Statistica Sinica Preprint No: SS-2023-0288	
Title	Kernel Mode-Based Regression under Random Truncation
Manuscript ID	SS-2023-0288
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0288
Complete List of Authors	Tao Wang and Weixin Yao
Corresponding Authors	Tao Wang
E-mails	taow@uvic.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Kernel Mode-Based Regression under Random Truncation

Tao Wang[a] Weixin Yao[b]

a. University of Victoria *b. University of California Riverside*

Abstract: We propose to estimate a parametric regression with truncated data built on the mode value, where the dependent variable is subject to left truncation by another random variable. We construct a kernel mode-based objective function with a constant bandwidth for estimation and suggest a modified mode expectation-maximization algorithm to numerically estimate the model. The asymptotic normal distribution of the proposed estimator is derived under mild conditions. To efficiently construct confidence intervals for the resulting estimator, we develop a mode-based empirical likelihood method, where the asymptotic distribution of the empirical log-likelihood ratio is shown to follow a chi-square distribution. Furthermore, by combining the kernel mode-based objective function with the SCAD penalty, a variable selection procedure for the parameters is introduced and its oracle property is established. Monte Carlo simulations and real data analysis related to housing market are presented to show the finite sample performance of the developed estimation and variable selection procedures.

Key words and phrases: Empirical likelihood, Mode-based regression, Random truncation, Robust estimation, Variable selection.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

1. Introduction

The concept of the mode is attractive as the value of highest probability density. The mode can be defined without moment conditions and could provide another understanding of the data, i.e., capture the "most likely" values. Owing to these appealing features, regression models based on the mode value, denoted as $Mode(Y \mid \mathbf{X})$ for random variables $(Y, \mathbf{X})$ (modal regression), have received significant attention recently (Yao and Li, 2014; Ullah et al., 2022, 2023), which can provide a valuable alternative to existing regressions. In addition, mode-based regression can be utilized as an alternative to robust regression to achieve robustness and efficiency in the presence of outliers or heavy-tailed distributions (Wang and Li, 2021; Wang, 2024). Due to space constraints, we provide more explanations on the distinction between modal regression and mode-based regression, as well as their respective advantages, in the supplementary file. However, all existing methods related to mode-based regression assume that the data are fully observed, which may be unrealistic in practical applications.

Truncated data, occurring when sample observations are restricted to some intervals, have been extensively studied in the literature. There exists a large number of research in mean or quantile regression to handle the truncated data since the early works of Amemiya (1973) and Hausman

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

and Wise (1977). However, most of these estimation procedures tend to be complicated and computationally intensive due to the fact that $\mathbb{E}(Y \mid \mathbf{X}) \neq \mathbb{E}(Y \mid \mathbf{X}, Y \geq y^*)$, where y^* is fixed at a prespecified value. Lee (1989, 1993) proposed estimating the regression line for truncated data by imposing a mode restriction on the error distribution. Because $Mode(Y \mid \mathbf{X}) = Mode(Y \mid \mathbf{X}, Y \geq y^*)$ generally holds for a known truncated point y^*, we can conduct mode estimation directly with observable data points to recover the "most likely" relationship between Y and $\mathbf{X}$. Nevertheless, the objective function operated in Lee (1989, 1993) is challenging to estimate in consequence of the presence of indicator function and maximum operator.

In this paper, we mainly concentrate on investigating parametric regression built on the mode value through a kernel objective function, in which the dependent variable is subject to left truncation by another random variable. Different from fixed truncation, random truncation corresponds to a biased sampling, where observations of variables $(Y, \mathbf{X})$ are interfered by another independent random variable T such that all three quantities of $Y, \mathbf{X}$, and T are observable only if $Y \geq T$. Many researchers have delved into randomly truncated data in the context of mean or quantile regression by using a weighted estimation procedure; see Wang (1989), He and Yang (2003), Zhou (2011) and the references therein. In this paper, we extend the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

weighted method to mode-based linear regression for randomly truncated data. Since there is no explicit expression for the resulting mode-based estimator, we also suggest a modified mode expectation-maximization (MEM) algorithm to numerically estimate the model.

As shown in Section 3, estimating the asymptotic covariance of the mode-based estimator numerically poses a challenge due to the presence of unknown terms. Also, in small sample sizes, confidence intervals formulated on asymptotic normality may experience significant coverage challenges, especially when the data distribution is nonnormal. To reliably construct confidence intervals, we utilize the empirical likelihood method (Owen, 1988, 1990) placed on the suggested kernel mode-based estimation. Compared to the normal approximation method, the developed empirical likelihood procedure can avoid the plug-in estimation for the limit variance and determine the shape of confidence intervals completely by data. The resulting empirical log-likelihood ratio is proved to satisfy the standard nonparametric Wilks' theorem, leading to a mode-based confidence interval.

Furthermore, for the practical selection of important covariates, we propose an efficient mode-based variable selection procedure by leveraging the Smoothly Clipped Absolute Deviation (SCAD) penalty (Fan and Li, 2001) with randomly truncated data. By inheriting the properties of the suggested

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

kernel mode-based estimation, the new variable selection procedure exhibits good robustness and efficiency. To circumvent the computationally intensive cross-validation approach, an extended Bayesian information criterion (BIC) is employed to consistently select the regularization parameter in the SCAD penalty. Because of the irregular of the SCAD penalty at the origin, we combine the suggested MEM algorithm with local quadratic approximation to develop a penalized MEM algorithm for numerically estimating.

The rest of the paper is organized as follows. In Section 2, we devote to the presentation of the applicability of mode value for truncated data with fixed truncation. In Section 3, we propose a parametric truncated regression model established on the mode value for randomly truncated data. In Section 4, we develop a mode-based variable section procedure with randomly truncated data. In Section 5, we present numerical studies to illustrate the finite sample performance of the suggested estimation and variable selection procedures. We conclude the paper in Section 6. All additional numerical and technical results are relegated to the supplementary file.

2. Motivation for Mode-Based Estimation

As a measure of center, the mode has the advantage of robustness since it concentrates on the majority of data points. Moreover, the mode, unlike

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

the mean, can be accurately approximated provided that the truncation is not excessively severe. We illustrate the use of mode by focusing on a parametric regression for truncated data within the latent variable framework.

Consider a truncated parametric regression model

$$Y = \mathbf{X}^T\boldsymbol{\beta} + \varepsilon, \tag{2.1}$$

where $Y \in \mathbb{R}$ is the dependent variable, $\mathbf{X} \in \mathbb{R}^p$ is the vector of covariates, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ is an unknown vector of parameters, and ε is the independent and identically distributed (i.i.d.) error with an unknown distribution function that may depend on the covariates $\mathbf{X}$. To capture the conditional mode estimator, we assume that $Mode(\varepsilon \mid \mathbf{X}) = 0$. The dependent variable and covariates are only revealed if the dependent variable $Y \geq y^*$, where y^* is a known truncation point. For simplicity, we let $y^* = 0$ in this section.

Let ε denote the unobserved random variable without truncation and ε^* represent the unobserved random variable with truncation. Therefore, $\varepsilon^* = \varepsilon \mid \varepsilon \geq -\mathbf{X}^T\boldsymbol{\beta}$. Generally, we will have $\mathbb{E}(\varepsilon^*) \neq \mathbb{E}(\varepsilon)$, but $Mode(\varepsilon^*) = Mode(\varepsilon)$ with the assumption of the existence of a global unique mode when truncation is not beyond the mode (see Figure S1 in the supplementary file). We then obtain the following lemma to support the identification of $\boldsymbol{\beta}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Lemma 1. *Suppose that* $Mode(\varepsilon \mid \mathbf{X}) = 0$, $P[Mode(Y \mid \mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta} > 0] > 0$, *and* $\mathbb{E}[\mathbf{X}\mathbf{X}^T \mid Mode(Y \mid \mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta} > 0]$ *is positive definite. Consequently, the parameter* $\boldsymbol{\beta}$ *in* (2.1) *can be identified based on the mode value since*

$$\begin{aligned}&\mathbb{E}[\mathbf{X}\mathbf{X}^T \mid Mode(Y \mid \mathbf{X}) > 0]^{-1}\mathbb{E}[\mathbf{X}Mode(Y \mid \mathbf{X}) \mid Mode(Y \mid \mathbf{X}) > 0]\\ =&\mathbb{E}[\mathbf{X}\mathbf{X}^T \mid \mathbf{X}^T\boldsymbol{\beta} > 0]^{-1}\mathbb{E}[\mathbf{X}\mathbf{X}^T\boldsymbol{\beta} \mid \mathbf{X}^T\boldsymbol{\beta} > 0] = \boldsymbol{\beta}.\end{aligned}$$

Lemma 1 indicates that mode-based estimation can yield a consistent estimator for truncated data, which does not depend on the functional form of the distribution of the residuals. We thereupon define the corresponding estimator that can be interpreted as the solution of a population analogy to the above identification lemma. Suppose that $\{Y_i, \mathbf{X}_i\}_{i=1}^n$ are the observed i.i.d. samples from the conditional distribution of $(Y^*, \mathbf{X})$ given the event $Y^* \geq y^* = 0$, we can maximize the following function to estimate $\boldsymbol{\beta}$

$$Q_n(\boldsymbol{\beta}) = \frac{1}{nh_0}\sum_{i=1}^{n} K\left(\frac{Y_i - \mathbf{X}_i^T\boldsymbol{\beta}}{h_0}\right), \tag{2.2}$$

where $K(\cdot)$ is a kernel function and h_0 is a bandwidth. As argued by Yao and Li (2014) and Ullah et al. (2021, 2022, 2023), the choice of kernel function does not play much important role in mode estimation. We thus choose

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Gaussian kernel for $K(\cdot)$ in this paper for conducting numerical analysis. Further elaboration is provided in the supplementary file to elucidate how (2.2) facilitates the attainment of the mode-based estimator.

