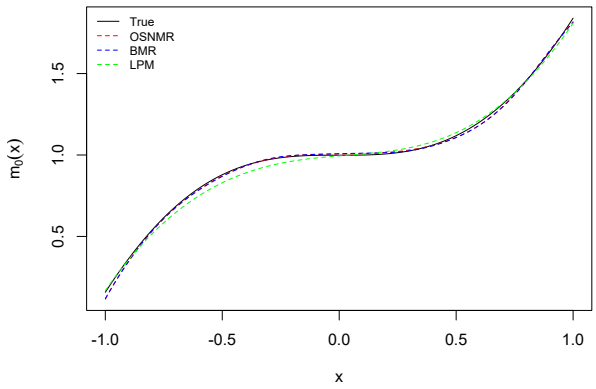
For Case 1, all three methods perform well, and the differences among them are relatively small. LPM gives slightly smaller MSE values, especially when the sample size becomes large, while OSNMR



(a) $n = 100$



(b) $n = 200$

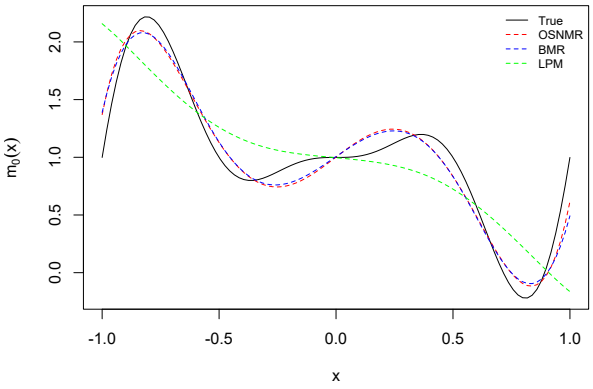


(c) $n = 400$

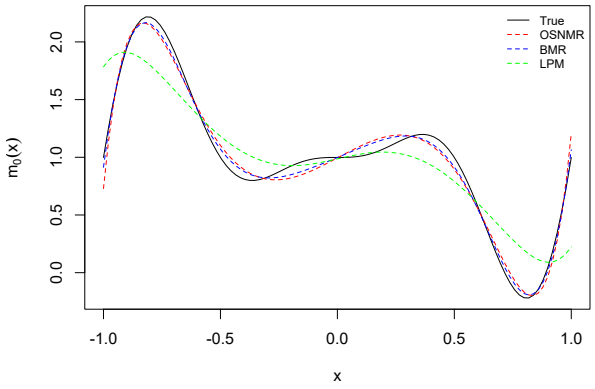
Figure 2. Estimated modal regression functions for Case 2 under different sample sizes.

Table 1. Average MSE and SD in parentheses for the three estimators under different cases and sample sizes.

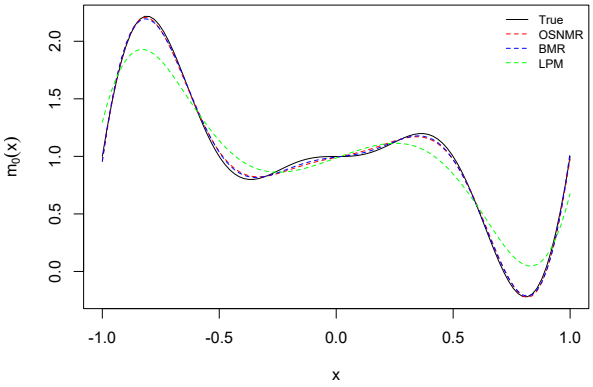
Case	n	OSNMR	BMR	LPM
Case 1	100	0.0464 (0.1955)	0.0456 (0.1998)	0.0455 (0.1575)
	200	0.0252 (0.1459)	0.0240 (0.1466)	0.0188 (0.1063)
	400	0.0130 (0.1054)	0.0127 (0.1068)	0.0080 (0.0727)
Case 2	100	0.0581 (0.2210)	0.0526 (0.2138)	0.0160 (0.1084)
	200	0.0291 (0.1592)	0.0298 (0.1612)	0.0090 (0.0831)
	400	0.0194 (0.1209)	0.0164 (0.1198)	0.0051 (0.0641)
Case 3	100	0.1065 (0.2590)	0.1097 (0.2645)	0.1421 (0.1509)
	200	0.0459 (0.1752)	0.0368 (0.1731)	0.0802 (0.1117)
	400	0.0208 (0.1296)	0.0190 (0.1296)	0.0297 (0.0772)



(a) $n = 100$

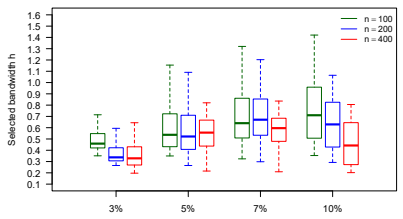


(b) $n = 200$

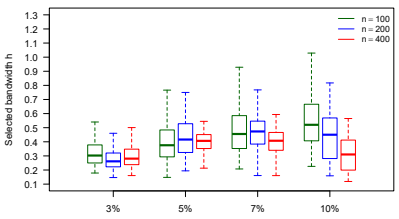


(c) $n = 400$

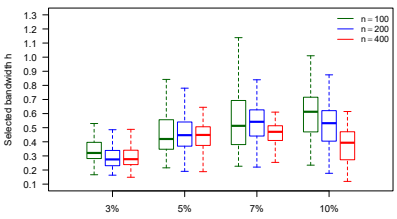
Figure 3. Estimated modal regression functions for Case 3 under different sample sizes.



