Journal of Multivariate Analysis 125 (2014) 36–49

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Model checking in Tobit regression via nonparametric smoothing☆

Hira L. Koul, Weixing Song *, Shan Liu

Michigan State University, United States
Kansas State University, United States

ARTICLE INFO

Article history:
Received 2 February 2013
Available online 18 December 2013

AMS subject classifications:
primary 62G08
secondary 62G10

Keywords:
Tobit regression model
Zheng's test
Consistency and local power

ABSTRACT

This paper proposes a class of lack-of-fit tests for checking the adequacy of a presumed parametric form of the regression function in Tobit regression models. This class of tests is a weighted adaptation of the Zheng's test for fitting a parametric regression model. The asymptotic null distributions of the underlying test statistics are shown to be normal. Moreover, the consistency of these tests against some fixed alternatives and asymptotic power against some local nonparametric alternatives are also derived. An optimal test within the proposed class of tests against a given sequence of nonparametric local alternatives is identified. A finite sample simulation shows some superiority of some of the proposed tests, compared to some of the existing tests.

1. Introduction

The Tobit regression model was originally designed for investigating the relationship between the household expenditures on durable goods and the income. It is now a frequently used tool for modeling censored or truncated variables in many areas such as econometrics, biometrics, agricultural study and engineering fields. Early development of the Tobit regression model can be traced back to Tobin [25]. The survey paper of Amemiya [3] provided a comprehensive introduction to the Tobit model, together with a detailed discussion on the commonly used estimation procedures. Many empirical examples can be found in Amemiya [3], Blaylock and Blisard [5], McConnel and Zetzman [20], Lichtenberg and Shapiro [18], Adesina and Zinnah [1], Ekstrand and Carpenter [9] and the references therein.

To be specific, suppose the relationship between a random scalar Y^* and a random vector X can be fitted by a classical regression model, $Y^* = m(X) + \varepsilon$, where $m(x) = E[Y^*|X = x]$. In some cases, Y^* can only be observed if Y^* falls into a certain range. The Tobit regression model assumes Y^* can be observed if its value is above a certain threshold y_0 which is usually assumed to be known, that is, one actually observes $Y = Y^*I(Y^* > y_0) + y_0I(Y^* \le y_0)$. Without the loss of generality, we shall take $y_0 = 0$ throughout this paper.

The standard Tobit regression model assumes that $m(x) = \alpha + \beta' x$, and the error ε has a normal distribution with mean 0 and unknown variance σ_ε^2. The existing literature on the standard Tobit regression model mainly focuses on the estimation of the unknown regression parameters $(\alpha, \beta')'$ and the error variance σ_ε^2. Under the normality assumption of the error term ε, Amemiya [2] and Heckman [12,13] proposed consistent estimators for $\theta = (\alpha, \beta', \sigma_\varepsilon^2)'$. These estimators are not consistent if the normality assumption does not hold. A robust estimator of θ was proposed by Powell [21] based on the least absolute deviations. Lewbel and Linton [17] and Zhou [33] developed nonparametric regression estimators for a class

☆ Research supported in part by the NSF DMS Collaborative Grants 1205271 and 1205276.

* Corresponding author at: Michigan State University, United States.
E-mail address: weixing@ksu.edu (W. Song).

0047-259X/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.11.017

of general nonparametric Tobit regression models, but the hard-to-interpret nature of nonparametric procedures still makes the parametric modeling the first choice for practitioners.

Generally speaking, the choice of the parametric forms of $m(x)$ is either based on some empirical evidence or simply for the sake of mathematical convenience. Misspecification of the function form of $m(x)$ often leads to misleading conclusions. For example, Horowitz and Neumann [14] showed that the violation of the linearity assumption can produce inconsistent estimators of the parameters and biased prediction of the survival time in censored regression models. From both theoretical and practical point of views, it is necessary to develop formal tests to check the adequacy of the selected regression model.

The literature on the lack-of-fit testing in Tobit regression models is relatively scarce, compared to that in the classical regression setup. Wang [26] proposed a simple nonparametric test in Tobit median regression model in which the median of the random error is assumed to be zero, while the null hypothesis is restricted to linear regression functions only. Song [23] developed a lack-of-fit testing procedure for a more general null hypothesis, not limited to linear regression functions, by assuming that $E(\varepsilon) = 0$. The proposed test is based on the Khamaladze type transformation of a certain marked residual process. The most restrictive assumption in Song [23] is that the predictor variable X must be one-dimensional. Following a few of the significant works such as Härdle and Mammen [11] and Koul and Ni [16] in the classical regression models, Song and Zhang [24] developed a class of lack-of-fit tests based on a class of empirical L_2-distances between a nonparametric estimator and a parametric estimator of the regression function being fitted under the null hypothesis. The function form being tested may not be limited to linear, and the predictor can be multidimensional. In this paper, we apply the idea from Zheng [32] to construct a class of new lack-of-fit tests for the Tobit regression model. Due to its computational simplicity and good performance, the idea from Zheng [32] has been also implemented in several other contexts, see, e.g., Dette and von Lieres und Wilkau [7], and Wang and Zhou [27], among others. Lopez and Patilea [19] developed a nonparametric lack-of-fit test using Zheng's idea for a parametric mean regression model when the response variable is censored at random, which is quite different from our current setup.

Although Wang [26] conjectured that her test can detect local alternatives that converge to the null at a certain rate, similarly as in testing the validity of the mean regression functions, the explicit asymptotic power function was not given. In the current paper, we derive an explicit expression for the asymptotic power under certain local alternatives and describe an optimal test within the class of proposed tests for a given local alternative.