Maximizing the objective function (2.2) can be equivalently formulated as solving the following estimating equation

$$\frac{1}{nh_0^2}\sum_{i=1}^{n} K^{(1)}\left(\frac{Y_i - \mathbf{X}_i^T\hat{\boldsymbol{\beta}}}{h_0}\right)\mathbf{X}_i = \mathbf{0}_{p\times 1}, \tag{2.3}$$

where $\hat{\boldsymbol{\beta}}$ is the resulting mode-based estimator and $K^{(1)}(\cdot)$ represents the first derivative with respect to $\boldsymbol{\beta}$. Note that $K^{(1)}(\cdot)$ is a bounded score function since it will go to zero when the tuning parameter h approaches infinity, providing support for kernel mode-based estimation to achieve robustness. Because the corresponding asymptotic result of $\hat{\boldsymbol{\beta}}$ can be treated as a special case of Theorem 2 in Section 3, we omit the detailed theoretical illustration here. In the supplementary file, we demonstrate that the proposed kernel mode-based estimation can produce more accurate estimates compared to least squares estimation through a simulation example.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

3. Mode-Based Estimation under Random Truncation

In contrast to the fixed truncation discussed in Section 2, in this section, we approximately treat random truncation effects under mode content and utilize truncated data to estimate the parameter $\boldsymbol{\beta}$ defined in (2.1).

3.1 Model and Identification

Let $(Y_k, T_k, \mathbf{X}_k)$, $1 \leq k \leq N$, be a sequence of i.i.d. random vectors, where Y_k is independent of T_k and the sample size N is deterministic but unknown. Suppose that Y and T have, respectively, unknown distribution functions F and G. Under random truncation, some observations would be missing and only a subsequence $\{(Y_{k_i}, T_{k_i}, \mathbf{X}_{k_i}) : 1 \leq i \leq n\}$ can be observed. The size of the actually observed sample, $n \leq N$, is binomially distributed, i.e., $0 < \alpha = P(Y \geq T) < 1$. Without possible confusion, we shall denote the observable subsequence by $(U_i, V_i, \mathbf{W}_i) : i = 1, \cdots, n$ subject to $U_i \geq V_i$, where $U_i = Y_{k_i}$, $V_i = T_{k_i}$, and $\mathbf{W}_i = \mathbf{X}_{k_i}$. Conditional on the value of n, those observed random variables are still i.i.d.. Note that if we multiple each variable by -1, the model can be converted to a right-truncated regression.

Let $F(y) = P(Y \leq y)$, $G(t) = P(T \leq t)$, and $F(y, \mathbf{x}) = P(Y \leq y, \mathbf{X} \leq \mathbf{x})$. Denote (a_F, b_F) as the support of Y or F, where $a_F = \inf\{y : F(y) > 0\}$ and $b_F = \sup\{y : F(y) < 1\}$, and (a_G, b_G) as the support of T or G, where

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

$a_G = \inf\{t : G(t) > 0\}$ and $b_G = \sup\{t : G(t) < 1\}$. Random truncation restricts the observation range of Y and T, under which the conditional distributions $F_0(y) = P(Y \le y \mid Y \ge a_G)$ and $G_0(t) = P(T \le t \mid T \le b_F)$ can be estimated nonparametrically. To ensure $F_0 = F$, we assume that $a_G \le a_F$ is satisfied, whereas G is identifiable, i.e., $G = G_0$, only when $b_G \le b_F$ and $\int_{a_F}^{\infty} dF/G < \infty$ (necessary but not sufficient); see Woodroofe (1985).

We denote any distribution function that is affiliated with the truncated random variables by a superscript * in what follows. Because the original data $(Y_k, T_k, \mathbf{X}_k), 1 \le k \le N$ are i.i.d., the observed data $\{U_i, V_i, \mathbf{W}_i\}_{i=1}^n$ still remain i.i.d. with a common distribution $F^*(u, v, \mathbf{w}) = P[U \le u, V \le v, \mathbf{W} \le \mathbf{w}] = P[Y \le u, T \le v, \mathbf{X} \le \mathbf{w} \mid Y \ge T] = \alpha^{-1} \int_{a_G \le x \le u} \int_{z \le w} G(\mathbf{x} \wedge v)\, dF(\mathbf{x}, z)$, where $\mathbf{x} \wedge v = \min(\mathbf{x}, v)$. According to Stute (1993) and He and Yang (2003), the distribution functions of U, V, and $\mathbf{W}$ involved with a truncated random variable are defined as $F^*(u) = P(U \le u) = \frac{1}{\alpha} \int_{-\infty}^{u} G(y)\, F(dy)$, $G^*(v) = P(V \le v) = \frac{1}{\alpha} \int_{-\infty}^{\infty} G(v \wedge y) F(dy)$, $F^*(u, \mathbf{w}) = P(U \le u, \mathbf{W} \le \mathbf{w}) = \frac{1}{\alpha} \int_{-\infty}^{u} \int_{-\infty}^{w} G(y) F(dy, d\mathbf{x})$, and are estimated by their corresponding empirical distribution functions

$$
\begin{aligned}
F_n^*(u) &= n^{-1} \sum_{i=1}^{n} I_{\{U_i \le u\}}, \quad G_n^*(v) = n^{-1} \sum_{i=1}^{n} I_{\{V_i \le v\}}, \\
F_n^*(u, \mathbf{w}) &= n^{-1} \sum_{i=1}^{n} I_{\{U_i \le u, \mathbf{W}_i \le \mathbf{w}\}},
\end{aligned}
\tag{3.1}
$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

in which $I_{\{\cdot\}}$ is the indicator function. Define $R(\cdot)$ by $R(y) = G^*(y) - F^*(y)$ $= \alpha^{-1} G(y) \left[1 - F\left(y^-\right)\right]$ with the corresponding empirical estimator being

$$R_n(y) = n^{-1} \sum_{i=1}^{n} I_{\{V_i \le y \le U_i\}} = G_n^*(y) - F_n^*\left(y^-\right), \tag{3.2}$$

where $F\left(y^-\right)$ represents the left-continuous version of $F(y)$.

Constructing a mode-based estimate for $\boldsymbol{\beta}$ requires estimating the unknown terms $F(y)$, $G(t)$, and α, which are crucial components of the mode-based estimation framework. According to Woodroofe (1985), the product-limit estimates, F_n and G_n provided below, are asymptotically optimal nonparametric estimators for $F(y)$ and $G(t)$

$$F_n(y) = 1 - \prod_{U_i \le y} \left[1 - \frac{F_n^*\{U_i\}}{R_n(U_i)}\right] \text{ and } G_n(t) = 1 - \prod_{V_i > t} \left[1 - \frac{G_n^*\{V_i\}}{R_n(V_i)}\right], \tag{3.3}$$

where the curly bracket $g\{t\}$ denotes the difference $g(t) - g(t^-)$, in which $g(t^-)$ represents the left-continuous version of $g(t)$, and an empty product is set to equal one. Without ties among the U's and V's, (3.3) simplifies to become $F_n(y) = 1 - \prod_{U_i \le y} \left[1 - \frac{1}{nR_n(U_i)}\right]$ and $G_n(t) = 1 - \prod_{V_i > t} \left[1 - \frac{1}{nR_n(V_i)}\right]$.

To estimate $\boldsymbol{\beta}$ in (2.1), we shall first estimate α by cause of the unknown

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

value of N. Following He and Yang (1998), we use the estimator

$$\alpha_n(y) = \frac{G_n(y)\left[1 - F_n\left(y^-\right)\right]}{R_n(y)} \tag{3.4}$$

to estimate α for any y such that $R_n(y) > 0$, which is shown to be independent of Y. Furthermore, they showed $\alpha_n \to \alpha$ almost surely as $n \to \infty$. Thereafter, one can obtain the nonparametric estimate of $F(y, \mathbf{x})$ as follows

$$F_n(y, \mathbf{x}) = \alpha_n \int_{u \le y} \int_{\mathbf{w} \le \mathbf{x}} \frac{1}{G_n(u)} F_n^*(du, d\mathbf{w}). \tag{3.5}$$

We then come back to our main problem, which is to estimate the mode-based coefficient $\boldsymbol{\beta}$ under random left truncation. According to Kemp and Santos Silva (2012) and Yao and Li (2014), when we observe $\{Y_i, \mathbf{X}_i\}_{i=1}^n$ without truncation, we should maximize (2.2) to obtain the mode-based estimate, which can be clearly reexpressed as in the integral form

$$\int \frac{1}{h} K\left(\frac{y - \mathbf{x}^T\boldsymbol{\beta}}{h}\right) d\hat{F}_n(y, \mathbf{x}), \tag{3.6}$$

where $\hat{F}_n(y, \mathbf{x})$ is the empirical distribution of $\{Y_i, \mathbf{X}_i\}_{i=1}^n$. In the left trunca-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

tion case, we replace $\hat{F}_n(y, \mathbf{x})$ in (3.6) by $F_n(y, \mathbf{x})$ defined in (3.5) and obtain

$$\int \frac{1}{h} K\left(\frac{y - \mathbf{x}^T\boldsymbol{\beta}}{h}\right) dF_n(y, \mathbf{x}) = \int_{u \le y} \int_{\mathbf{w} \le \mathbf{x}} \frac{\alpha_n}{G_n(u)} \frac{1}{h} K\left(\frac{u - \mathbf{w}^T\boldsymbol{\beta}}{h}\right) dF_n^*(u, \mathbf{w}).$$

Based on the above results, we can finally estimate the unknown parameter vector $\boldsymbol{\beta}$ by maximizing a weighted kernel objective function

$$Q_n(\boldsymbol{\beta}) = \frac{\alpha_n}{nh} \sum_{i=1}^{n} \frac{1}{G_n(U_i)} K\left(\frac{U_i - \mathbf{W}_i^T\boldsymbol{\beta}}{h}\right), \tag{3.7}$$

where the resulting mode-based estimator is defined as $\hat{\boldsymbol{\beta}}$. It is interesting to point out that the objective function (3.7) reduces to the traditional kernel mode-based objective function if no (or fixed) truncation (i.e., $\alpha = 1$) is present in the data. We subsequently have the following theorem.