(a) Case 1

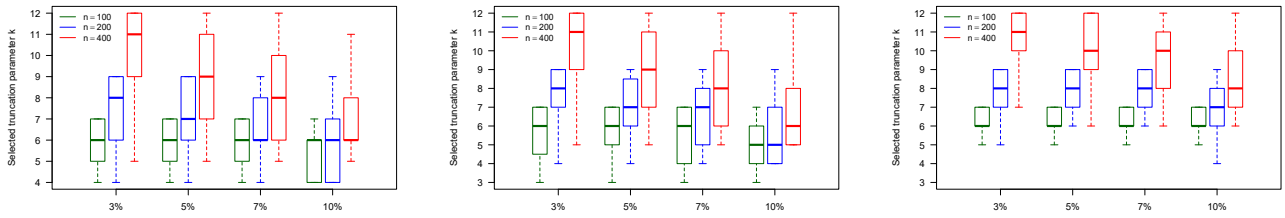


(b) Case 2



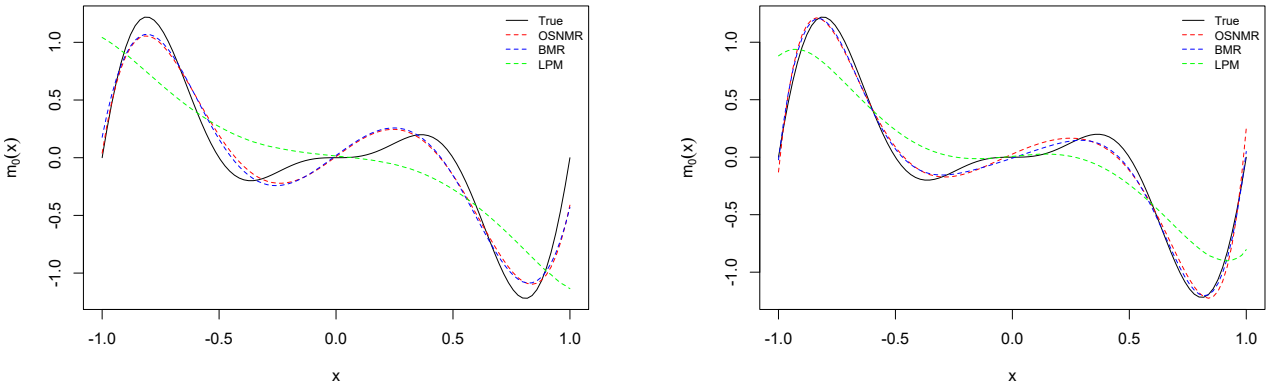
(c) Case 3

Figure 4. Sensitivity analysis of the selected bandwidth h for OSNMR under different sample sizes.

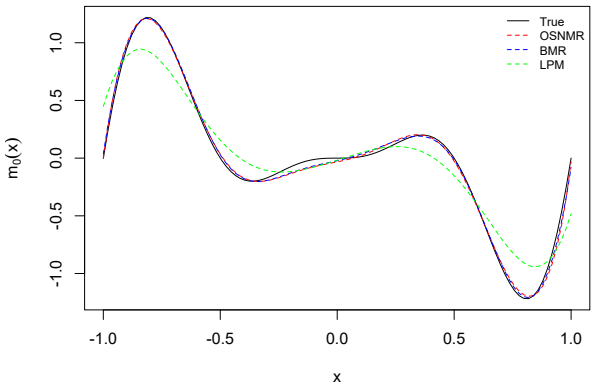


(a) Case 1 **(b)** Case 2 **(c)** Case 3

Figure 5. Sensitivity analysis of the selected truncation parameter k for OSNMR under different sample sizes.



(a) $n = 100$ **(b)** $n = 200$



(c) $n = 400$

Figure 6. Estimated modal regression functions for Case 3 under different sample sizes with a heavy-tailed distribution.

Table 2. Average computing time (seconds) for the three estimators under different cases and sample sizes.

Case	n	OSNMR	BMR	LPM
Case 1	100	0.0523	0.0616	1.6147
	200	0.0706	0.1043	3.0915
	400	0.2049	0.2735	5.9773
Case 2	100	0.0589	0.0910	0.3899
	200	0.1184	0.1670	0.8108
	400	0.3439	0.4442	1.8527
Case 3	100	0.0641	0.1004	0.5199
	200	0.1230	0.1742	1.1187
	400	0.3294	0.4286	2.7736

Note: Computations were performed in R version 4.4.2 on Windows 11 with an Intel Core i7-12700K CPU and 64 GB RAM.

Table 3. Empirical coverage probabilities of 95% CIs for OSNMR and BMR.

	Case 1			Case 2			Case 3		
Method	$x = -0.8$	$x = 0.5$	$x = 0.8$	$x = -0.8$	$x = 0.5$	$x = 0.8$	$x = -0.8$	$x = 0.5$	$x = 0.8$
				$n = 100$					
OSNMR	94.5%	96.0%	96.0%	93.5%	93.5%	94.0%	90.5%	86.0%	91.5%
BMR	93.0%	95.5%	93.5%	94.5%	94.5%	94.5%	91.0%	88.0%	92.5%
				$n = 200$					
OSNMR	94.0%	93.5%	95.5%	95.5%	96.0%	94.0%	93.5%	88.0%	93.5%
BMR	94.0%	93.0%	96.0%	94.5%	94.0%	94.5%	94.0%	89.5%	94.5%
				$n = 400$					
OSNMR	95.5%	94.5%	96.0%	94.0%	93.5%	95.0%	94.5%	92.5%	95.0%
BMR	94.0%	94.0%	97.0%	95.0%	92.5%	94.5%	94.0%	92.5%	95.0%

Table 4. Average MSE and SD in parentheses for the three estimators under Case 3 with heavy-tailed $t(3)$ errors.

n	OSNMR	BMR	LPM
100	0.3513 (0.5295)	0.3205 (0.5172)	0.2189 (0.3066)
200	0.1961 (0.4029)	0.1830 (0.4026)	0.1376 (0.2368)
400	0.1073 (0.2981)	0.1032 (0.3037)	0.0662 (0.1797)

and BMR produce very similar fitted curves. In Case 2, where the true modal function is smoother and less oscillatory, LPM again has the smallest MSE, but OSNMR and BMR still follow the true curve closely. For the more oscillatory function in Case 3, the advantage of global approximation becomes clearer. OSNMR and BMR give smaller MSE values than LPM, especially for $n = 100$ and $n = 200$, and their fitted curves track the main turning points of the true function more accurately. Although LPM gives smaller MSE values in the first two relatively smooth cases, this improvement comes with a noticeably higher computational cost.