Other testing methods could also be adapted to deal with the model adequacy checking in the Tobit regression models. For example, the consistent test proposed by Dette [6] and the modified version of the Härdle and Mammen [11] (HM) test constructed in Zhang and Dette [31]. In Section 4, a finite sample comparison of a member of the proposed class of tests, denoted by KSL, is made with that of Dette (D) and HM tests adapted to the current setup. In Table 4 of Section 4, it is observed that the empirical level of the D-test is much smaller than that of the KSL and HM tests, while it is comparable for these two tests. The empirical power of HM is somewhat better than that of the KSL test, while both of these tests dominate the D-test. For more details see Section 4. Although the KSL-test appears to be somewhat inferior to the HM-test in this limited simulation, it is still a competitive candidate for model checking due to its relatively simple form and computational ease, in particular when the number of predictors d is large.

The paper is organized as follows. Section 2 contains the description of the proposed test statistics, and a list of technical assumptions needed for the asymptotic results. The main results are presented in Section 3, including the asymptotic null distributions, the consistency and the local power of the tests; simulation studies are presented in Section 4, and the proofs of the main results are deferred to Section 5.

Throughout this paper, we will use f_v, F_v to denote respectively, density and distribution function (d.f.) of a random variable v, and $\to_D$ denotes the convergence in distribution.

2. Test statistics

Consider the classical regression model $Y = \mu(X) + \varepsilon$, where X is a d-dimensional random predictor, ε is the error term such that $E(\varepsilon|X) = 0$, so that $\mu(x) = E(Y|X = x)$. Let $\Theta \subset \mathbb{R}^q$ and $m(x, \theta), x \in \mathbb{R}^d, \theta \in \Theta$ be a family of parametric functions, and consider the problem of testing

$$\mathcal{H}_0 : \mu(x) = m(x, \theta), \quad \forall x \in \mathbb{R}^d, \text{ and for some } \theta \in \Theta, \quad \text{vs.} \quad \mathcal{H}_1 : \mathcal{H}_0, \text{ is not true.}$$

Zheng [32] used the following idea to propose a test of $\mathcal{H}_0$. Let $e := Y - m(X, \theta)$ and f denote the density of X. Then, $E(e|X) = \mu(X) - m(X, \theta)$. Hence, under $\mathcal{H}_0$, $E[eE(e|X)f(X)] = 0$, while under $\mathcal{H}_1$, $E[eE(e|X)f(X)] = E\{[\mu(X) - m(X, \theta)]^2 f(X)\} > 0$. So an empirical version of $E[eE(e|X)f_X(X)]$ can be used as a building block for a test statistic. Based on the sample $\{(X_i, Y_i) : i = 1, 2, \ldots, n\}$ from the above model, one obtains an estimate $\hat{\theta}$ of θ under $\mathcal{H}_0$ using existing procedures in the literature. After replacing e_i by $\hat{e}_i = Y_i - m(X_i, \hat{\theta})$, $E(e_i|X_i)$ and $f_X(X_i)$ by their Nadaraya–Watson kernel estimators, Zheng's test statistic is a standardized version of the statistic

$$\frac{1}{n(n-1)} \sum_{i \neq j}^{n} K_h(X_i - X_j)\hat{e}_i\hat{e}_j,$$

where K is the kernel function, $K_h(\cdot) := h^{-d}K(\cdot/h)$, and h is the bandwidth. Zheng established consistency of this test. The asymptotic null distribution of the standardized version of the above statistics along with its power against local alternatives can be derived by the aid of the central limit theorem for degenerate U-statistics of Hall [10].

Now consider the Tobit regression model

$$Y^* = m(X) + \varepsilon, \qquad Y = Y^* I(Y^* > 0),$$

and the problem of testing

$$H_0 : m(x) = m(x, \theta_0), \quad \forall x \in \mathbb{R}^d, \text{ and for some } \theta_0 \in \Theta, \quad \text{vs.} \quad H_1 : H_0 \text{ is not true.}$$

Because Y^* can be observed only when $Y^* > 0$, it is not feasible to directly apply Zheng's test in the Tobit regression model for testing H_0. We need to build a new regression model which can show a certain dependence between an observable quantity and the regression function $m(x)$. A natural way to find this dependence is to consider the conditional expectation $g(x) := E(Y|X = x)$. Throughout this paper, we shall assume that the density function f_ε is known for the sake of simplicity and model identifiability, but f_ε need not be a Gaussian density. It might appear restrictive to assume that the distribution of f_ε is known, but in real applications, in particular for most econometric literature, this is indeed a common practice.

Now, let $Q_j(z) = \int_z^\infty u^j f_\varepsilon(u)du,\ j = 0, 1,\ z \in \mathbb{R}$. Then

$$g(x) = E(Y|X = x) = m(x)Q_0(-m(x)) + Q_1(-m(x)). \tag{2.1}$$

This motivates us to consider the following regression model

$$Y = g(X) + \xi = m(X)Q_0(-m(X)) + Q_1(-m(X)) + \xi. \tag{2.2}$$

Clearly, ξ and $g(X)$ are uncorrelated.

An attractive feature of the model (2.2) is that, as a functional of $m(x)$, g is strictly increasing, provided $F_\varepsilon(-a) < 1$, for all $a \in \mathbb{R}$. This follows because by Lemma 5.1, $\partial(aQ_0(-a) + Q_1(-a))/\partial a = 1 - F_\varepsilon(-a) > 0$. Therefore, testing for H_0 is equivalent to testing for

$$\mathcal{K}_0 : g(x) = g(x, \theta) \quad \text{for some } \theta \in \Theta, \quad \text{versus} \quad \mathcal{K}_1 : \mathcal{K}_0 \text{ is not true}$$

for regression model (2.2), where $g(x, \theta)$ is the same as $g(x)$ with $m(x)$ replaced by $m(x, \theta)$.

Another way to proceed is to base a test on $I(Y = 0)$. Note that $E(I(Y = 0)|X = x) = 1 - Q_0(-m(x)) = F_\varepsilon(-m(x))$, which as a functional of $m(x)$, is also strictly monotone provided now that F_ε is strictly increasing. As a result, one might consider the binary regression model

$$I(Y = 0) = 1 - Q_0(-m(X)) + \eta, \tag{2.3}$$

where η denotes the error. Because this model only uses the truncation information of Y^*, not the full observed data, the corresponding test may not be as powerful as the one based on the model (2.2), which is also confirmed by the simulation studies of Section 4. Throughout this paper, we will only focus on the model (2.2), except for a brief discussion on the finite sample performance of a test based on the model (2.3) given in Section 4.