Theorem 1. *The kernel mode-based objective function* (3.7) *can be utilized to achieve the estimator of* $Q_N(\boldsymbol{\beta}) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{Y_i - \mathbf{X}_i^T\boldsymbol{\beta}}{h}\right)$ *without truncation as* $n \to \infty$, *that is,* $|Q_n(\boldsymbol{\beta}) - Q_N(\boldsymbol{\beta})| = o_p(1)$ *and*

$$\mathbb{E}\left[\frac{\alpha}{h} \frac{1}{G(U)} K\left(\frac{U - \mathbf{W}^T\boldsymbol{\beta}}{h}\right) \mid \mathbf{X}\right] = \mathbb{E}\left[\frac{1}{h} K\left(\frac{Y - \mathbf{X}^T\boldsymbol{\beta}}{h}\right) \mid \mathbf{X}\right],$$

which implies that the true parameter vector $\boldsymbol{\beta}_0$ *under the left truncated assumption satisfies that* $\boldsymbol{\beta}_0 = \arg\max \mathbb{E}\left[\frac{1}{h} \frac{\alpha}{G(U)} K\left(\frac{U - \mathbf{W}^T\boldsymbol{\beta}}{h}\right)\right] = \mathbb{E}\left[\frac{1}{h} K\left(\frac{Y - \mathbf{X}^T\boldsymbol{\beta}}{h}\right)\right]$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Algorithm 1: MEM Algorithm under Random Truncation

Data: Sample observations $\{(U_i, V_i, \mathbf{W}_i)\}_{i=1}^n$ and bandwidth h.
Result: Final kernel mode-based estimate $\hat{\boldsymbol{\beta}}$.
while *two consecutive solutions are not close enough, i.e.,* $\|\hat{\boldsymbol{\beta}}^{(m)} - \hat{\boldsymbol{\beta}}^{(m-1)}\| > 10^{-4}$ **do**
  if *current estimate* $\hat{\boldsymbol{\beta}}^{(m)}$ *with iterative indicator* $m \geq 1$ **then**
    E-Step: Calculate weight $\pi(i \mid \hat{\boldsymbol{\beta}}^{(m)})$ with

$$\frac{[G_n(U_i)]^{-1} K\left(\frac{U_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}}^{(m)}}{h}\right)}{\sum_{i=1}^n [G_n(U_i)]^{-1} K\left(\frac{U_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}}^{(m)}}{h}\right)} \propto \frac{1}{G_n(U_i)} K\left(\frac{U_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}}^{(m)}}{h}\right),$$

    which is nonnegative and sums to one.
    M-Step: Update the estimate with log-maximization

$$\begin{aligned}&\arg\max \sum_{i=1}^n \pi(i \mid \hat{\boldsymbol{\beta}}^{(m)}) \log\left[\frac{1}{G_n(U_i)} K\left(\frac{U_i - \mathbf{W}_i^T \boldsymbol{\beta}}{h}\right)\right] \\ =&(\mathbf{W}^T \Phi \mathbf{W})^{-1} \mathbf{W}^T \Phi \mathbf{U},\end{aligned}$$

    where $\mathbf{U} = (U_1, \cdots, U_n)^T$ is an $n \times 1$ vector, $\mathbf{W} = (\mathbf{W}_1^T, \cdots, \mathbf{W}_n^T)^T$ is an $n \times p$ matrix, and Φ is an $n \times n$ diagonal matrix with diagonal elements $\{\pi(i \mid \hat{\boldsymbol{\beta}}^{(m)})\}_{i=1}^n$.
  end
end

Different from mean regression, we do not have an explicit expression for the proposed mode-based estimator. To numerically estimate the model, we suggest a modified MEM algorithm shown above, which includes E-Step for calculating weights and M-Step for maximizing the log-objective function. The rationale behind this algorithm is that the outliers are likely to suffer large residuals and hence get downweighted through E-Step. Following Yao

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

and Li (2014), we can show that the log-$Q_n(\boldsymbol{\beta})$ in M-Step does not decrease after each iteration. As a result, the sequence of estimates generated by the algorithm will monotonically converge towards at least a local maximum. On account of the use of Gaussian kernel, the updated values of unknown parameters can have a closed form in M-Step, which renders the MEM algorithm highly stable and flexible. The simulation studies in Section 5 indicate that the convergence is typically achieved within 50 iterations. In practice, the initial value $\hat{\boldsymbol{\beta}}^{(0)}$ can be chosen as the median or Huber estimate. Notice that for numerical estimation, we allow for the existence of local modes in the data (see simulation study in the supplementary file). To ensure the achieving of global maximum, we can try different initial estimates and compare the values of kernel mode-based objective function.

3.2 Asymptotic Properties

The development of statistical properties for the resulting mode-based estimator is not trivial, as there exist truncation effects in the objective function and the bandwidth h is treated as a constant. To investigate the theoretical proprieties, we impose some technical conditions outlined below.

C1 The true value of parameter $\boldsymbol{\beta}_0$ lies within the interior of the known compact parameter space, which is a subset of $\mathbb{R}^p$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

C2 The distribution functions F and G are continuous and $a_G \leq a_F$. Additionally, $\int \frac{dF}{1-F} < \infty$, $\int \frac{dF}{G^2} < \infty$, and $\mathbb{E}[\|\mathbf{X}\|^2/G(Y)] < \infty$, where the integral sign $\int$ denotes integration from $-\infty$ to $+\infty$.

C3 The kernel function $K(\cdot) : \mathbb{R} \to \mathbb{R}$ is a nonnegatively symmetric density function with bounded support and integrates to one.

C4 There exists a constant $C > 0$ such that $\mathbb{E}\{\sup_{\mathcal{Y}:|\mathcal{Y}-\varepsilon|<C}|K_h^{(3)}(\mathcal{Y})|\} < \infty$, where $K_h(\cdot) = h^{-1}K(\cdot/h)$ is the rescaled kernel and $K_h^{(c)}(\cdot)$ denotes the cth derivative of $K_h(\cdot)$.

C5 The first derivative of the kernel objective function satisfies $\mathbb{E}[K_h^{(1)}(\varepsilon) \mid \mathbf{X}] = 0$. Also, the functions $\mathbb{E}[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]$ and $\mathbb{E}\{[K_h^{(1)}(\varepsilon)]^2/G(Y) \mid \mathbf{X}\}$ are all continuous in relation to $\mathbf{X}$. Furthermore, $\mathbb{E}[K_h^{(2)}(\varepsilon) \mid \mathbf{X}] < 0$ and $\mathbb{E}\{[K_h^{(1)}(\varepsilon)]^2/G(Y) \mid \mathbf{X}\}$ is finite for any $h > 0$.

C6 The bandwidth h is a constant and is independent of sample size n.

C7 There is a constant $s > 2$ such that $\mathbb{E}\|\mathbf{X}\|^{2s} < \infty$. Also, $\mathbb{E}\{[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]\mathbf{X}\mathbf{X}^T\}$ is finite and non-singular.

Due to space limitations, comments pertaining to the aforementioned conditions are included in the supplementary file. After that, the following

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

asymptotic results can be obtained, where for $j, k = 1, \cdots, p$, we denote

$$\sigma_{jk} = \alpha \Big\{ \int \frac{K_h^{(1)}(\varepsilon) X_j K_h^{(1)}(\varepsilon) X_k}{G(Y)} F(dY, dX) + \int \frac{\int_{Y \le s} K_h^{(1)}(\varepsilon) X_j F(dY, dX) \int_{Y \le s} K_h^{(1)}(\varepsilon) X_k F(dY, dX)}{[1 - F(s)] G^2(s)} G(ds) \Big\}.$$

Theorem 2. *With the conditions C1-C7 held, when $n \to \infty$,*

(i) the mode-based maximum of $Q_n(\boldsymbol{\beta})$ occurs at $\hat{\boldsymbol{\beta}}$ such that

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n^{-1/2});$$

(ii) the mode-based estimator satisfying the consistency result in (i) has the following asymptotic property

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left\{[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]\mathbf{X}\mathbf{X}^T\right\}^{-1} \Sigma \mathbb{E}\left\{[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]\mathbf{X}\mathbf{X}^T\right\}^{-1}\right),$$

where "$\xrightarrow{d}$" denotes convergence in distribution and Σ is a $p \times p$ positive definite matrix with the element σ_{jk}.

When Y is absent of truncation (or has fixed truncation), that is all data can be fully observed, then $G_n(U_i) = 1$ and $\alpha = 1$. In this case, the

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

asymptotic normality in Theorem 2 would be reduced to

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}\left(0, \{\mathbb{E}[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]\}^{-2}\mathbb{E}\{[K_h^{(1)}(\varepsilon)]^2 \mid \mathbf{X}\}\mathrm{Cov}^{-1}(\mathbf{X})\right)$$

with a finite covariance matrix $\mathrm{Cov}(\mathbf{X})$. It can be seen that the resulting large sample properties do not depend on any moment conditions on the random error, which enables resulting mode-based estimator to be robust against outliers or heavy-tailed distributions. This intuitively suggests that $\hat{\boldsymbol{\beta}}$ enjoys a $\sqrt{n}$-consistency property even when $\mathrm{Var}(\varepsilon)$ is infinite.

To select the bandwidth h in a data-driven way, we employ a cross-validation (CV) approach specifically tailored to the kernel mode-based function

$$\mathrm{CV}\,(h) = \sum_{i=1}^{n} K_h\left(U_i - \mathbf{W}_i^T\hat{\boldsymbol{\beta}}_{[-i]}\right), \tag{3.8}$$

where $\hat{\boldsymbol{\beta}}_{[-i]}$ is the solution from (3.7) after deleting the ith subject. When utilizing a Gaussian kernel for $K(\cdot)$, (3.8) can be interpreted as leaving out the ith projection error from the estimation process. We shall demonstrate that $\mathrm{CV}(h)$ performs well in our numerical examples listed in Section 5.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

3.3 Mode-Based Empirical Likelihood

In small samples, normal approximation confidence intervals for $\boldsymbol{\beta}$ might lack accuracy, as certain unknown terms in the asymptotic variance need to be estimated. Alternatively, we propose an empirical likelihood method to construct confidence intervals by establishing mode-based auxiliary random vectors, which has the ability to restrict side information and automatically determine the geometry of confidence intervals; see Chen and Van Keilegom (2009). This empirical likelihood approach enables us to make inferences for any linear combination of coefficients.