The computing times in Table 2 show a clear difference among the three methods. OSNMR and BMR require comparable computational time, both increasing only moderately as the sample size grows. In contrast, LPM is consistently more time-consuming because it estimates the modal regression function point by point. This difference is particularly visible in Case 1, where LPM takes much longer than the two global methods. Taken together, the simulation results suggest that OSNMR provides a useful balance between estimation accuracy and computational efficiency. It performs similarly to BMR in most settings and is much faster than LPM, while still giving competitive fitted modal curves. Finally, Table 3 reports the empirical coverage probabilities of the 95% confidence intervals for OSNMR and BMR at $x = -0.8$, 0.5, and 0.8. In general, the coverage probabilities of both methods are close to the nominal level of 95%. For the smoother functions in Cases 1 and 2, the results are quite stable across sample sizes and evaluation points. For the more oscillatory function in Case 3, the coverage probabilities are slightly lower in some cases, especially at $x = 0.5$ when the sample size is small. This is expected, since the local shape of the true function is more difficult to estimate in this case. When the sample size increases, the coverage probabilities move closer to 95%, which further supports the finite-sample reliability of the proposed OSNMR estimator.

Table 5. Average MSE and SD in parentheses for the sensitivity analysis of OSNMR under different choices of C.

Case	n	3%	5%	7%	10%
Case 1	100	0.0452 (0.1934)	0.0442 (0.1935)	0.0353 (0.1714)	0.0440 (0.1761)
	200	0.0406 (0.1849)	0.0260 (0.1482)	0.0205 (0.1323)	0.0241 (0.1372)
	400	0.0294 (0.1571)	0.0141 (0.1094)	0.0126 (0.1032)	0.0180 (0.1235)
Case 2	100	0.0605 (0.2308)	0.0593 (0.2324)	0.0415 (0.1897)	0.0306 (0.1650)
	200	0.0494 (0.2102)	0.0283 (0.1576)	0.0250 (0.1430)	0.0190 (0.1306)
	400	0.0324 (0.1688)	0.0167 (0.1181)	0.0144 (0.1112)	0.0177 (0.1271)
Case 3	100	0.1289 (0.2947)	0.1172 (0.2662)	0.0856 (0.2317)	0.0780 (0.2141)
	200	0.0663 (0.2272)	0.0446 (0.1751)	0.0346 (0.1522)	0.0335 (0.1488)
	400	0.0398 (0.1821)	0.0215 (0.1329)	0.0183 (0.1238)	0.0256 (0.1443)

To further examine the robustness of the proposed method, we include an additional heavy-tailed setting for Case 3. Specifically, we keep the regression function and generate the error term from $\epsilon \sim t(3)$. All other simulation settings remain the same. The corresponding results are shown in Table 4 and Figure 6. The heavy-tailed error clearly makes the problem harder, so the MSE values are larger than those in the previous cases. Nevertheless, the proposed OSNMR estimator still performs in a stable way. Its MSE decreases from 0.3513 for $n = 100$ to 0.1073 for $n = 400$, and the fitted curves in Figure 6 gradually approach the true modal regression function as the sample size increases. This indicates that

OSNMR still has good robustness when the error distribution has heavier tails.

We also assess the influence of the tuning constant in the CV criterion through a sensitivity analysis for OSNMR. In this part, we take $C = 3\%, 5\%, 7\%$, and 10%, respectively, and repeat the simulation $R = 200$ times for each case and each sample size. From Table 5, the estimation errors do not change sharply when C varies from 3% to 10%. In most settings, the MSE decreases as the sample size increases, which is consistent with the main simulation results. The choice $C = 5\%$ gives a balanced performance and is therefore used as the default value in the main numerical study. For a fixed value of C, Figure 4 shows that the selected bandwidth h tends to become smaller as the sample size increases. Figure 5 shows that the truncation parameter k generally increases with the sample size, meaning that more basis terms can be used when more data are available. At the same time, changing C does not lead to a drastic shift in either h or k. These findings indicate that the tuning constant c is related to the bias–variance trade-off of the estimator, while the proposed method is not overly sensitive to the specific choice of C within the considered range. In general, a smaller value of c tends to reduce smoothing bias but may increase estimation variance, whereas a larger value of c tends to produce smoother estimates with lower variance but relatively larger bias.

6. Real data analysis

We apply the proposed OSNMR method to the speed-flow data, a classical dataset in transportation studies. Speed-flow diagrams are commonly used to describe how vehicle speed changes with traffic flow. In this dataset, traffic flow is measured in vehicles per lane per hour, and speed is measured in miles per hour. The observations were collected from a four-lane Californian freeway using loop detectors, with each point recorded over a 30-second interval. This dataset was originally analyzed in the modal regression literature by Einbeck and Tutz [7], who emphasized that the speed-flow relationship may display different traffic regimes and therefore cannot always be well-represented by ordinary mean regression.

The same dataset was also considered by Yang et al. [8] in their study of B-spline modal regression. Different from the multimodal branch estimation perspective of Einbeck and Tutz [7], Yang et al. [8] used the data to compare single-valued modal regression estimators, including the B-spline modal regression estimator and the local polynomial modal estimator. This setting is closer to the focus of the present paper, since our theoretical development is also based on a well-defined single-valued conditional mode.

In this application, we use the full sample from the R package `hdrcde`. The speed-flow data contain 1318 observations. We compare the proposed OSNMR estimator with BMR, LPM, and mean regression (MR). Since the traffic-flow variable is measured on a relatively large original scale, the Legendre orthonormal basis is used in OSNMR after an internal location-scale transformation of the covariate. This transformation is used only to construct the basis functions and does not change the scale on which the fitted curves are displayed. The final results are plotted against the original traffic-flow variable. The purpose of this application is not to recover all possible traffic regimes, but to compare how the proposed OSNMR method performs relative to BMR and LPM when estimating the dominant modal trend in the speed-flow relationship.