Let $\hat{\theta}_n$ be any $\sqrt{n}$-consistent estimator of θ_0 under H_0, $\hat{\xi}_i = Y_i - g(X_i, \hat{\theta}_n)$, K be a symmetric density function and let $h = h_n$ be a sequence of window widths, and $w(x)$ be an positive measurable function. Then, following Zheng, a class of test statistics, one for each w, for testing $\mathcal{K}_0$ vs. $\mathcal{K}_1$, hence for testing the hypotheses H_0 versus H_1, is

$$V_n = \frac{1}{n(n-1)h^d} \sum_{i \neq j} K\left(\frac{X_i - X_j}{h}\right) \hat{\xi}_i \hat{\xi}_j w(X_i) w(X_j). \tag{2.4}$$

The presence of the weight function w in V_n allows for choosing an optimal test against a sequence of local alternatives among this class of tests.

We shall show that the asymptotic null distribution of $nh^{d/2}V_n$ is normal with mean 0 and variance

$$\sigma^2 = 2 \int K^2(u)du \int \tau^4(x) f_X^2(x) w^4(x)dx, \tag{2.5}$$

where $\tau^2(x) = E[(Y - g(X, \theta))^2|X = x]$. A consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{2}{n(n-1)} \sum_{i \neq j}^{n} \frac{1}{h^d} K^2\left(\frac{X_i - X_j}{h}\right) \hat{\xi}_i^2 \hat{\xi}_j^2 w^2(X_i) w^2(X_j). \tag{2.6}$$

The proposed test statistic for testing K_0 is $T_n = nh^{d/2}V_n/\hat{\sigma}$, with the large values of T_n being significant.

We shall now describe the assumptions needed to derive the asymptotic results of these test statistics.

(C1) The error density f_ε is bounded, $E(\varepsilon) = 0$, and $E(\varepsilon^4) < \infty$; ε and X are independent.

(C2) $\tau^2(x) = E[(Y - g(X))^2|X = x]$, $\nu^4(x) = E[(Y - g(X))^4|X = x]$ are continuously differentiable with respect to x, and the derivatives are bounded by a measurable function $b(x)$ such that $Eb^2(X) < \infty$.

(C3) The density function f_X of X and its first-order derivatives are uniformly bounded.

(C4) $m(x,\theta)$ is continuously differentiable with respect to θ, the vector of derivatives $\dot{m}(x,\theta)$ satisfies $E\|\dot{m}(X;\theta_0)\|^4 < \infty$ and the following: for any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_0,

$$\sup_{1\le i\le n} |m(X_i,\hat{\theta}_n) - m(X_i,\theta_0) - (\hat{\theta}_n-\theta_0)'\dot{m}(X_i,\theta_0)| = O_p(1/n).$$

(C5) The kernel function K is continuous, bounded, symmetric around 0, and $\int x^2K(x)dx < \infty$.

(C6) The bandwidth $h \to 0$, $nh^d \to \infty$, as $n \to \infty$.

(C7) The weight function w is continuous and $E\{[\tau^8(X) + \|\dot{m}(X,\theta_0)\|^4]w^8(X)\} < \infty$.

Conditions (C2) and (C3) are the same as the Assumption 1 in Zheng [32], and are very typical in nonparametric smoothing literature. Condition (C4) plays a similar role as the Assumption 2 in Zheng [32] to guarantee the negligibility of the higher order term in some Taylor expansions used when deriving the asymptotic results of these test statistics. The kernel function in Condition (C5) and the bandwidth h in Condition (C6) are the most commonly used ones in nonparametric literature.

3. Main results

We shall assume that there always exists a $\sqrt{n}$-consistent estimator for the parameter θ in the regression function under H_0. The theorem below states the asymptotic null distribution of the test statistics T_n. In the sequel, all limits are taken as $n \to \infty$, unless mentioned otherwise.

Theorem 3.1. *Suppose* (C1)–(C7) *hold. Then under* H_0, $T_n = nh^{d/2}V_n/\hat{\sigma} \to_D N(0,1)$, *where* V_n *is defined in* (2.4) *and* $\hat{\sigma}$ *in* (2.6).

Hence, the asymptotic size of the test that rejects H_0, whenever $|T_n| > z_{1-\alpha/2}$, is α, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)100\%$ percentile of the standard normal distribution.

Next, consider the consistency of this test. Let $\gamma \notin \{m(x,\theta), x \in \mathbb{R}^d, \theta \in \Theta\}$ be a measurable real valued function on $\mathbb{R}^d$ with $E\gamma^2(X) < \infty$. Consider the alternate hypothesis

$$H_a : m(x) = \gamma(x), \quad \forall x.$$

To prove the consistency of the proposed tests against H_a, we have to consider the asymptotic behavior of $\hat{\theta}_n$ under the alternative H_a. In the classical regression setup, Jennrich [15], Wu [30], and White [28,29] showed that, under some mild regularity conditions, the nonlinear least squares estimator converges in probability and is asymptotically normal even in the presence of model misspecification. Similarly, for the regression model (2.2), if we define $\theta_a = \text{argmin}_\theta E_\gamma[Y - g(X,\theta)]^2$, where E_γ is the expectation under H_a, then under H_a and some regularity conditions on the $m(x,\theta)$, we can show that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$. We will not pursue a rigorous verification of this claim here. The relevant discussion can be found in Zheng [32]. Let $g_\gamma(x)$ denote the function $g(x)$ where $m(x)$ is replaced by $\gamma(x)$.

Theorem 3.2. *Suppose all the conditions in Theorem 3.1 hold with* θ_0 *replaced by* θ_a, *and* $E\{[g_\gamma(X) - g(X,\theta_a)]^2 f_X(X)\} > 0$. *Then, for any* $0 < \alpha < 1$, *the test that rejects* H_0 *whenever* $|T_n| > z_{1-\alpha/2}$ *is consistent against the alternatives* H_a.