Taking into account the influence of truncated data, for a fixed $\boldsymbol{\beta}$, we establish a weighted auxiliary random vector as follows

$$\Xi_i(\boldsymbol{\beta}) = G^{-1}(U_i)\, K_h^{(1)}(\varepsilon_i)\, \mathbf{W}_i. \tag{3.9}$$

Let $p_1, p_2, \cdots, p_n$ denote nonnegative numbers summing to unity. Building upon the result from Theorem 1 such that $\mathbb{E}[\alpha G^{-1}(U_i)K_h^{(1)}(\varepsilon_i)\mathbf{W}_i \mid \mathbf{X}_i] = \mathbb{E}[K_h^{(1)}(\varepsilon_i)\mathbf{X}_i \mid \mathbf{X}_i]$, we can get $\mathbb{E}[\alpha\Xi_i(\boldsymbol{\beta}) \mid \mathbf{X}_i] = \alpha\mathbb{E}[\Xi_i(\boldsymbol{\beta}) \mid \mathbf{X}_i] = \mathbb{E}[K_h^{(1)}(\varepsilon_i)\ \mathbf{X}_i \mid \mathbf{X}_i] = \mathbf{X}_i\mathbb{E}[K_h^{(1)}(\varepsilon_i) \mid \mathbf{X}_i]$. With the condition $\mathbb{E}[K_h^{(1)}(\varepsilon_i) \mid \mathbf{X}_i] = 0$ for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in condition C5, we can arrive at $\mathbb{E}[\Xi_i(\boldsymbol{\beta}) \mid \mathbf{X}_i] = 0$. This result holds if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, which follows from the fact that

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

$\mathbb{E}[K_h^{(1)}(\varepsilon_i) \mid \mathbf{X}_i] = 0$ is unique to the true parameter $\boldsymbol{\beta}_0$. Thereupon, the empirical log-likelihood ratio statistic of $\boldsymbol{\beta}$ can be defined as

$$\mathcal{L}_0(\boldsymbol{\beta}_0) = -2\max\left\{\sum_{i=1}^{n}\log(np_i) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\Xi_i(\boldsymbol{\beta}_0) = 0\right\}.$$

Since $\mathcal{L}_0(\boldsymbol{\beta}_0)$ contains unknown terms, a natural way is to replace them by their estimates. Using the standard Lagrange multiplier method such that $\sum_{i=1}^{n}\log(np_i) - \gamma(1-\sum_{i=1}^{n}p_i) - \boldsymbol{\lambda}_{\boldsymbol{\beta}_0}^T\sum_{i=1}^{n}p_i\Xi_i(\boldsymbol{\beta}_0)$, where γ and $\boldsymbol{\lambda}_{\boldsymbol{\beta}_0}$ are Lagrange multipliers, we can obtain $\gamma = -n$ and $p_i = 1/\{n[1+\boldsymbol{\lambda}_{\boldsymbol{\beta}}^T\Xi_i(\boldsymbol{\beta}_0)]\}$. Therefore, the empirical log-likelihood ratio evaluated at true parameter value is

$$\mathcal{L}(\boldsymbol{\beta}_0) = 2\sum_{i=1}^{n}\log\left\{1+\boldsymbol{\lambda}_{\boldsymbol{\beta}_0}^T\Xi_i(\boldsymbol{\beta}_0)\right\}, \tag{3.10}$$

where $\boldsymbol{\lambda}_{\boldsymbol{\beta}_0}$, a $p\times 1$ vector of Lagrange multipliers, is the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Xi_i(\boldsymbol{\beta}_0)}{1+\boldsymbol{\lambda}_{\boldsymbol{\beta}}^T\Xi_i(\boldsymbol{\beta}_0)} = 0 \tag{3.11}$$

by putting $p_i = 1/\{n[1+\boldsymbol{\lambda}_{\boldsymbol{\beta}}^T\Xi_i(\boldsymbol{\beta}_0)]\}$ into the constraint $\sum_{i=1}^{n}p_i\Xi_i(\boldsymbol{\beta}_0) = 0$. In addition, with the constraint $p_i \geq 0$ and $\sum_{i=1}^{n}p_i = 1$, the function $\sum_{i=1}^{n}\log(np_i) = n\log(n) + \sum_{i=1}^{n}\log(p_i)$ is maximized by $p_i = 1/n$, which is derived by solving the Lagrange function $\sum_{i=1}^{n}\log(p_i) - \gamma(1-\sum_{i=1}^{n}p_i)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Combining this with the result $p_i = 1/\{n[1 + \boldsymbol{\lambda}_{\boldsymbol{\beta}}^T \Xi_i(\boldsymbol{\beta}_0)]\}$, we can arrive at $\boldsymbol{\lambda}_{\boldsymbol{\beta}}^T \Xi_i(\boldsymbol{\beta}_0) = 0$. Therefore, according to (3.11), we have $n^{-1} \sum_{i=1}^n \Xi_i(\boldsymbol{\beta}_0) = 0$, leading to $\mathbb{E}[\Xi_i(\boldsymbol{\beta}_0)] = 0$ in expectation, given that $\Xi_i(\boldsymbol{\beta}_0)$ are independent random variables. Consequently, $\mathcal{L}(\boldsymbol{\beta}_0)$ will be maximized by $\hat{\boldsymbol{\beta}}$ obtained from (3.7). We then have the following theorem.

Theorem 3. *(Wilks' Theorem) Suppose the same conditions as Theorem 2 are satisfied. As $n \to \infty$, the limiting distribution of $\mathcal{L}(\boldsymbol{\beta}_0)$ is $\mathcal{L}(\boldsymbol{\beta}_0) \xrightarrow{d} \chi_p^2$, where χ_p^2 is the chi-square distribution with p degrees of freedom.*

Theorem 3 indicates that the proposed empirical log-likelihood ratio has a chi-square distribution, which is free of any tuning parameter. Therefore, the confidence intervals for $\boldsymbol{\beta}_0$ with asymptotically coverage probability $1 - \alpha_c$ can be defined as $I_{n,\alpha_c} = \{\boldsymbol{\beta} : \mathcal{L}(\boldsymbol{\beta}) \leq C_{\alpha_c}\}$ with C_{α_c} satisfying $P(\chi_p^2 \leq C_{\alpha_c}) = 1 - \alpha_c$. The confidence intervals do not depend on an explicit estimate of $\mathbb{E}\{[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]\mathbf{X}\mathbf{X}^T\}^{-1} \Sigma \mathbb{E}\{[K_h^{(2)}(\varepsilon) \mid \mathbf{X}]\mathbf{X}\mathbf{X}^T\}^{-1}$, which is a big advantage compared to confidence intervals built on Wald-type statistics.

4. Mode-Based Variable Selection under Random Truncation

In practice, it is frequently uncertain what the true model is, and the covariates may contain a large amount of extraneous information. To address this challenge, in this section, we propose a shrinkage procedure for (3.7), which

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

allows us to conduct estimation and variable selection simultaneously.

4.1 Mode-Based Penalized Estimation

Benefiting from the favorable properties of the SCAD penalty, we integrate mode-based estimation with the SCAD penalty for randomly truncated data, which is defined on $\mathbb{R}^+$ with respect to its first derivative

$$p_\lambda^{(1)}(t) = \lambda \left\{ I(t \leq \lambda) + \frac{(a\lambda - t)_+}{(a-1)\lambda} I(t > \lambda) \right\}, \ t > 0, \tag{4.1}$$

where $a > 2$ and $\lambda > 0$ are tuning parameters to control the amount of shrinkage, $I(\cdot)$ is the indicator function, and $(t)_+$ denotes the positive value of t. The larger the value of λ, the greater the amount of shrinkage. According to Fan and Li (2001), the SCAD penalty satisfies the conditions for unbiasedness, sparsity, and continuity, but is not differentiable at zero.

After incorporating the SCAD penalty into (3.7), the resulting penalized kernel mode-based objective function is established as follows

$$Q_n^p(\boldsymbol{\beta}) = \frac{\alpha_n}{nh} \sum_{i=1}^{n} \frac{1}{G_n(U_i)} K\left(\frac{U_i - \mathbf{W}_i^T \boldsymbol{\beta}}{h} \right) - \sum_{j=1}^{p} p_\lambda(|\beta_j|). \tag{4.2}$$

By maximizing the above objective function with an appropriate penalty parameter λ, we can obtain a sparse mode-based estimator of $\boldsymbol{\beta}$, denoted

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

as $\hat{\boldsymbol{\beta}}^p$, thereby facilitating variable selection. It is worth noting that the tuning parameter λ need not be uniform across all coefficients. We can easily set $\lambda_j = \lambda[\mathrm{Var}(\beta_j)]^{1/2}$ for $j = 1, \cdots, p$ to allow for different penalties on individual parameters. For simplicity, we concentrate on the illustration with a common λ. Additionally, alternative penalty functions can also be utilized here and lead to similar consistency results as outlined below.

We establish the oracle property of the SCAD penalized mode-based regression. Without loss of generality, we partition the true parameter vector as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, where $\boldsymbol{\beta}_1 \in \mathbb{R}^s$ comprises all nonzero components and $\boldsymbol{\beta}_2 \in \mathbb{R}^{p-s}$ contains all zero parameters. Similarly, the covariates are partitioned into two sets, with $\mathbf{X}_1$ representing the covariates corresponding to the first s elements of $\mathbf{X}$. We can then present the following theorem.

Theorem 4. *With the conditions C1-C7 satisfied, if* $\lambda \to 0$ *and* $\sqrt{n}\lambda \to \infty$ *as* $n \to \infty$*, we have*

(i) Selection Consistency: with probability tending to one, $\hat{\boldsymbol{\beta}}_2^p = \mathbf{0}_{(p-s)\times 1}$.

(ii) Asymptotic Normality: $\sqrt{n}(\hat{\boldsymbol{\beta}}_1^p - \boldsymbol{\beta}_{1,0}) \xrightarrow{d} \mathcal{N}(0, \Delta^{-1}\Sigma_1\Delta^{-1})$*, where* $\Delta = \mathbb{E}\{[K_h^{(2)}(\varepsilon) \mid \mathbf{X}_1]\mathbf{X}_1\mathbf{X}_1^T\}$ *and* Σ_1 *is a* $s \times s$ *positive definite matrix with the element* σ_{jk} *associated with covariates* $\mathbf{X}_1$.