For the three modal regression methods, the hyperparameters are selected using the cross-validation criterion described in Section 3.2. In this application, the constant in the CV criterion is defined as $c = C\{\max(Y) - \min(Y)\}$, where Y denotes the speed variable. We set $C = 5\%$, so that c corresponds

to 5% of the empirical range of speed. This choice is consistent with the simulation study and is used for all modal regression estimators in the real data analysis. For OSNMR, the selected bandwidth is $h = 0.5351$, and the selected truncation parameter is $k = 8$. For BMR, the selected bandwidth is $h = 2.68$, with 11 internal knots. For LPM, the selected bandwidths are $h_1 = 0.64$ in the covariate direction and $h_2 = 1$ in the response direction. These selected values suggest that the three modal regression methods use comparable smoothing levels in the response direction, while differing mainly in how they approximate the regression curve. From Figure 7, the mean regression curve does not fully characterize the dominant relationship between speed and traffic flow. A similar observation has also been made in previous studies using this dataset, see [7, 20]. Overall, the three modal regression estimators capture a similar dominant modal pattern in the speed-flow data. OSNMR gives a fitted curve that is close to that of BMR, which is reasonable since both methods are global approximation methods. LPM also captures the main trend, but it requires pointwise estimation and is therefore much more computationally demanding. In addition, we display the pointwise 95% bootstrap confidence intervals for the OSNMR fitted curve, based on $B = 100$ bootstrap replications.

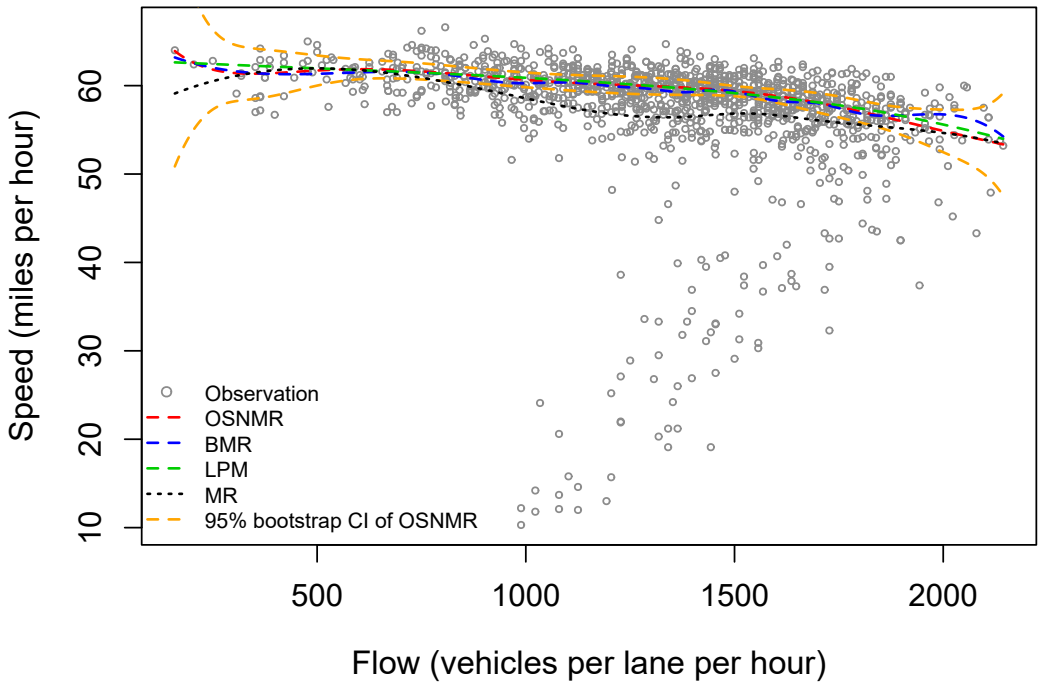


Figure 7. Fitted modal regression curves for the speed-flow data. The orange dashed curves denote the pointwise 95% bootstrap confidence intervals for the OSNMR fitted curve based on $B = 100$ bootstrap replications.

Remark 2. *The empirical coverage probabilities reported in the simulation study are based on Monte Carlo standard deviations across repeated simulations. This construction is useful for finite sample evaluation, but it is not directly applicable to a single real dataset, because the true data-generating*

mechanism and repeated samples are unavailable in practice. For the real data analysis, we use a bootstrap resampling method to assess the uncertainty of the fitted OSNMR curve.

Let $\{(Y_i, X_i), i = 1, \ldots, n\}$ *denote the observed data. For the bth bootstrap replication,* $b = 1, \ldots, B$*, we obtain a bootstrap sample* $\{(Y_i^{*(b)}, X_i^{*(b)}), i = 1, \ldots, n\}$ *by sampling the observed pairs with replacement. The OSNMR estimator is then refitted to the bootstrap sample, giving the bootstrap fitted curve* $\widehat{m}^{*(b)}(x)$*. Repeating this procedure B times gives B bootstrap estimates* $\widehat{m}^{*(1)}(x), \ldots, \widehat{m}^{*(B)}(x)$*. Alternatively, following the normal approximation idea commonly used in bootstrap inference, we compute the bootstrap standard error* $\widehat{\sigma}_B(x) = \left[\frac{1}{B-1}\sum_{b=1}^{B}\left\{\widehat{m}^{*(b)}(x) - \overline{\widehat{m}}^{*}(x)\right\}^2\right]^{1/2}$*, where* $\overline{\widehat{m}}^{*}(x) = \frac{1}{B}\sum_{b=1}^{B}\widehat{m}^{*(b)}(x)$*. Then a pointwise* $(1-\alpha)$ *confidence interval can also be written as* $\widehat{m}(x) \pm z_{1-\alpha/2}\widehat{\sigma}_B(x)$*, where* $z_{1-\alpha/2}$ *is the* $(1-\alpha/2)$ *quantile of the standard normal distribution. Similar bootstrap ideas have been used in nonparametric function estimation [30, 31].*

Under the same computing environment (R version 4.4.2 on Windows 11, Intel Core i7-12700K CPU, 64 GB RAM), the average computation times are reported in Table 6. OSNMR achieves competitive efficiency, comparable to BMR and substantially faster than LPM. To further evaluate fitting and prediction performance, we randomly split the full data into a training set and a test set. In each split, $n_{\text{training}} = 918$ observations are used to fit the model, and the remaining $n_{\text{test}} = 400$ observations are used for prediction evaluation. The prediction performance is measured by the mean absolute deviation, $\text{MAD}_{\text{pred}} = n_{\text{test}}^{-1}\sum_{i=1}^{n_{\text{test}}}\left|y_i^{\text{test}} - \hat{y}_i^{\text{test}}\right|$, where $\hat{y}_i^{\text{test}}$ denotes the predicted value for the ith test observation. Similarly, the fitting performance on the training set is measured by $\text{MAD}_{\text{fit}} = n_{\text{training}}^{-1}\sum_{i=1}^{n_{\text{training}}}\left|y_i^{\text{training}} - \hat{y}_i^{\text{training}}\right|$. The reported values in Table 6 are averages over 100 replications. From Table 6, the three modal regression methods have smaller MAD values than MR, indicating that modal regression is more suitable for capturing the dominant trend of the speed-flow relationship. OSNMR gives slightly larger MAD values than BMR but smaller values than LPM and MR. More importantly, OSNMR has a much shorter average computation time than LPM, while maintaining fitting and prediction performance close to BMR. These results suggest that OSNMR provides a good balance between estimation accuracy and computational efficiency in this real data application.