Next, we describe the asymptotic power of the proposed tests against sequences of local nonparametric alternatives. For this purpose, let $\delta(x)$ be a continuous function such that $E\delta^2(X)w^2(X) < \infty$, and consider the following sequence of local alternatives

$$H_{\ell oc} : m(x) = m(x,\theta_0) + \delta(x)/\sqrt{nh^{d/2}}. \quad (3.1)$$

We continue to assume that the estimators $\hat{\theta}_n$ used in the test statistic are satisfying $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ without a rigorous justification. We have

Theorem 3.3. *Suppose all the conditions in Theorem 3.1 hold. Then under* $H_{\ell oc}$ *in* (3.1), $T_n \to_D N(\mu, 1)$, *where* $\mu = EQ_0^2(-m(X,\theta_0))\delta^2(X)w^2(X)f_X(X)/\sigma$ *and* σ *is defined in* (2.5).

Optimal w: From Theorem 3.3 we conclude that the asymptotic power of the proposed test is $1-\Phi(z_{\alpha/2}-\mu)+\Phi(-z_{\alpha/2}-\mu)$ which is an increasing function of μ. Thus, the weight function w that will maximize the power is the one that maximizes μ. But by the Cauchy–Schwarz inequality,

$$\mu = \frac{\int Q_0^2(-m(x,\theta_0))\delta^2(x)w^2(x)f_X^2(x)dx}{\sqrt{2\int K^2(u)du \int(\tau^2(x))^2 f_X^2(x)w^4(x)dx}} \le \sqrt{\frac{\int Q_0^4(-m(x,\theta_0))\delta^4(x)f_X^2(x)/(\tau^2(x))^2dx}{2\int K^2(u)du}},$$

with equality holding, if and only if, $w(x) \propto Q_0(-m(x,\theta_0))\delta(x)/\tau^2(x)$, for all x. Because μ is unique for all w's which are different up to a multiple, we may take the optimal w to be $w(x) = Q_0(-m(x,\theta_0))\delta(x)/\tau^2(x)$. Clearly, this weight function w is unknown because of θ_0, but one can estimate it by replacing θ_0 with any consistent estimator.

If the regression model (2.3) is used for model checking, and the first order derivative of the density function f_ε is bounded, then under the local alternative hypothesis (3.1), one can show that $T_n \to_D N(\mu, 1)$ with $\mu = Ef_\varepsilon^2(-m(X,\theta_0))\delta^2(X)w^2(X)f_X(X)/\sigma$ and σ is defined in (2.5). It is easily seen that here the optimal weight function $w(x) = f_\varepsilon(-m(x,\theta_0))\delta(x)/\tau^2(x)$.

As we mentioned in Section 1, some other existing tests of the lack-of-fit of a regression function in ordinary regression models can also be modified to check the function form in the Tobit regression model. Let

$$v_{ij} = \frac{K((X_i - X_j)/h)}{\sum_{k=1}^n K((X_i - X_k)/h)}, \qquad c_{ij} = \sum_{k=1}^n v_{ki}v_{kj}, \quad i, j = 1, 2, \ldots, n.$$

The empirical version of Härdle and Mammen [11]'s test (HM-test), proposed by Zhang and Dette [31], is to reject the null hypothesis whenever $nh^{d/2}\hat{\Gamma}_{n,HM}^{-1/2}|T_{n,HM} - \hat{C}_{n,HM}| > z_{\alpha/2}$, where

$$T_{n,HM} = \frac{1}{n}\sum_{i=1}^n\left(\sum_{j=1}^n v_{ij}\hat{\xi}_j\right)^2, \qquad \hat{C}_{n,HM} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n v_{ji}^2\hat{\xi}_i^2, \qquad \hat{\Gamma}_{n,HM} = 2h^d\sum_{i\neq j} c_{ij}^2\hat{\xi}_i^2\hat{\xi}_j^2.$$

The large sample results of this test statistic under the null and alternative hypotheses can be obtained in a similar way as in Zhang and Dette [31].

Dette [6]'s test (D-test) is based on the difference between the least squares variance estimate from the fitted regression model and a nonparametric kernel variance estimate. The test statistic has the form $nh^{d/2}\hat{\Gamma}_{n,D}^{-1/2}|T_{n,D} - \hat{C}_{n,D}|$, where

$$T_{n,D} = \frac{1}{n}\sum_{i=1}^n\hat{\xi}_i^2 - \frac{1}{n}\sum_{i=1}^n\left(Y_i - \sum_{j=1}^n v_{ij}Y_j\right)^2, \qquad \hat{\Gamma}_{n,D} = 2ch^d\sum_{i\neq j} c_{ij}^2\hat{\xi}_i^2\hat{\xi}_j^2,$$
$$\hat{C}_{n,D} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n v_{ji}^2(\hat{\xi}_i - \hat{\xi}_j)^2 + \frac{1}{n}\sum_{i=1}^n\sum_{j\neq k\neq i} v_{ij}v_{ik}\hat{\xi}_i^2,$$

and

$$c = \frac{\int\left(2K(u) - \int K(u+v)K(v)dv\right)^2 du}{\int K^2(u)du}.$$

The asymptotic normality under the null hypothesis and the consistency under certain fixed alternative hypotheses can be shown using a similar argument as in Dette [6] for the random design.

In the classical regression case where (Y, X) are fully observed, Zhang and Dette [31] found out that by choosing the weighting functions properly, and adopting equal bandwidths, HM-test is more powerful than Zheng [32]'s test, and is more powerful than the D-test. In the next section, we arrive at a similar conclusion in a simulation study in the current setup when we compare the proposed test corresponding to $w \equiv 1$ with the D and HM tests.