The penalty function becomes singular at the origin due to the condition $\lambda \to 0$, which in turn offers the penalized estimator sparsity property.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

The condition $\sqrt{n}\lambda \to \infty$ implies that if the tuning parameter converges to zero at a speed slower than $n^{-1/2}$, the corresponding penalized estimator can be $n^{1/2}$-consistent. Therefore, the suggested method has the ability to consistently produce spare solutions for mode-based regression coefficients.

4.2 Computational Algorithm

To implement the aforementioned variable selection procedure, we need to obtain the appropriate tuning parameters a and λ in the process of computation, which can control the degree of robustness and efficiency of the proposed estimator. To reduce intensive computation and guarantee consistent variable selection, we follow Fan and Li (2001) to choose $a = 3.7$ from the Bayesian point of view and utilize the following extended BIC to select λ

$$\mathrm{BIC}(\lambda) = \log\left[\frac{\alpha_n}{nh}\sum_{i=1}^{n}\frac{1}{G_n\left(U_i\right)}K\left(\frac{U_i - \mathbf{W}_i^T\hat{\boldsymbol{\beta}}^p}{h}\right)\right] - \frac{\log(n)}{n}df_\lambda, \qquad (4.3)$$

where df_λ is the degrees of freedom, i.e., the number of nonzero elements of $\hat{\boldsymbol{\beta}}^p$ for any candidate penalty parameter λ. The suggested BIC takes into account both the number of unknown parameters and the complexity of the model space, where the first term in (4.3) is an "artificial" likelihood, sharing essential properties of a parametric log-likelihood, and the second

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

term measures the model complexity. Thereby, the tuning parameter λ can be chosen as $\arg\max_\lambda \mathrm{BIC}(\lambda)$, which is supported by the following theorem.

Theorem 5. *Under the conditions C1-C7, the tuning parameter $\hat{\lambda}$ selected by $BIC(\lambda)$ can choose the true model with probability approaching one.*

Despite the excellent statistical properties of the SCAD penalized mode-based estimator, the maximization of the SCAD penalized objective function is not easy because it is irregular at the origin and does not have a second derivative at some points. To circumvent this difficulty, we take the local quadratic approximation for the SCAD penalty suggested by Fan and Li (2001). Suppose that we can get an estimate $\beta_j^{(m)}$ in the mth step that is close to the true parameter β_j. If $\beta_j^{(m)}$ is near 0, then set $\hat{\beta}_j = 0$. Otherwise, the SCAD penalty can be locally approximated by a quadratic function as

$$p_\lambda(|\beta_j|) \approx p_\lambda(|\beta_j^{(m)}|) + \frac{1}{2}\{p_\lambda^{(1)}(|\beta_j^{(m)}|)/|\beta_j^{(m)}|\}(\beta_j^2 - \beta_j^{(m)2}) \text{ for } \beta_j \approx \beta_j^{(m)},$$

where $p_\lambda^{(1)}(\cdot)$ represents the first derivative of $p_\lambda(\cdot)$. Following the same procedures outlined in Algorithm 1, we can propose a modified MEM algorithm for SCAD penalized mode-based regression by replacing $p_\lambda(|\beta_j|)$ with the above equation and ignoring irrelevant constants; see Algorithm 2. Starting from an initial estimate, we iterate the E-Step and M-Step until

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

a convergence criterion is met. The convergence property of the proposed penalized MEM Algorithm 2 is investigated in the supplementary file.

Algorithm 2: MEM Algorithm for Variable Selection

Data: Sample observations $\{(U_i, V_i, \mathbf{W}_i)\}_{i=1}^n$, bandwidth h, and tuning parameter λ.

Result: Final penalized kernel mode-based estimate $\hat{\boldsymbol{\beta}}^p$.

while *two consecutive solutions are not close enough, i.e.,* $\|\hat{\boldsymbol{\beta}}^{p(m)} - \hat{\boldsymbol{\beta}}^{p(m-1)}\| > 10^{-4}$ **do**

  if *current estimate* $\hat{\boldsymbol{\beta}}^{p(m)}$ *with iterative indicator* $m \geq 1$ **then**

    E-Step: Calculate weight $\pi(i \mid \hat{\boldsymbol{\beta}}^{p(m)})$ with

$$\frac{(G_n(U_i))^{-1} K\left(\frac{U_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}}^{p(m)}}{h}\right)}{\sum_{i=1}^n (G_n(U_i))^{-1} K\left(\frac{U_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}}^{p(m)}}{h}\right)} \propto \frac{1}{G_n(U_i)} K\left(\frac{U_i - \mathbf{W}_i^T \hat{\boldsymbol{\beta}}^{p(m)}}{h}\right),$$

    which is nonnegative and sums to one.

    M-Step: Update the estimate with log-maximization

$$\arg\max \sum_{i=1}^n \pi(i \mid \hat{\boldsymbol{\beta}}^{p(m)}) \log\left[\frac{1}{G_n(U_i)} K\left(\frac{U_i - \mathbf{W}_i^T \boldsymbol{\beta}}{h}\right)\right]$$
$$- \frac{n}{2} \sum_{j=1}^p \left\{\frac{p_\lambda^{(1)}(|\hat{\beta}_j^{p(m)}|)}{|\hat{\beta}_j^{p(m)}|}\right\} \beta_j^2 = (\mathbf{W}^T \Phi \mathbf{W} + n\Sigma_\lambda(\hat{\boldsymbol{\beta}}^{p(m)}))^{-1} \mathbf{W}^T \Phi \mathbf{U},$$

    where Φ is an $n \times n$ diagonal matrix with diagonal elements $\pi(\cdot)$, and $\Sigma_\lambda(\hat{\boldsymbol{\beta}}^{(m)}) = \text{diag}\{p_\lambda^{(1)}(|\hat{\beta}_1^{p(m)}|)/|\hat{\beta}_1^{p(m)}|, \cdots, p_\lambda^{(1)}(|\hat{\beta}_p^{p(m)}|)/|\hat{\beta}_p^{p(m)}|\}$ for nonvanished $\hat{\boldsymbol{\beta}}^{(m)}$.

  end

end

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

5. Numerical Examples

We present numerical examples to illustrate the finite sample performance of the suggested estimation and variable selection procedures in this section. Note that R_n in (3.2) affects the product-limit estimates of F and G, and may approach zero within the range of the data due to the nature of random truncation, resulting in unreasonable estimates of $F_n(y)$ and $G_n(t)$. In accordance with the approach in Woodroofe (1985), we solve this problem by substituting $\max\{R_n(y), 1/n + 1/n^2\}$ for R_n in the subsequent analysis.

5.1 Monte Carlo Experiments

We assume that N is fixed and n, the observed sample size, is random under different data generating process (DGP). Certainly, one can also assume that n is fixed and N is random. The total number of generated random samples is $\{200, 400, 600, 1000\}$ and the number of replications is 400. The estimators from oracle estimation, the proposed estimation, and the naive estimation using observations which are assumed not truncated are compared. To illustrate the robustness and efficiency of the proposed estimator, we also report the results from Huber (with tuning parameter 1.345), median, and least squares (LS) estimations under random truncation.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

DGP 1 We consider the following linear regression model

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i, \ i = 1, \cdots, N, \tag{5.1}$$

where $\beta_1 = 0$, $\beta_2 = 1$, and X_i is generated from a uniform distribution $U[0, 2]$. The truncating variable T_i is independently generated from a normal distribution with mean 0 and variance 1, resulting in approximately 22% median truncated data. To illustrate robustness and efficiency, the error term ε_i is assumed to follow four different distributions: (i) normal distribution $\mathcal{N}(0, 1)$; (ii) Student's t distribution with 3 degrees of freedom, $t(3)$, representing the heavy-tailed distribution; (iii) Laplace distribution, $L_p(0, 1)$; and (iv) mixture of normal distributions, $0.9\mathcal{N}(0, 1)+0.1\mathcal{N}(0, 10^2)$, where the 10% data from $\mathcal{N}(0, 10^2)$ are most likely to be outliers.

The estimation results are reported in Table 1, where the average estimate, standard error (SE), and mean squared error (MSE) are presented based on 400 simulations. As observed, the performance of the oracle estimator is the best, and the proposed mode-based estimator consistently produces smaller SE and MSE than the naive estimator. Even when compared to Huber and median estimations, the developed kernel mode-based estimation can have better performance in respect of SE and MSE. Furthermore,

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

LS estimator performs slightly better than kernel mode-based estimator with normally distributed errors. This is expected since the Gaussian kernel approximation $\exp(-t^2/h)$ tends to resemble t^2/h for a large bandwidth h. Nevertheless, the suggested mode-based estimator is superior to LS estimator for all other error distributions, reflecting the robustness property of mode-based estimation. All of these results hold true for both smaller ($N = 200$) and larger ($N = 1000$) sample sizes. With increasing sample sizes, the MSEs for all estimators become smaller. Notably, for naive estimator, the bias does not decrease as the sample size N increases. To validate the accuracy of SE, we compare it to the sample standard deviation (SD) of the estimators for both oracle and proposed mode-based regression coefficients. Figure S5 in the supplementary file illustrates that the ratio is close to one, indicating that the proposed estimation method performs reasonably well. We also plot empirical density functions and boxplots of the resulting kernel mode-based estimators in the supplementary file to illustrate the asymptotic normality property. As shown in Figure S6, when the sample sizes grow, the SE exhibits a diminishing pattern in magnitude and the distribution tends to approach normality for all error distributions.