Table 6. Fitting and prediction performance for the speed-flow data.

Method	MAD_{fit}	MAD_{pred}	Mean time (s)
OSNMR	3.4049	3.4453	0.5804
BMR	3.3941	3.4395	0.7580
LPM	3.4215	3.4511	37.7912
MR	3.9322	3.9993	0.0047

7. Conclusions and discussion

In this paper, we proposed a nonparametric modal regression method based on orthogonal series expansion. The proposed OSNMR estimator approximates the unknown modal regression function by a finite number of orthonormal basis functions, which turns the original nonparametric estimation problem into a finite-dimensional optimization problem. This construction keeps the flexibility of nonparametric modal regression while giving a relatively simple and transparent estimation form.

We compared the proposed method with the B-spline modal regression estimator and the local polynomial modal regression estimator through simulation studies and a real data application. The results suggest that OSNMR can provide competitive fitting performance. In particular, its computational cost is close to that of BMR and is much lower than that of LPM, which is based on pointwise estimation. This makes the proposed method a useful alternative when the sample size is moderate or large and a global estimate of the modal regression function is desired.

The selection of hyperparameters remains an important issue in modal regression, especially because the estimation criterion is based on kernel density estimation. In this paper, we used a cross-validation criterion tailored to modal regression. The criterion selects hyperparameters by favoring fitted curves whose fixed-width neighborhoods contain a larger proportion of observations. The simulation and application results show that this criterion works well in practice for selecting the bandwidth and the truncation parameter.

We also established the main asymptotic properties of the proposed estimator under suitable regularity conditions, including its convergence rate and asymptotic normality. These results provide theoretical support for the proposed orthogonal series modal regression framework.

There are still several directions worth further study. The present paper focuses on the unimodal setting, where the conditional mode is assumed to be uniquely defined. In many real applications, however, the conditional distribution may have more than one local mode. Extending the proposed orthogonal series approach to nonparametric multimodal regression, where several modal branches are estimated simultaneously, is an interesting and important topic for future research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors thank the Editor and the anonymous reviewers for their valuable comments and suggestions, leading to significant improvement of this article. This work was supported by the Zhejiang Provincial Natural Science Foundation [Grant No. LMS25A010001], the Strategic Discipline Team of New Economic Statistical Monitoring and Intelligent Decision-Making Research, and the First Class Discipline of Zhejiang-A (Zhejiang University of Finance and Economics-Statistics).

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. M. J. Lee, Mode regression, *J. Econom.*, **42** (1989), 337–349. https://doi.org/10.1016/0304-4076(89)90057-2
2. W. Yao, B. G. Lindsay, R. Li, Local modal regression, *J. Nonparametric Stat.*, **24** (2012), 647–663. https://doi.org/10.1080/10485252.2012.678848

3. W. Yao, L. Li, A new regression model: Modal linear regression, *Scand. J. Stat.*, **41** (2014), 656–671. https://doi.org/10.1111/sjos.12054
4. Y. C. Chen, C. R. Genovese, R. J. Tibshirani, L. Wasserman, Nonparametric modal regression, *Ann. Stat.*, **44** (2016), 489–514. https://doi.org/10.1214/15-AOS1373
5. Y. C. Chen, Modal regression using kernel density estimation: A review, *Wiley Interdiscip. Rev.-Comput. Stat.*, **10** (2018), e1431. https://doi.org/10.1002/wics.1431
6. T. Wang, Non-parametric estimator for conditional mode with parametric features, *Oxf. Bull. Econ. Stat.*, **86** (2024), 44–73. https://doi.org/10.1111/obes.12577
7. J. Einbeck, G. Tutz, Modelling beyond regression functions: An application of multimodal regression to speed–flow data, *J. R. Stat. Soc. Ser. C-Appl. Stat.*, **55** (2006), 461–475. https://doi.org/10.1111/j.1467-9876.2006.00547.x
8. L. Yang, W. Yuan, S. Wang, Modal regression models based on B-splines, *Comput. Stat.*, **40** (2025), 225–248. https://doi.org/10.1007/s00180-024-01487-0
9. A. Ullah, T. Wang, W. Yao, Nonlinear modal regression for dependent data with application for predicting COVID-19, *J. R. Stat. Soc. Ser. A-Stat. Soc.*, **185** (2022), 1424–1453. https://doi.org/10.1111/rssa.12849
10. X. Jing, J. S. Cho, Forecasting the confirmed COVID-19 cases using modal regression, *J. Forecast.*, **44** (2025), 1578–1601. https://doi.org/10.1002/for.3261
11. H. Yuan, S. Xiang, W. Yao, A new bandwidth selection method for nonparametric modal regression based on generalized hyperbolic distributions, *Comput. Stat.*, **39** (2024), 1729–1746. https://doi.org/10.1007/s00180-023-01435-4
12. D. J. Henderson, C. F. Parmeter, R. R. Russell, Modes, weighted modes, and calibrated modes: Evidence of clustering using modality tests, *J. Appl. Econom.*, **23** (2008), 607–638. https://doi.org/10.1002/jae.1023
13. Z. C. Han, Y. Y. Zhao, Nonparametric modal regression with mixed variables and application to analyze the GDP data, *J. Comput. Appl. Math.*, **445** (2024), 115841. https://doi.org/10.1016/j.cam.2024.115841
14. M. J. Lee, Quadratic mode regression, *J. Econom.*, **57** (1993), 1–19. https://doi.org/10.1016/0304-4076(93)90056-B
15. G. Kemp, J. Santos Silva, Regression towards the mode, *J. Econom.*, **170** (2012), 92–101. https://doi.org/10.1016/j.jeconom.2012.03.002
16. H. Zhou, X. Huang, Nonparametric modal regression in the presence of measurement error, *Electron. J. Stat.*, **10** (2016), 3579–3620. https://doi.org/10.1214/16-EJS1210
17. X. Li, X. Huang, Linear mode regression with covariate measurement error, *Can. J. Stat.*, **47** (2019), 262–280. https://doi.org/10.1002/cjs.11492
18. S. Xiang, W. Yao, Nonparametric statistical learning based on modal regression, *J. Comput. Appl. Math.*, **409** (2022), 114130. https://doi.org/10.1016/j.cam.2022.114130
19. H. Ota, K. Kato, S. Hara, Quantile regression approach to conditional mode estimation, *Electron. J. Stat.*, **13** (2019), 3120–3160. https://doi.org/10.1214/19-EJS1607