4. Simulation studies

To assess the finite sample performance of the proposed tests, Monte Carlo simulations are conducted in this section. Linear regression functions with $d = 1$ and $d = 2$ are chosen to serve as the null models, a variety of quadratic components are added to the linear terms to serve as the alternative models. The significance level is chosen to be 0.05, and the sample sizes considered are $n = 100, 300, 500$ for all simulations. For each setup, the test is repeated 1000 times, the empirical level and power are calculated by $\#\{|T_n| \geq 1.96\}/1000$. The simulation setups are exactly the same as in Song and Zhang [24].

In the following simulation studies, we consider the proposed tests corresponding to the uninformative weight function $w(x) \equiv 1$ and the optimal weight function.

Simulation 1: Here, the data are generated from the model

$$Y^* = \alpha + \beta X + \gamma X^2 + \varepsilon, \qquad Y = \max\{Y^*, 0\}. \tag{4.1}$$

In the simulation, $X \sim N(0, 1)$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, the true regression parameters are chosen to be $\alpha = 1$, $\beta = 1$ and $\sigma_\varepsilon^2 = 1$. We used a standard normal density function as the kernel function, and $h = n^{-1/5}$ as the bandwidth. Data from the model with $\gamma = 0$ are used to study the empirical size, while data from the models with $\gamma = 0.1, 0.2, 0.3, 0.5$ are used to study the empirical powers. Under the current setup, we can see that theoretically $P(\varepsilon \leq -1 - X) \approx 24\%$ observations of Y^* are truncated below 0 when $\gamma = 0$. The `censReg` function in the R package `censReg` is used to calculate the estimates of all unknown parameters.

Table 1
$d = 1$. Empirical powers.

		100	300	500	100	300	500
$\gamma = 0.0$	Test I	0.006	0.004	0.005	0.040	0.035	0.055
	Test II	0.010	0.006	0.011	0.050	0.065	0.060
$\gamma = 0.1$	Test I	0.016	0.049	0.091	0.075	0.140	0.225
	Test II	0.008	0.014	0.035	0.085	0.080	0.095
$\gamma = 0.2$	Test I	0.086	0.447	0.778	0.270	0.600	0.900
	Test II	0.015	0.095	0.191	0.065	0.180	0.350
$\gamma = 0.3$	Test I	0.353	0.924	0.996	0.580	0.970	1.000
	Test II	0.029	0.225	0.528	0.110	0.385	0.645
$\gamma = 0.5$	Test I	0.903	1.000	1.000	0.980	1.000	1.000
	Test II	0.136	0.710	0.970	0.375	0.860	1.000

For the sake of convenience, we denote the test based on model (2.2) as Test I, and the one based on model (2.3) as Test II in which $\hat{\xi}_i$ is replaced by $I(Y_i = 0) - F_\varepsilon(-m(X_i, \hat{\theta}))$. First, the uninformative weight function $w(x) = 1$ is considered. The simulation results are presented in the left hand side of Table 1. The simulation shows that the empirical levels are all less than the nominal levels in all the chosen cases, hence the proposed tests are conservative. This is very common for nonparametric smoothing tests. The test has small powers against the alternative models for small sample sizes, but the power improves as sample sizes get larger. As discussed in Section 2, Test II is much less powerful than Test I.

In general, bootstrap provides more accurate approximation to the distribution of the test statistic than the asymptotic normal distribution. Since the distribution of ε is normal with mean 0 and unknown variance σ_ε^2, so a viable parametric bootstrap algorithm can be developed as follows:

(1) Obtain the estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}_\varepsilon^2$ for α, β and σ_ε^2 based on the sample (Y_i, X_i), $i = 1, 2, \ldots, n$, and calculate the test statistic T_n;
(2) At stage b, generate n i.i.d. random numbers from $N(0, \hat{\sigma}_\varepsilon^2)$, denote them as ε_i^b, and calculate $Y_i^{*b} = \hat{\alpha} + \hat{\beta} X_i + \varepsilon_i^b$ and $Y_i^b = \max\{Y_i^{*b}, 0\}$, $i = 1, 2, \ldots, n$;
(3) Estimate $\alpha, \beta, \sigma_\varepsilon^2$ from data (Y_i^b, X_i), denote these estimators as $\hat{\alpha}^b$, $\hat{\beta}^b$ and $\hat{\sigma}_\varepsilon^b$, and calculate the test statistic T_n^b based on the residuals

$$\hat{\xi}_i^b = Y_i^b - [(\hat{\alpha}^b + \hat{\beta}^b X_i)\Phi((\hat{\alpha}^b + \hat{\beta}^b X_i)/\hat{\sigma}_\varepsilon^b) + \hat{\sigma}_\varepsilon^b \phi((\hat{\alpha}^b + \hat{\beta}^b X_i)/\hat{\sigma}_\varepsilon^b)]$$

in Test I, and $\hat{\xi}_i^b = I(Y_i^b = 0) - \Phi(-(\hat{\alpha}^b + \hat{\beta}^b X_i)/\hat{\sigma}_\varepsilon^b)$ in Test II;
(4) Repeat step (2)–(3) B times, and find out the 2.5-th and 97.5-th percentiles from $T_n^1, T_n^2, \ldots, T_n^B$.
(5) Reject the null hypothesis if T_n is less than the 2.5-th or larger than the 97.5-th percentiles found in Step (4).

In our simulation studies, B is chosen to be 200, and in each scenario, the above bootstrap algorithm is repeated 200 times. The consistency of the above bootstrap algorithm could be justified by adapting the method in Beran [4] and in Section 4.5 from Shao and Tu [22]. This bootstrap algorithm was also used by Drukker [8] for bootstrapping a conditional moments test for normality in the Tobit regression model. The right hand side of Table 1 reports the simulation results. As expected, all the empirical levels are close to the nominal level 0.05, and the powers are all larger than the ones reported in the left panel of Table 1.