Table 1: Monte Carlo Results-DGP 1 (Estimation)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

	N=200			N=400			N=600			N=1000		
Method	β_2	SE(β_2)	MSE(β_2)	β_2	SE(β_2)	MSE(β_2)	β_2	SE(β_2)	MSE(β_2)	β_2	SE(β_2)	MSE(β_2)
						$\mathcal{N}(0,1)$						
Oracle	0.9989	0.0601	0.0036	0.9989	0.0436	0.0019	0.9994	0.0355	0.0013	1.0042	0.0262	0.0007
Proposed	1.0044	0.0773	0.0060	0.9977	0.0526	0.0028	1.0001	0.0424	0.0018	1.0030	0.0330	0.0011
Naive	1.1816	0.0585	0.0364	1.1838	0.0427	0.0356	1.1832	0.0359	0.0349	1.1845	0.0264	0.0347
Huber	1.0027	0.1336	0.0178	0.9979	0.0926	0.0086	0.9996	0.0749	0.0056	1.0041	0.0574	0.0033
Median	1.0016	0.1645	0.0270	0.9974	0.1093	0.0119	0.9976	0.0898	0.0080	1.0082	0.0728	0.0054
LS	0.9976	0.0632	0.0040	0.9985	0.0442	0.0020	0.9995	0.0364	0.0013	1.0040	0.0271	0.0007
						$t(3)$						
Oracle	1.0056	0.0790	0.0063	0.9922	0.0543	0.0030	1.0006	0.0437	0.0019	1.0032	0.0333	0.0011
Proposed	1.0101	0.0868	0.0076	0.9954	0.0590	0.0035	0.9998	0.0481	0.0023	1.0024	0.0359	0.0013
Naive	1.2469	0.0683	0.0656	1.2437	0.0482	0.0617	1.2411	0.0378	0.0596	1.2443	0.0320	0.0607
Huber	1.0013	0.1057	0.0111	0.9916	0.0723	0.0053	0.9987	0.0626	0.0039	1.0041	0.0464	0.0022
Median	0.9879	0.1586	0.0252	0.9894	0.1059	0.0113	1.0062	0.0901	0.0081	1.0043	0.0701	0.0049
LS	0.9932	0.1774	0.0314	0.9914	0.1186	0.0141	1.0022	0.1004	0.0101	0.9987	0.0754	0.0057
						$L_p(0,1)$						
Oracle	1.0021	0.0737	0.0054	1.0075	0.0527	0.0028	1.0030	0.0414	0.0017	1.0009	0.0322	0.0010
Proposed	1.0005	0.0747	0.0056	1.0063	0.0509	0.0026	1.0024	0.0423	0.0018	0.9999	0.0324	0.0010
Naive	1.2290	0.0650	0.0567	1.2285	0.0476	0.0545	1.2249	0.0377	0.0520	1.2259	0.0295	0.0519
Huber	1.0004	0.0864	0.0075	1.0042	0.0644	0.0042	1.0038	0.0463	0.0022	1.0019	0.0403	0.0016
Median	0.9909	0.1398	0.0196	1.0023	0.0913	0.0083	1.0030	0.0718	0.0052	1.0006	0.0581	0.0034
LS	0.9943	0.1573	0.0247	1.0039	0.1101	0.0121	1.0036	0.0850	0.0072	1.0023	0.0647	0.0042
						$0.9\mathcal{N}(0,1)+0.1\mathcal{N}(0,10^2)$						
Oracle	0.9937	0.0809	0.0066	0.9981	0.0564	0.0032	1.0008	0.0466	0.0022	0.9985	0.0377	0.0014
Proposed	0.9929	0.0890	0.0080	0.9991	0.0594	0.0035	1.0015	0.0508	0.0026	0.9995	0.0410	0.0017
Naive	1.3223	0.0776	0.1099	1.3208	0.0547	0.1059	1.3184	0.0454	0.1034	1.3197	0.0357	0.1035
Huber	0.9906	0.1270	0.0162	0.9997	0.0814	0.0066	1.0016	0.0718	0.0051	1.0005	0.0562	0.0032
Median	0.9918	0.1758	0.0309	1.0049	0.1114	0.0124	1.0010	0.0950	0.0090	0.9931	0.0733	0.0054
LS	0.9946	0.2054	0.0421	1.0156	0.1361	0.0187	1.0010	0.1152	0.0132	0.9913	0.0902	0.0082

To evaluate the performance of mode-based empirical likelihood (EL) estimation, we report the simulated coverage probabilities (CP) of confidence intervals for β_2 and the average lengths (AL) of confidence intervals at the nominal level 95% in Table 2. We also include results from the normal approximation method for comparison. Comparing to the normal approximation procedure, it is evident that the suggested EL estimation consis-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

tently results in shorter AL and produces empirical CP closer to the nominal level 95% regardless of the error distribution, albeit with some slight undercoverage. Conversely, the CP from other estimations with nonnormal errors generally fall below the nominal level. Furthermore, the AL from all estimations decrease as the sample size N increases. When outliers or heavy-tailed distributions are present, the proposed mode-based EL estimation outperforms other methods by producing smaller AL, except for the oracle estimation, which translates into greater power and more precise estimates.

Table 2: Monte Carlo Results-DGP 1 (Empirical Likelihood)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

		$\mathcal{N}(0,1)$		$t(3)$		$L_p(0,1)$		Mixture $\mathcal{N}$	
Method	N	CP	AL	CP	AL	CP	AL	CP	AL
Oracle-EL	200	0.9350	0.4807	0.9425	0.5920	0.9400	0.6368	0.9350	0.6809
	400	0.9400	0.3688	0.9450	0.5135	0.9450	0.5582	0.9425	0.6174
	600	0.9425	0.2835	0.9475	0.4230	0.9450	0.4864	0.9450	0.5232
	1000	0.9475	0.2287	0.9550	0.3606	0.9475	0.4254	0.9450	0.4617
Proposed-EL	200	0.9225	0.4938	0.9400	0.6228	0.9375	0.6746	0.9300	0.7163
	400	0.9375	0.3826	0.9425	0.5409	0.9400	0.5859	0.9350	0.6410
	600	0.9400	0.3266	0.9425	0.4584	0.9425	0.5149	0.9375	0.5546
	1000	0.9450	0.2798	0.9475	0.3724	0.9450	0.4287	0.9400	0.4725
Huber-EL	200	0.9175	0.5107	0.9375	0.6894	0.9300	0.7430	0.9225	0.7826
	400	0.9300	0.4228	0.9400	0.5910	0.9350	0.6638	0.9250	0.7136
	600	0.9375	0.3616	0.9425	0.4888	0.9375	0.5802	0.9350	0.6243
	1000	0.9400	0.3190	0.9450	0.3946	0.9425	0.4765	0.9400	0.5374
Median-EL	200	0.9100	0.5624	0.9325	0.7289	0.9225	0.7818	0.9200	0.8259
	400	0.9200	0.4943	0.9350	0.6571	0.9300	0.7156	0.9225	0.7560
	600	0.9325	0.3817	0.9400	0.5225	0.9350	0.6337	0.9250	0.6821
	1000	0.9375	0.3286	0.9425	0.4342	0.9425	0.5459	0.9375	0.5953
LS-EL	200	0.9400	0.4638	0.9250	0.7958	0.8925	1.3242	0.8525	1.3067
	400	0.9450	0.3230	0.9275	0.7117	0.9100	1.1550	0.9075	1.1961
	600	0.9475	0.2679	0.9325	0.6321	0.9175	0.8365	0.9125	0.8920
	1000	0.9525	0.2138	0.9350	0.5236	0.9225	0.7069	0.9175	0.7781
Normal Approximation	200	0.9012	0.5039	0.9277	0.6580	0.9209	0.8603	0.9142	1.0361
	400	0.9170	0.4160	0.9286	0.5819	0.9236	0.7686	0.9189	0.9240
	600	0.9277	0.3582	0.9310	0.5130	0.9301	0.6915	0.9210	0.8520
	1000	0.9321	0.3130	0.9350	0.4679	0.9327	0.6150	0.9278	0.7930

DGP 2 To illustrate the variable selection procedure, we consider

$$Y_i = \sum_{j=1}^{8} X_{j,i}\beta_j + \varepsilon_i,\ i = 1, \cdots, N, \tag{5.2}$$

where $\boldsymbol{\beta} = (0.5, 1, 1.5, 2, 0, 0, 0, 0)$ and $\mathbf{X}_i$ is generated from a multivariate normal distribution of dimension 8 with mean 0 and pairwise covariance

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

$0.5^{|j-k|}$ for $j, k = 1, \cdots, 8$. The error distributions considered here are identical to those in DGP 1. To assess finite sample performance, we compute the generalized mean squared error (GMSE) of the parameter, which is defined as $\text{GMSE}(\hat{\boldsymbol{\beta}}) = \mathbb{E}[(\mathbf{X}^T\hat{\boldsymbol{\beta}} - \mathbf{X}^T\boldsymbol{\beta}_0)^2] = \mathbb{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T\Sigma_{\mathbf{X}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]$ and $\Sigma_{\mathbf{X}}$ is the covariance matrix of $\mathbf{X}$. As the covariates $\mathbf{X}$ are centralized, we have $\Sigma_{\mathbf{X}} = \mathbb{E}(\mathbf{X}\mathbf{X}^T)$ and $\text{GMSE}(\hat{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\mathbb{E}(\mathbf{X}\mathbf{X}^T)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. In addition, several indicators related to variable selection are also reported in the following table, where C stands for the average number of nonzero coefficients correctly estimated to be nonzero, IC indicates the average number of zero coefficients incorrectly estimated to be nonzero, U-Fit represents the proportion of simulations excluding any nonzero coefficient, C-Fit denotes the percentage of simulations selecting exact subset model, and O-Fit shows the proportion of simulations including all nonzero coefficients and some zero ones in 400 replications. Throughout the simulations, an estimate whose absolute value is less than 10^{-4} is set to be zero.

The simulation results are reported in Table 3, from which we can observe that the variable selection procedure based on the oracle method performs the best, while the naive method produces the worst results. The suggested kernel mode-based method can select all four true covariates in all scenarios and performs better than LS-based variable selection proce-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

dure for all nonnormal error distributions. It also outperforms both Huber and median-based variable selection procedures in all cases as they tend to select irrelevant variables more frequently, resulting in lower oracle proportions. When the error follows a standard normal distribution, the developed procedure is relatively effective compared with LS-based method in both model complexity and model error. Additionally, the overfitting values in Table 3 are always greater than zero for all selection procedures, suggesting that the SCAD penalty may have a tendency to overfit more than to underfit. This phenomenon might be alleviated by adjusting the tuning parameters more efficiently, which is deserved to be researched further in the future. In respect to GMSE, mode-based method has smaller magnitudes compared to other estimation methods when the error distribution is nonnormal. As the sample size N increases, the resulting GMSE grows smaller and C-Fit becomes larger, while the corresponding IC and O-Fit get smaller. Overall, the results demonstrate that the model selection outcome drew on the established method is satisfactory and that the selected model closely resembles the true model in terms of nonzero coefficients.