20. Y. Feng, J. Fan, J. A. K. Suykens, A statistical learning approach to modal regression, *J. Mach. Learn. Res.*, **21** (2020), 1–35. https://doi.org/10.5555/3455716.3455718

21. J. M. Krief, Semi-linear mode regression, *Econom. J.*, **20** (2017), 149–167. https://doi.org/10.1111/ectj.12088

22. A. Ullah, T. Wang, W. Yao, Modal regression for fixed effects panel data, *Empirical Econ.*, **60** (2021), 261–308. https://doi.org/10.1007/s00181-020-01999-w

23. A. Ullah, T. Wang, W. Yao, Semiparametric partially linear varying coefficient modal regression, *J. Econom.*, **235** (2023), 1001–1026. https://doi.org/10.1016/j.jeconom.2022.09.002

24. T. Wang, W. Yao, Nonparametric spatial mode-oriented regression, *Electron. J. Stat.*, **19** (2025), 4256–4311. https://doi.org/10.1214/25-EJS2427

25. T. Wang, W. Yao, Online kernel-based mode learning, *J. Comput. Graph. Stat.*, **34** (2025), 1498–1512. https://doi.org/10.1080/10618600.2025.2459287

26. S. Xiang, W. Yao, X. Cui, Advances in modal regression: From theoretical foundations to practical implementations, *Wiley Interdiscip. Rev.-Comput. Stat.*, **18** (2026), e70061. https://doi.org/10.1002/wics.70061

27. C. Dong, J. Gao, *Modern Series Methods in Econometrics and Statistics*, Springer, Switzerland, 2025. https://doi.org/10.1007/978-981-96-2822-3

28. J. Li, S. Ray, B. G. Lindsay, A nonparametric statistical approach to clustering via mode identification, *J. Mach. Learn. Res.*, **8** (2007), 1687–1723. https://doi.org/10.5555/1314498.1314555

29. H. Zhou, X. Huang, Bandwidth selection for nonparametric modal regression, *Commun. Stat.-Simul. Comput.*, **48** (2019), 968–984. https://doi.org/10.1080/03610918.2017.1402044

30. J. Z. Huang, C. O. Wu, L. Zhou, Varying-coefficient models and basis function approximations for the analysis of repeated measurements, *Biometrika*, **89** (2002), 111–128. https://doi.org/10.1093/biomet/89.1.111

31. W. Zhao, R. Zhang, J. Liu, Y. Lv, Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression, *Ann. Inst. Stat. Math.*, **66** (2014), 165–191. https://doi.org/10.1007/s10463-013-0410-4

32. W. K. Newey, D. McFadden, Large sample estimation and hypothesis testing, in *Handbook of Econometrics*, Elsevier, Amsterdam, (1994), 2111–2245. https://doi.org/10.1016/S1573-4412(05)80005-4

Appendix A

Throughout this Appendix, C denotes a generic positive constant that may vary from line to line. Let $\|\cdot\|$ denote the Euclidean norm for vectors and the corresponding matrix norm for matrices. Recall that $m(x) = m_k(x) + \gamma_k(x)$, $m_k(x) = \Psi_k(x)^\top \boldsymbol{\beta}_k$, where $\Psi_k(x) = (\psi_0(x), \ldots, \psi_{k-1}(x))^\top$ and $\gamma_k(x) = m(x) - m_k(x)$ is the truncation error. By Condition 4, $\|\gamma_k\|_{L^2} = O(k^{-s})$, $\sup_{x\in\mathcal{X}} |\gamma_k(x)| = O(a_k)$.

Let $Q(\boldsymbol{\beta}) = E\left[\phi_h\{Y - \Psi_k(X)^\top \boldsymbol{\beta}\}\right]$ be the population objective function, and let $\boldsymbol{\beta}_{k,h}$ denote the maximizer of $Q(\boldsymbol{\beta})$ in a neighborhood of $\boldsymbol{\beta}_k$. Define $B_{k,h}(x) = \Psi_k(x)^\top(\boldsymbol{\beta}_{k,h} - \boldsymbol{\beta}_k) - \gamma_k(x)$, where

$\gamma_k(x) = m(x) - m_k(x)$. Also define

$$\boldsymbol{H}_k = -E\left\{g''(0 \mid X)\Psi_k(X)\Psi_k(X)^\top\right\}$$

and

$$\boldsymbol{V}_k = \left[\int \{\phi'(u)\}^2 du\right] E\left\{g(0 \mid X)\Psi_k(X)\Psi_k(X)^\top\right\}.$$

Under Conditions 2 and 5, $\boldsymbol{H}_k$ is positive definite.

Proof of Theorem 1. To justify the local expansion around the population maximizer, we first verify that $\widehat{\boldsymbol{\beta}}_k$ is consistent for $\boldsymbol{\beta}_{k,h}$. Under Conditions 1–4, $Q(\boldsymbol{\beta})$ is continuous and is uniquely maximized at $\boldsymbol{\beta}_{k,h}$ in a neighborhood of $\boldsymbol{\beta}_k$. Moreover, Conditions 3, 5, and 6 imply the uniform convergence

$$\sup_{\boldsymbol{\beta}} |Q_n(\boldsymbol{\beta}) - Q(\boldsymbol{\beta})| = o_p(1).$$

Therefore, by the standard argmax theorem for extremum estimators (Theorem 2.1 of [32]),

$$\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h} = o_p(1).$$

For the increasing-dimensional series approximation considered here, this argument is understood under the growth restrictions imposed in Condition 6, which ensure stochastic equicontinuity over the local sieve parameter space. This consistency result ensures that $\widehat{\boldsymbol{\beta}}_k$ lies in a shrinking neighborhood of $\boldsymbol{\beta}_{k,h}$ with a probability approaching one, so the following Taylor expansion is valid.