For comparison purposes, we also conduct simulation studies based on the optimal weight functions. Let $z = (\alpha + \beta x)/\sigma_\varepsilon$. For model (2.2), the optimal weight function is $w(x) = \Phi(z)x^2/\tau^2(x)$, and $\tau^2(x) = \sigma_\varepsilon^2[1+z^2]\Phi(z) + \sigma_\varepsilon^2 z\phi(z) - \sigma_\varepsilon^2[z\Phi(z) + \phi(z)]^2$. For model (2.3), the optimal weight function is $w(x) = \phi(z)x^2/\tau^2(x)$, and $\tau^2(x) = \Phi(z)(1 - \Phi(z))$. Again, α, β and σ_ε are estimated as above. Simulation results are presented in Table 2. The left panel is for the test based on the model (2.2), and the right panel is for the test based on the model (2.3). It is easily seen that the empirical levels are closer to 0.05 than the ones reported in Table 1, and as expected, the empirical powers are larger than the corresponding ones in Table 1 for large sample sizes. Recall that the optimal weight functions are indeed "optimal" asymptotically, so it is no surprise to see the empirical powers in Table 2 being less than the ones reported in Table 1 for small sample sizes. Also, Test I in general is more powerful than Test II with the optimal weight.

Simulation 2: To see the performance of the proposed test for $d > 1$, we generated the data from the model

$$Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \gamma(X_1^2 + X_2^2) + \varepsilon, \qquad Y = \max\{Y^*, 0\}. \tag{4.2}$$

In the simulation, (X_1, X_2) is from a bivariate normal distribution with 0 mean vector, and identity covariance matrix, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, the true parameters are chosen to be $\alpha = \beta_1 = \beta_2 = \sigma_\varepsilon^2 = 1$. We choose the product of two standard normal density functions as the kernel function, and $h = n^{-1/7}$ as the bandwidth. Data from the model with $\gamma = 0$ are used to study the empirical size, while data from the models with $\gamma = 0.1, 0.2, 0.3, 0.5$ are used to study the empirical powers. In the current setup, we can see that theoretically $P(\varepsilon \leq -1 - X_1 - X_2) \approx 28\%$ observations of Y^* are truncated below 0 when $\gamma = 0$.

Table 2
Optimal weights.

	Test I			Test II		
	100	300	500	100	300	500
$\gamma = 0.0$	0.007	0.014	0.014	0.036	0.018	0.028
$\gamma = 0.1$	0.016	0.078	0.213	0.019	0.057	0.117
$\gamma = 0.2$	0.077	0.584	0.877	0.065	0.358	0.654
$\gamma = 0.3$	0.267	0.942	0.991	0.157	0.763	0.963
$\gamma = 0.5$	0.715	0.987	0.994	0.611	0.994	1.000

Table 3
$d = 2$. Empirical powers.

		100	300	500	100	300	500
$\gamma = 0.0$	Test I	0.002	0.008	0.010	0.040	0.035	0.055
	Test II	0.008	0.012	0.006	0.050	0.065	0.060
$\gamma = 0.1$	Test I	0.034	0.158	0.283	0.075	0.140	0.225
	Test II	0.010	0.023	0.038	0.085	0.080	0.095
$\gamma = 0.2$	Test I	0.247	0.838	0.985	0.270	0.600	0.900
	Test II	0.021	0.090	0.207	0.065	0.180	0.350
$\gamma = 0.3$	Test I	0.690	0.999	1.000	0.580	0.970	1.000
	Test II	0.036	0.235	0.494	0.110	0.385	0.645
$\gamma = 0.5$	Test I	0.997	1.000	1.000	0.980	1.000	1.000
	Test II	0.102	0.622	0.948	0.375	0.860	1.000

Table 4
$d = 2$. Empirical powers when testing simple null hypothesis.

	$d = 1$			$d = 2$		
	100	300	500	100	300	500
$\gamma = 0.0$	0.038	0.043	0.046	0.047	0.057	0.053
$\gamma = 0.1$	0.075	0.126	0.189	0.117	0.302	0.524
$\gamma = 0.2$	0.195	0.557	0.847	0.438	0.962	0.998
$\gamma = 0.3$	0.454	0.958	0.999	0.859	1.000	1.000
$\gamma = 0.5$	0.922	1.000	1.000	1.000	1.000	1.000

The simulation results for Test I and Test II are presented in the left panel of Table 3. The test again appears conservative, and the power increases with increasing sample size. We also did some simulations when X_1 and X_2 are weakly and moderately correlated. The results are not reported here because of their similarity to the left panel of Table 3. It is easily seen that Test II is less powerful than Test I. Similar to the one dimensional case, we also conduct a parametric bootstrap simulation and the results are shown in the right panel of Table 3. Clearly, the nominal level 0.05 is preserved in the bootstrap simulation and the power is much larger than the one shown in the left panel of Table 3.

For comparison purposes, a simulation study for simple null hypotheses based on Test I is also conducted. Using the same setups as in Simulation 1 and 2, but assuming that $\alpha = \beta = \beta_1 = \beta_2 = \sigma_\varepsilon^2 = 1$ are all known in the null models, we obtain the simulation results as shown in Table 4. Similar patterns as in the previous tables are also appeared in Table 4, but the empirical level is much closer to the nominal level 0.05 in all cases.

We also made a comparison study with Song and Zhang [24]'s (SZ) test procedure, which can be viewed as a variant of HM-test in which the weight function is so chosen that the denominator is not present in the test statistic, and the results are mixed. Based on the model (2.1) and the critical values from the normal theory, if all the parameters are unknown, then the proposed test is slightly more powerful than the SZ test when $d = 1$, but less powerful for the case of $d = 2$. If bootstrap critical values are used, then the SZ test is more powerful than the KSL test. If only truncation information is used, then the KSL test outperforms the SZ test for both $d = 1$ and $d = 2$ cases. This is also true when testing the simple hypotheses.