Table 3: Monte Carlo Results-DGP 2 (Variable Selection)

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

N	Method	GMSE	C	IC	U-Fit	C-Fit	O-Fit	N	GMSE	C	IC	U-Fit	C-Fit	O-Fit
						$\mathcal{N}(0,1)$								
	Oracle	0.2463	3.5961	0.1911	0.0000	0.7800	0.2200		0.1369	3.6579	0.1529	0.0000	0.8000	0.2000
	Proposed	0.2886	3.5534	0.2034	0.0000	0.7700	0.2300		0.1757	3.6435	0.1836	0.0000	0.7800	0.2200
200	Naive	0.6423	3.3883	0.2478	0.0100	0.7700	0.2200	400	0.4053	3.4121	0.2365	0.0200	0.7600	0.2200
	Huber	0.3372	3.4722	0.2335	0.0000	0.7300	0.2700		0.2023	3.5746	0.2022	0.0000	0.7500	0.2500
	Median	0.4123	3.4564	0.2831	0.0000	0.7100	0.2900		0.2484	3.5236	0.2554	0.0000	0.7400	0.2600
	LS	0.2119	3.6958	0.1764	0.0000	0.8100	0.1900		0.1105	3.7143	0.1415	0.0000	0.8200	0.1800
	Oracle	0.1035	3.7985	0.1182	0.0000	0.8200	0.1800		0.0772	3.8023	0.1075	0.0000	0.8500	0.1500
	Proposed	0.1246	3.7213	0.1337	0.0000	0.8000	0.2000		0.0815	3.7892	0.1142	0.0000	0.8200	0.1800
600	Naive	0.2961	3.5744	0.1793	0.0000	0.7700	0.2300	1000	0.2404	3.6275	0.1448	0.0000	0.7900	0.2100
	Huber	0.1567	3.6403	0.1542	0.0000	0.7700	0.2300		0.1183	3.7201	0.1258	0.0000	0.8000	0.2000
	Median	0.1766	3.5960	0.1837	0.0000	0.7600	0.2400		0.1357	3.6936	0.1354	0.0000	0.7800	0.2200
	LS	0.0923	3.8281	0.1132	0.0000	0.8500	0.1500		0.0543	3.9417	0.1011	0.0000	0.8700	0.1300
						$t(3)$								
	Oracle	0.3163	3.5291	0.2646	0.0000	0.7500	0.2500		0.1698	3.6118	0.2103	0.0000	0.7800	0.2200
	Proposed	0.3335	3.4760	0.2837	0.0000	0.7400	0.2600		0.1822	3.5845	0.2368	0.0000	0.7700	0.2300
200	Naive	0.7470	3.3126	0.3675	0.0600	0.7100	0.2300	400	0.4274	3.4076	0.3152	0.0600	0.7200	0.2200
	Huber	0.4194	3.4266	0.3157	0.0000	0.7100	0.2900		0.2275	3.5194	0.2865	0.0000	0.7400	0.2600
	Median	0.4673	3.4179	0.3225	0.0000	0.7000	0.3000		0.2613	3.5016	0.2941	0.0000	0.7300	0.2700
	LS	0.5821	3.3463	0.3437	0.0000	0.6800	0.3200		0.2896	3.4776	0.3035	0.0000	0.7200	0.2800
	Oracle	0.1339	3.7219	0.1437	0.0000	0.8000	0.2000		0.0925	3.7998	0.1162	0.0000	0.8300	0.1700
	Proposed	0.1512	3.6143	0.1563	0.0000	0.7900	0.2100		0.1078	3.7467	0.1282	0.0000	0.8100	0.1900
600	Naive	0.3226	3.5472	0.2077	0.0000	0.7500	0.2500	1000	0.2679	3.6132	0.1687	0.0000	0.7700	0.2300
	Huber	0.1761	3.5962	0.1762	0.0000	0.7700	0.2300		0.1329	3.7039	0.1333	0.0000	0.8000	0.2000
	Median	0.1955	3.5895	0.1923	0.0000	0.7600	0.2400		0.1525	3.6823	0.1407	0.0000	0.7900	0.2100
	LS	0.2147	3.5784	0.2018	0.0000	0.7600	0.2400		0.1849	3.6563	0.1539	0.0000	0.7800	0.2200
						$L_p(0,1)$								
	Oracle	0.2964	3.5752	0.2173	0.0000	0.7700	0.2300		0.1506	3.6344	0.1708	0.0000	0.7900	0.2100
	Proposed	0.3143	3.5327	0.2334	0.0000	0.7600	0.2400		0.1796	3.6154	0.2015	0.0000	0.7800	0.2200
200	Naive	0.7135	3.3472	0.3105	0.0600	0.7200	0.2200	400	0.4154	3.4100	0.2868	0.0200	0.7400	0.2400
	Huber	0.3978	3.4576	0.2658	0.0000	0.7400	0.2600		0.2129	3.5483	0.2446	0.0000	0.7500	0.2500
	Median	0.4454	3.4363	0.2745	0.0000	0.7300	0.2700		0.2502	3.5152	0.2619	0.0000	0.7400	0.2600
	LS	0.5194	3.3828	0.2931	0.0000	0.7400	0.2600		0.2723	3.5057	0.2751	0.0000	0.7500	0.2500
	Oracle	0.1279	3.7635	0.1229	0.0000	0.8100	0.1900		0.0853	3.8011	0.1123	0.0000	0.8400	0.1600
	Proposed	0.1316	3.6917	0.1426	0.0000	0.8000	0.2000		0.0924	3.7652	0.1205	0.0000	0.8200	0.1800
600	Naive	0.3164	3.5620	0.1983	0.0000	0.7600	0.2400	1000	0.2544	3.6181	0.1575	0.0000	0.7800	0.2200
	Huber	0.1662	3.6379	0.1639	0.0000	0.7800	0.2200		0.1239	3.7148	0.1294	0.0000	0.8000	0.2000
	Median	0.1864	3.5906	0.1887	0.0000	0.7700	0.2300		0.1462	3.6895	0.1387	0.0000	0.8000	0.2000
	LS	0.2026	3.5876	0.1905	0.0000	0.7600	0.2400		0.1639	3.6754	0.1471	0.0000	0.7900	0.2100
						$0.9\mathcal{N}(0,1)+0.1\mathcal{N}(0,10^2)$								
	Oracle	0.3786	3.5170	0.2823	0.0000	0.7300	0.2700		0.2208	3.6092	0.2313	0.0000	0.7700	0.2300
	Proposed	0.3964	3.4586	0.2904	0.0000	0.7100	0.2900		0.2461	3.5568	0.2580	0.0000	0.7600	0.2400
200	Naive	0.8113	3.2718	0.3825	0.0200	0.6800	0.3400	400	0.5429	3.3602	0.3341	0.0600	0.7100	0.2300
	Huber	0.4524	3.4153	0.3280	0.0000	0.7100	0.2900		0.2611	3.4825	0.3087	0.0000	0.7300	0.2700
	Median	0.4993	3.4069	0.3336	0.0000	0.7000	0.3000		0.2862	3.4603	0.3175	0.0000	0.7200	0.2800
	LS	0.6125	3.3256	0.3552	0.0000	0.7000	0.3000		0.3122	3.4162	0.3224	0.0000	0.7200	0.2800
	Oracle	0.1854	3.7055	0.1615	0.0000	0.7900	0.2100		0.1273	3.7524	0.1323	0.0000	0.8200	0.1800
	Proposed	0.1966	3.6791	0.1783	0.0000	0.7800	0.2200		0.1421	3.7117	0.1483	0.0000	0.8000	0.2000
600	Naive	0.3630	3.4428	0.2288	0.0000	0.7400	0.2600	1000	0.2985	3.5520	0.1931	0.0000	0.7500	0.2500
	Huber	0.2104	3.5305	0.1829	0.0000	0.7700	0.2300		0.1773	3.6445	0.1561	0.0000	0.7800	0.2200
	Median	0.2338	3.5109	0.1947	0.0000	0.7600	0.2400		0.1964	3.6052	0.1628	0.0000	0.7700	0.2300

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

5.2 Empirical Analysis: Housing Market

We evaluate the proposed method by using the clean air housing market dataset from census tracts of Boston from 1970 census in this subsection. The dataset, available at `http://lib.stat.cmu.edu/datasets/boston_corrected.txt`, is mainly used for investigating the effect of clean air on house price, containing 16 variables among 506 observations. For simplicity, we exclude all categorical variables from the dataset to conduct estimation using the suggested mode-based method and identify the model structure using the developed mode-based variable selection procedure. After that, we have the dependent variable CMEDV (corrected median value of owner-occupied homes) and the covariates LON (point longitudes in decimal degrees), LAT (point latitudes in decimal degrees), CRIM (crime rate by town), ZN (proportion of residential land zoned for large lots by town), INDUS (proportion non-retail business acres per town), NOX (nitrogen oxide concentration), RM (average number of rooms per dwelling), AGE (proportion of owner occupied homes built prior to 1940), DIS (weighted distances to five employment centers in Boston), TAX (property tax rate), PTRATIO (pupil-teacher ratio by town), B (black population proportion), and LSTAT (proportion of lower socioeconomic status population). Each variable is standardized prior to analysis. We assume a parametric regression

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

model between the response CMEDV and the 13 covariates, along with an intercept. Since the distribution of the dependent variable is not normal (see Figure S7 in the supplementary file), it is beneficial to conduct the suggested mode-based estimation to achieve robustness and efficiency. Additionally, Figure S8 in the supplementary file indicates high correlation among some variables, suggesting redundant information. Thereupon, it is necessary to conduct variable selection to determine the independent effect of each important variable. Due to the absence of truncation in the data, to illustrate the proposed method, we suppose that the dataset is truncated by an exponential distribution with mean 13, resulting in approximately 90% truncation rate. For comparison, we report results from the suggested mode-based estimation, naive method, and Huber, median, and mean (LS) estimations.