We first write the estimation error as the sum of a coefficient estimation error and a truncation error. Since $\hat{m}_k(x) = \Psi_k(x)^\top \hat{\boldsymbol{\beta}}_k$ and $m(x) = \Psi_k(x)^\top \boldsymbol{\beta}_k + \gamma_k(x)$, we have

$$\hat{m}_k(x) - m(x) = \Psi_k(x)^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) - \gamma_k(x).$$

Taking the L^2 norm gives

$$\|\hat{m}_k - m\|_{L^2} \le \left\|\Psi_k(\cdot)^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)\right\|_{L^2} + \|\gamma_k\|_{L^2}.$$

By the orthonormality of $\{\psi_j\}_{j\ge 0}$, for any $\boldsymbol{a} \in \mathbb{R}^k$, $\left\|\Psi_k(\cdot)^\top \boldsymbol{a}\right\|_{L^2} = \|\boldsymbol{a}\|$. Therefore,

$$\|\hat{m}_k - m\|_{L^2} \le \|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| + \|\gamma_k\|_{L^2}.$$

It remains to control $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\|$. Let

$$Q_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \phi_h\left(Y_i - \Psi_k(X_i)^\top \boldsymbol{\beta}\right)$$

and

$$Q(\boldsymbol{\beta}) = E\left[\phi_h\left(Y - \Psi_k(X)^\top \boldsymbol{\beta}\right)\right].$$

Let $\boldsymbol{\beta}_{k,h}$ be the maximizer of the population objective $Q(\boldsymbol{\beta})$ in a neighborhood of $\boldsymbol{\beta}_k$. We decompose $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k = (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h}) + (\boldsymbol{\beta}_{k,h} - \boldsymbol{\beta}_k)$.

We first consider the deterministic bias term $\boldsymbol{\beta}_{k,h} - \boldsymbol{\beta}_k$. Since

$$Y = m(X) + \epsilon = \Psi_k(X)^\top \boldsymbol{\beta}_k + \gamma_k(X) + \epsilon,$$

for any $\boldsymbol{\beta}$ near $\boldsymbol{\beta}_k$,

$$Y - \Psi_k(X)^\top \boldsymbol{\beta} = \epsilon + \gamma_k(X) - \Psi_k(X)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_k).$$

Thus, conditional on $X = x$,

$$E\left[\phi_h\{Y - \Psi_k(X)^\top \boldsymbol{\beta}\} \mid X = x\right] = \int \phi(u) g\left(\Psi_k(x)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_k) - \gamma_k(x) + hu \mid x\right) du.$$

By Conditions 2–4, $g(t \mid x)$ is smooth near $t = 0$, $g'(0 \mid x) = 0$, and $g''(0 \mid x) < 0$ uniformly in x. A Taylor expansion around zero, together with the symmetry of ϕ, gives that the population objective is locally maximized at a point satisfying

$$\|\boldsymbol{\beta}_{k,h} - \boldsymbol{\beta}_k\| = O(h^2 + k^{-s}).$$

Here the term h^2 is the kernel smoothing bias, and the term k^{-s} is caused by the truncation error γ_k.

We next control the stochastic term $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h}$. Define the score function

$$S_n(\boldsymbol{\beta}) = \frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} \phi' \left(\frac{Y_i - \Psi_k(X_i)^\top \boldsymbol{\beta}}{h} \right) \Psi_k(X_i).$$

Since $\hat{\boldsymbol{\beta}}_k$ maximizes $Q_n(\boldsymbol{\beta})$, it satisfies the first-order condition $S_n(\widehat{\boldsymbol{\beta}}_k) = 0$. Expanding $S_n(\hat{\boldsymbol{\beta}}_k)$ around $\boldsymbol{\beta}_{k,h}$ yields $0 = S_n(\boldsymbol{\beta}_{k,h}) - J_n(\widetilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h})$, where $\widetilde{\boldsymbol{\beta}}$ lies between $\hat{\boldsymbol{\beta}}_k$ and $\boldsymbol{\beta}_{k,h}$, and $J_n(\boldsymbol{\beta}) = -\frac{\partial^2 Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}$.

Under Conditions 1–6, the negative Hessian satisfies $\|J_n(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{H}_k\| = o_p(1)$, uniformly in a shrinking neighborhood of $\boldsymbol{\beta}_{k,h}$, where $\boldsymbol{H}_k = -E\left\{g''(0 \mid X) \Psi_k(X) \Psi_k(X)^\top\right\}$.

By Conditions 2 and 5, $\boldsymbol{H}_k$ is positive definite.

$$\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h} = \boldsymbol{H}_k^{-1} S_n(\boldsymbol{\beta}_{k,h}) + o_p\left(\|S_n(\boldsymbol{\beta}_{k,h})\|\right).$$

We now evaluate the stochastic order of $S_n(\boldsymbol{\beta}_{k,h})$. Since $\boldsymbol{\beta}_{k,h}$ is the population maximizer, $E\{S_n(\boldsymbol{\beta}_{k,h})\} = 0$. Moreover, using the change of variable $u = \epsilon/h$, the boundedness of $g(0 \mid x)$, and Condition 5,

$$E\left[\left\| \frac{1}{h^2} \phi'\left(\frac{\epsilon}{h}\right) \Psi_k(X) \right\|^2\right] \le \frac{C}{h^3} E\|\Psi_k(X)\|^2 = O\left(\frac{k}{h^3}\right).$$

Therefore, $\|S_n(\boldsymbol{\beta}_{k,h})\| = O_p\left(\sqrt{\frac{k}{nh^3}}\right)$. It follows that

$$\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h}\| = O_p\left(\sqrt{\frac{k}{nh^3}}\right).$$

Combining this with the population bias bound gives $\|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| = O_p\left(\sqrt{\frac{k}{nh^3}} + h^2 + k^{-s}\right)$.