We also conducted a finite sample comparison study between the HM test, the D-test described in Section 3, and the test proposed in this paper corresponding to $w(x) = 1$, labeled as the KSL-test, using the same model as in Simulation 1. The bandwidth $h = n^{-2.3/5}$ is used in the simulation for all three tests, with such a choice, $nh^2 \to \infty$, and $nh^2\sqrt{h} \to 0$. The simulation results are shown in Table 5. In addition to sample size 100, 300, 500, we also conduct the simulation for $n = 800, 1000$. The simulation results show that, when the tests are based on model (2.2), the empirical level of the D-test is much smaller than those of KSL-test and HM-test, which indicates that the former test is more conservative than the latter two. As far as the power is concerned, the HM-test performs somewhat better than the KSL-test, and the KSL-test is more powerful than the D-test. A similar pattern appears when the tests are constructed from the model (2.3), with a few exceptions between the KSL-test and D-test when $\gamma = 0.3, 0.5$ and the sample sizes are around 800 and 1000. As expected, all the tests based on model (2.2) are more powerful than the ones based on model (2.3). Although the KSL-test is somewhat

Table 5
Comparisons between HM, KSL, and D-tests.

	Test I					Test II				
	100	300	500	800	1000	100	300	500	800	1000
	HM-Test									
$\gamma = 0.0$	0.023	0.031	0.040	0.044	0.039	0.023	0.017	0.032	0.037	0.034
$\gamma = 0.1$	0.028	0.064	0.078	0.155	0.198	0.026	0.042	0.065	0.074	0.084
$\gamma = 0.2$	0.099	0.346	0.626	0.897	0.964	0.037	0.137	0.235	0.485	0.594
$\gamma = 0.3$	0.250	0.861	0.985	1.000	1.000	0.079	0.368	0.679	0.932	0.978
$\gamma = 0.5$	0.835	0.999	1.000	1.000	1.000	0.262	0.892	0.996	1.000	1.000
	KSL-Test									
$\gamma = 0.0$	0.026	0.029	0.031	0.026	0.034	0.025	0.024	0.026	0.035	0.036
$\gamma = 0.1$	0.026	0.059	0.061	0.092	0.113	0.021	0.033	0.054	0.050	0.039
$\gamma = 0.2$	0.098	0.252	0.408	0.659	0.783	0.031	0.070	0.093	0.154	0.200
$\gamma = 0.3$	0.210	0.702	0.939	0.997	0.999	0.048	0.117	0.237	0.394	0.494
$\gamma = 0.5$	0.791	0.998	1.000	1.000	1.000	0.096	0.372	0.649	0.900	0.951
	D-Test									
$\gamma = 0.0$	0.007	0.009	0.008	0.013	0.011	0.008	0.009	0.014	0.021	0.027
$\gamma = 0.1$	0.004	0.012	0.019	0.047	0.055	0.004	0.011	0.026	0.034	0.026
$\gamma = 0.2$	0.016	0.069	0.204	0.417	0.582	0.003	0.025	0.053	0.110	0.182
$\gamma = 0.3$	0.043	0.336	0.656	0.922	0.976	0.009	0.053	0.141	0.365	0.533
$\gamma = 0.5$	0.249	0.849	0.977	0.997	1.000	0.016	0.243	0.616	0.942	0.987

inferior to the HM-test in this limited simulation, it is still a competitive candidate for model checking due to its relative computational ease, especially in the case when the number of predictors d is large.

Another interesting phenomenon found in the above simulation study is that with the smaller bandwidth $h = n^{-2.3/5}$, the empirical levels are much closer to the nominal level 0.05 compared to the ones using the larger bandwidth $h = n^{-1/5}$.

5. Proofs of the main results

The proof of Theorem 3.1 is facilitated by the lemmas stated below.

Lemma 5.1. *Under condition* (C1), $aQ_0(-a) + Q_1(-a) = \int_{-a}^{\infty}[1 - F_\varepsilon(u)]du$, *for all* $a \in \mathbb{R}$.

Proof. (C1) implies $E|\varepsilon| < \infty$. Hence $\lim_{a\to\infty} a[1 - F_\varepsilon(a)] = 0$. This fact and integration by parts shows

$$\begin{aligned} aQ_0(-a) + Q_1(-a) &= a[1 - F_\varepsilon(-a)] + \int_{-a}^{\infty} uf_\varepsilon(u)du \\ &= a[1 - F_\varepsilon(-a)] - \int_{-a}^{\infty} ud[1 - F_\varepsilon(u)] \\ &= a[1 - F_\varepsilon(-a)] - u[1 - F_\varepsilon(u)]\Big|_{-a}^{\infty} + \int_{-a}^{\infty}[1 - F_\varepsilon(u)]du \\ &= \int_{-a}^{\infty}[1 - F_\varepsilon(u)]du. \end{aligned}$$

Hence the lemma. □

Lemma 5.2. *Under condition* (C1), *there exists a constant* B *such that,*

$$|g(-m(x, \theta_1)) - g(-m(x, \theta_2)) - [m(x, \theta_1) - m(x, \theta_2)]Q_0(-m(x, \theta_2))| \le B[m(x, \theta_1) - m(x, \theta_2)]^2, \quad \forall x \in \mathbb{R}^d.$$

Proof. A Taylor expansion of g function and an application of Lemma 5.1, together with the boundedness of f_ε, imply the result. □

To find out the asymptotic distribution of the test statistic, we also need Theorem 1 of Hall [10] which is reproduced here for the sake of completeness.

Lemma 5.3. *Let* Z_i, $i = 1, 2, \ldots, n$ *be i.i.d. random vectors, and let*

$$U_n = \sum_{1\le i<j\le n} H_n(Z_i, Z_j), \qquad M_n(x, y) = EH_n(Z_1, x)H_n(Z_1, y),$$

where H_n is a sequence of measurable functions symmetric under permutation, with

$$E[H_n(Z_1,Z_2)|Z_1]=0,\ \text{a.s.}\quad\text{and}\quad EH_n^2(Z_1,Z_2)<\infty\quad\text{for each } n\geq 1.$$

If $[EM_n^2(Z_1,Z_2)+n^{-1}H_n^4(Z_1,Z_2)]/[EH_n^2(Z_1,Z_2)]^2\to 0$, *then* U_n *is asymptotically normally distributed with mean zero and variance* $n^2EH_n^2(Z_1,Z_2)/2$.