The estimation results are presented in Table 4, where the numbers above the brackets represent estimates and those in the brackets indicate empirical likelihood confidence intervals. Analyzing the table reveals that ignoring the truncation issue leads to kernel mode-based estimates differing in signs and magnitudes compared to the suggested and other existing estimates. In comparison to mean estimation, the proposed mode-based estimation can capture some distinguish features of the data. For instance, the resulting mode-based estimate for AGE is negative, while mean estimation

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

yields a positive estimate. Practically, one would anticipate AGE to negatively impact housing prices. Therefore, mode-based estimation provides more reasonable estimates that align with real-world intuition. Moreover, the empirical likelihood confidence intervals exhibit asymmetry around the estimates. Mode-based estimation provides the shortest confidence intervals while LS estimation offers the largest ones, which is expectable given that the dependent variable does not adhere to a normal distribution. Additionally, mode-based procedure can complement existing selection methods by identifying unique covariates that reveal the "most likely" (mode) effect. For example, the covariate CRIM is not selected by Huber, median, or mean estimation, while it is chosen by mode-based estimation, consistent with existing literature suggesting that crime rate influences housing values.

Table 4: Empirical Analysis Results

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

	Estimation					Variable Selection				
Covariates	Proposed	Naive	Huber	Median	LS	Proposed	Naive	Huber	Median	LS
LON	-7.9569 [-7.9746,-7.7685]	-6.7217 [-6.7558, -6.5286]	-7.8459 [-7.8082, -7.5956]	-7.0354 [-7.0869, -6.7669]	-7.1680 [-7.2513, -6.8378]	✓	✓	✓	✓	✓
LAT	1.7016 [1.6549, 1.7341]	3.6225 [3.5726, 3.6559]	2.5055 [2.4660, 2.5480]	1.9315 [1.8733, 1.9644]	1.9641 [1.9153, 2.0191]	✓	✓	✓	✓	×
CRIM	-0.1362 [-0.1962, -0.1320]	0.1458 [0.1155, 0.1937]	-0.0959 [-0.1334, -0.0651]	-0.0930 [-0.1208, -0.0480]	-0.0657 [-0.1089, -0.0278]	✓	✓	×	×	×
ZN	0.0195 [0.0131, 0.0408]	0.0111 [0.0046, 0.0371]	0.0184 [0.0153, 0.0446]	0.0278 [0.0170, 0.0516]	0.0379 [0.0207, 0.0589]	×	×	×	✓	✓
INDUS	-0.0175 [-0.0197, -0.0117]	-0.0501 [-0.0524, -0.0472]	-0.0630 [-0.0698, -0.0568]	-0.0715 [-0.0770, -0.0592]	-0.0574 [-0.0639, -0.0408]	×	×	×	×	×
NOX	-2.1120 [-2.1739, -2.1107]	2.4160 [2.3666, 2.4401]	-1.5836 [-1.6182, -1.5490]	-9.1068 [-9.1428, -9.0703]	-11.9698 [-12.0128, -11.9191]	✓	✓	✓	✓	✓
RM	5.1308 [5.0871, 5.1696]	6.9705 [6.9130, 7.0064]	6.3533 [6.3154, 6.4046]	4.9160 [4.8780, 4.9753]	4.0584 [4.0112, 4.1140]	✓	✓	✓	✓	✓
AGE	-0.0387 [-0.0425, -0.0272]	-0.0606 [-0.0727, -0.0493]	-0.0410 [-0.0522, -0.0330]	0.0027 [0.0008, 0.0223]	0.0031 [0.0009, 0.0296]	×	×	×	×	×
DIS	-0.6032 [-0.6447, -0.5809]	-0.6469 [-0.6815, -0.6094]	-0.8137 [-0.8445, -0.7710]	-0.8843 [-0.8449, -0.7648]	-1.3886 [-1.4354, -1.3427]	✓	✓	✓	✓	✓
TAX	-0.0038 [-0.0047, -0.0026]	-0.0043 [-0.0054, -0.0017]	-0.0063 [-0.0075, -0.0046]	-0.0002 [-0.0036, -0.0001]	-0.0002 [-0.0039, -0.0001]	×	×	×	×	×
PTRATIO	-0.5096 [-0.5562, -0.4728]	-0.4354 [-0.4817, -0.3890]	-0.5550 [-0.5927, -0.5031]	-0.6906 [-0.7321, -0.6398]	-0.7573 [-0.8305, -0.7279]	✓	✓	✓	✓	✓
B	0.0132 [0.0082, 0.0133]	0.0227 [0.0186, 0.0270]	0.0120 [0.0089, 0.0155]	0.0119 [0.0083, 0.0155]	0.0088 [0.0046, 0.0124]	×	×	×	×	✓
LSTAT	-0.2232 [-0.2320, -0.1495]	-0.0089 [-0.0193, -0.0022]	-0.2130 [-0.2687, -0.1750]	-0.3922 [-0.4320, -0.3264]	-0.5390 [-0.5957, -0.4829]	✓	✓	✓	✓	✓

6. Concluding Remarks

In this paper, we investigate parametric mode-based regression within a randomly truncated framework. We propose an estimator built on the kernel objective function, demonstrating its consistency and asymptotic normality. We also develop a mode-based empirical likelihood estimation procedure to construct confidence intervals. The established nonparametric version of Wilks' theorem ensures that the constructed empirical likelihood confidence interval has asymptotically correct coverage probability. We suggest a pe-

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

nalized kernel mode-based objective function for simultaneous estimation and variable selection with randomly truncated data, where the oracle property is demonstrated. The numerical results underscore the effectiveness of the proposed estimation and variable selection procedures.

The research presented in this paper can be extended in several directions. For example, due to dataset restrictions, we analyze the real data with an artificial truncation. However, in practice, it is more common to encounter datasets that are both left-truncated and right-censored; see Zhou and Yip (1999) and Su and Wang (2012). It would be intriguing to investigate the integration of both left-truncated and right-censored data within the kernel mode-based estimation framework. We discuss the detailed estimation techniques for handling right-censoring alongside left-truncation in the supplementary file. In addition, addressing doubly truncated data, where both the left and right endpoints of the study window are truncated, presents another promising avenue for future research. These extensions offer exciting opportunities for further exploration and innovation in the field of survival analysis and mode estimation.

Supplementary Material

The supplementary file contains additional numerical and technical results.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Acknowledgements

We are deeply grateful to the Co-Editor Yi-Hau Chen, Associate Editor, and two anonymous referees for their constructive comments, leading to the substantial improvement of the paper. We would also like to thank Bo Honoré, Aman Ullah, and the seminar participants at the UC Riverside, University of Washington, and University of Iowa for their helpful comments. Tao Wang's research is supported by SSHRC-IDG grant (430-2023-00149) and UVic-SSHRC Explore grant (2023-2024), and Weixin Yao's research is supported by NSF grant (DMS-2210272).

References

Amemiya, T. (1973). Regression Analysis When the Dependent Variable is Truncated Normal. *Econometrica*, 41, 997-1016.

Chen, S. X. and Van Keilegom, I. (2009). A Review on Empirical Likelihood Methods for Regression. *TEST*, 18, 415-447.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96, 1348-1360.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Hausman, J. A. and Wise, D. A. (1977). Social Experimentation, Truncated Distributions, and Efficient Estimation. *Econometrica*, 45, 919-938.

He, S. and Yang, G. L. (1998). Estimation of the Truncation Probability in the Random Truncation Model. *The Annals of Statistics*, 26, 1011-1027.

He, S. and Yang, G. L. (2003). Estimation of Regression Parameters with Left Truncated Data. *Journal of Statistical Planning and Inference*, 117, 99-122.

Kemp, G. C. R. and Santos Silva, J. M. C. (2012). Regression towards the Mode. *Journal of Econometrics*, 170, 92-101.

Lee, M. J. (1989). Mode Regression. *Journal of Econometrics*, 42, 337-349.

Lee, M. J. (1993). Quadratic Model Regression. *Journal of Econometrics*, 57, 1-19.

Owen, A. B. (1988). Empirical Likelihood Ratio Confidence Intervals for A Single Functional. *Biometrika*, 75, 237-249.

Owen, A. B. (1990). Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics*, 18, 90-120.

Stute, W. (1993). Almost Sure Representations of the Product-Limit Estimator for Truncated Data. *The Annals of Statistics*, 21, 146-156.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Su, Y.-R. and Wang, J.-L. (2012). Modeling Left-Truncated and Right-Censored Survival Data with Longitudinal Covariates. *The Annals of Statistics*, 40, 1465-1488.

Ullah, A., Wang, T., and Yao, W. (2021). Modal Regression for Fixed Effects Panel Data. *Empirical Economics*, 60, 261-308.

Ullah, A., Wang, T., and Yao, W. (2022). Nonlinear Modal Regression for Dependent Data with Application for Predicting COVID-19. *Journal of the Royal Statistical Society Series A*, 185, 1424-1453.

Ullah, A., Wang, T., and Yao, W. (2023). Semiparametric Partially Linear Varying Coefficient Modal Regression. *Journal of Econometrics*, 1001-1026.

Wang, M. C. (1989). A Semiparametric Model for Randomly Truncated Data. *Journal of the American Statistical Association*, 84, 742-748.

Wang, K. and Li, S. (2021). Robust Distributed Modal Regression for Massive Data. *Computational Statistics & Data Analysis*, 160, 107225.

Wang, T. (2024). Nonlinear Kernel Mode-Based Regression for Dependent Data. *Journal of Time Series Analysis*, 45, 189-213.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0288

Woodroofe, M. (1985). Estimating a Distribution Function with Truncated Data. *The Annals of Statistics*, 13, 163-177.

Yao, W. and Li, L. (2014). A New Regression Model: Modal Linear Regression. *Scandinavian Journal of Statistics*, 41, 656-671.

Zhou, W. (2011). A Weighted Quantile Regression for Randomly Truncated Data. *Computational Statistics and Data Analysis*, 55, 554-566.

Zhou, Y. and Yip, P. S. (1999). A Strong Representation of the Product-Limit Estimator for Left Truncated and Right Censored Data. *Journal of Multivariate Analysis*, 69, 261-280.

a. Department of Economics and Department of Mathematics and Statistics (by courtesy), University of Victoria, Victoria, BC V8W 2Y2, Canada. E-mail: taow@uvic.ca

b. Department of Statistics, University of California, Riverside, CA 92521, USA. E-mail: weixin.yao@ucr.edu