Finally, using $\|\gamma_k\|_{L^2} = O(k^{-s})$, we obtain

$$\|\hat{m}_k - m\|_{L^2} = O_p\left(\sqrt{\frac{k}{nh^3}} + h^2 + k^{-s}\right).$$

This completes the proof. □

Proof of Theorem 2. Fix an interior point $x \in \mathcal{X}$. We start from the decomposition

$$\hat{m}_k(x) - m(x) = \Psi_k(x)^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h}) + \{\Psi_k(x)^\top(\boldsymbol{\beta}_{k,h} - \boldsymbol{\beta}_k) - \gamma_k(x)\}.$$

Define $B_{k,h}(x) = \Psi_k(x)^\top(\boldsymbol{\beta}_{k,h} - \boldsymbol{\beta}_k) - \gamma_k(x)$. Then

$$\hat{m}_k(x) - m(x) - B_{k,h}(x) = \Psi_k(x)^\top(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h}).$$

From the Taylor expansion of the score around $\boldsymbol{\beta}_{k,h}$, as shown in the proof of Theorem 1,

$$\Psi_k(x)^\top \left\{\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_{k,h} - \boldsymbol{H}_k^{-1} S_n(\boldsymbol{\beta}_{k,h})\right\} = o_p\left((nh^3)^{-1/2}\right),$$

where $S_n(\boldsymbol{\beta}_{k,h}) = -\frac{1}{n}\sum_{i=1}^n \frac{1}{h^2}\phi'\left(\frac{Y_i - \Psi_k(X_i)^\top \boldsymbol{\beta}_{k,h}}{h}\right)\Psi_k(X_i)$. Therefore,

$$\widehat{m}_k(x) - m(x) - B_{k,h}(x) = \Psi_k(x)^\top \boldsymbol{H}_k^{-1} S_n(\boldsymbol{\beta}_{k,h}) + o_p\left((nh^3)^{-1/2}\right).$$

Now write the leading term as

$$\Psi_k(x)^\top \boldsymbol{H}_k^{-1} S_n(\boldsymbol{\beta}_{k,h}) = \frac{1}{n}\sum_{i=1}^n \xi_{ni}(x),$$

where

$$\xi_{ni}(x) = -\frac{1}{h^2}\Psi_k(x)^\top \boldsymbol{H}_k^{-1}\Psi_k(X_i)\phi'\left(\frac{Y_i - \Psi_k(X_i)^\top \boldsymbol{\beta}_{k,h}}{h}\right).$$

By the definition of $\boldsymbol{\beta}_{k,h}$, the leading term is centered up to a negligible higher-order term. Hence, $E\{\xi_{ni}(x)\} = o(1)$.

We next compute the variance. Since $Y_i - \Psi_k(X_i)^\top \boldsymbol{\beta}_{k,h} = \epsilon_i + o_p(1)$ in a neighborhood of the population maximizer, the leading variance is obtained from

$$E\left[\frac{1}{h^4}\{\phi'(\epsilon_i/h)\}^2 \Psi_k(X_i)\Psi_k(X_i)^\top\right].$$

Using the change of variable $u = \epsilon_i/h$ and the smoothness of $g(t \mid X_i)$ around zero,

$$E\left[\frac{1}{h^4}\{\phi'(\epsilon_i/h)\}^2 \Psi_k(X_i)\Psi_k(X_i)^\top\right] = \frac{1}{h^3}\left[\int \{\phi'(u)\}^2 du\right] E\left[g(0 \mid X)\Psi_k(X)\Psi_k(X)^\top\right] + o\left(\frac{1}{h^3}\right).$$

Thus,

$$\mathrm{Var}\left\{\Psi_k(x)^\top \boldsymbol{H}_k^{-1} S_n(\boldsymbol{\beta}_{k,h})\right\} = \frac{1}{nh^3}\Psi_k(x)^\top \boldsymbol{H}_k^{-1} \boldsymbol{V}_k \boldsymbol{H}_k^{-1}\Psi_k(x) + o\left(\frac{1}{nh^3}\right),$$

where $\boldsymbol{V}_k = \left[\int \{\phi'(u)\}^2 du\right] E\left\{g(0 \mid X)\Psi_k(X)\Psi_k(X)^\top\right\}$.

It remains to verify the central limit theorem. The summands $\xi_{ni}(x)$ are independent across i. The boundedness assumptions on ϕ', the smoothness and boundedness of $g(t \mid x)$ near zero, and the growth restriction on ζ_k ensure the Lindeberg condition. In particular, the condition $\frac{\zeta_k^2 \log n}{nh^3} \to 0$ prevents any

single observation from dominating the normalized sum. Therefore, by the Lindeberg central limit theorem for triangular arrays,

$$\frac{\sqrt{nh^3}\Psi_k(x)^\top \boldsymbol{H}_k^{-1} S_n(\boldsymbol{\beta}_{k,h})}{\left[\Psi_k(x)^\top \boldsymbol{H}_k^{-1} \boldsymbol{V}_k \boldsymbol{H}_k^{-1} \Psi_k(x)\right]^{1/2}} \xrightarrow{d} N(0,1).$$

Combining this result with the linear expansion above gives

$$\frac{\sqrt{nh^3}\left\{\widehat{m}_k(x) - m(x) - B_{k,h}(x)\right\}}{\left[\Psi_k(x)^\top \boldsymbol{H}_k^{-1} \boldsymbol{V}_k \boldsymbol{H}_k^{-1} \Psi_k(x)\right]^{1/2}} \xrightarrow{d} N(0,1).$$

Since $|B_{k,h}(x)| = O(h^2 + a_k)$, the undersmoothing condition implies that $\sqrt{nh^3}B_{k,h}(x) = o(1)$. Under the undersmoothing condition $\sqrt{nh^3}\{h^2 + a_k\} \to 0$, the bias term is asymptotically negligible and

$$\frac{\sqrt{nh^3}\left\{\hat{m}_k(x) - m(x)\right\}}{\left[\Psi_k(x)^\top \boldsymbol{H}_k^{-1} \boldsymbol{V}_k \boldsymbol{H}_k^{-1} \Psi_k(x)\right]^{1/2}} \xrightarrow{d} N(0,1).$$

This completes the proof. □