Proof of Theorem 3.1. Let $\xi_i = Y_i - g(X_i,\theta_0)$, $K_h(y) := h^{-d}K(y/h)$, and $w_{ij} := w(X_i)w(X_j)$, $1\leq i,j\leq n$. Then $\hat{\xi}_i = Y_i - g(X_i,\hat{\theta}_n) = \xi_i - [g(X_i,\hat{\theta}_n)-g(X_i,\theta_0)]$. Hence V_n can be written as the sum of the following three terms

$$V_{1n}=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)\xi_i\xi_j w_{ij},$$

$$V_{2n}=-\frac{2}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)\xi_i[g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)]w_{ij},$$

$$V_{3n}=\frac{2}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)[g(X_i,\hat{\theta}_n)-g(X_i,\theta_0)][g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)]w_{ij}.$$

Denote $Z_i=(X_i,\xi_i)$, V_{1n} can be written in a U-statistic form with

$$H_n(Z_i,Z_j)=K_h(X_i-X_j)\xi_i\xi_j w_{ij}.$$

Under the null hypothesis, $E[H_n(Z_1,Z_2)|Z_1]=0$, so V_{1n} is a degenerate U-statistic. We shall apply Lemma 5.3 to show the asymptotic normality of V_{1n}. For this purpose, we need to investigate the asymptotic behavior of $EM_n^2(Z_1,Z_2)$, $EH_n^4(Z_1,Z_2)$, and $EH_n^2(Z_1,Z_2)$, where $M_n(x,y)=EH_n(Z_1,x)H_n(Z_1,y)$ is defined as in Lemma 5.3. Consider

$$\begin{aligned}
EM_n^2(Z_1,Z_2) &= E(E[H_n(Z_3,Z_1)H_n(Z_3,Z_2)|Z_1,Z_2])^2\\
&= E\left(E\left[\frac{1}{h^{2d}}K\left(\frac{X_3-X_1}{h}\right)K\left(\frac{X_3-X_2}{h}\right)w(X_1)w(X_2)w^2(X_3)\xi_1\xi_2\xi_3^2\middle|Z_1,Z_2\right]\right)^2\\
&= \frac{1}{h^{4d}}E\left(\xi_1\xi_2 w(X_1)w(X_2)\int K\left(\frac{x_3-X_1}{h}\right)K\left(\frac{x_3-X_2}{h}\right)w^2(x_3)\tau^2(x_3)f_X(x_3)dx_3\right)^2\\
&= \frac{1}{h^{2d}}\iint \tau^2(x_1)\tau^2(x_2)w^2(x_1)w^2(x_2)\\
&\quad\times\left[\int K(u)K\left(u+\frac{x_1-x_2}{h}\right)w^2(x_1+hu)\cdot\tau^2(x_1+hu)f_X(x_1+hu)du\right]^2 f_X(x_1)f_X(x_2)dx_1dx_2\\
&= \frac{1}{h^d}\int\left[\int K(u)K(u+v)du\right]^2 dv\int[\tau^2(x)]^4 w^8(x)f_X^4(x)dx+o(1/h^d).
\end{aligned}$$

For $EH_n^2(Z_1,Z_2)$, we have

$$\begin{aligned}
EH_n^2(Z_1,Z_2) &= E(E[H_n^2(Z_1,Z_2)|X_1,X_2])\\
&= \frac{1}{h^{2d}}\int K^2\left(\frac{x_1-x_2}{h}\right)\tau^2(x_1)\tau^2(x_2)w^2(x_1)w^2(x_2)f_X(x_1)f_X(x_2)dx_1dx_2\\
&= \frac{1}{h^d}\int K^2(u)du\int(\tau^2(x))^2 w^4(x)f_X^2(x)dx+o(1/h^m).
\end{aligned}$$

Similarly,

$$EH_n^4(Z_1,Z_2)=\frac{1}{h^{3d}}\int K^4(u)du\int(\tau^2(x))^4 w^8(x)f_X^2(x)dx+o(1/h^{3d}).$$

Therefore, from (C6), we obtain

$$\frac{EM_n^2(Z_1,Z_2)+n^{-1}EH_n^4(Z_1,Z_2)}{[EH_n^2(Z_1,Z_2)]^2}=\frac{O(1/h^d)+O(1/(nh^{3d}))}{O(1/h^{2d})}=O(h^d)+O(1/(nh^d))\to 0.$$

Hence Lemma 5.3 is applicable and

$$nh^{d/2}V_{1n}\to_D N(0,\sigma^2), \tag{5.1}$$

where σ^2 is defined in (2.5).

Now consider V_{2n}. Let

$$d_{ni} = g(X_j, \hat{\theta}_n) - g(X_j, \theta_0) - [m(X_i, \hat{\theta}_n) - m(X_i, \theta_0)]Q_0(-m(X_i, \theta_0)), \tag{5.2}$$

$$\delta_{ni} = m(X_i, \hat{\theta}_n) - m(X_i, \theta_0) - (\hat{\theta}_n - \theta_0)'\dot{m}(X_i, \theta_0). \tag{5.3}$$

Then, V_{2n} can be written as the sum $V_{2n1} + V_{2n2}$, where

$$V_{2n1} = -\frac{2}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)\xi_i d_{nj} w_{ij},$$

$$V_{2n2} = -\frac{2}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)\xi_i [m(X_j, \hat{\theta}_n) - m(X_j, \theta_0)]Q_0(-m(X_j, \theta_0)) w_{ij}.$$

Applying Lemma 5.2 for $\theta_1 = \hat{\theta}_n, \theta_2 = \theta_0$, and $x = X_i, i = 1, 2, \ldots, n$, we have

$$|V_{2n1}| \le \frac{4B}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i|\delta_{nj}^2 w_{ij} + \frac{4B}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i|[(\hat{\theta}_n - \theta_0)'\dot{m}(X_j, \theta_0)]^2 w_{ij}$$
$$= A_{n1} + A_{n2}.$$

By Condition (C4), A_{n1} is bounded above by

$$O_p\left(\frac{1}{n^2}\right) \cdot \frac{1}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i| w_{ij}.$$

Note that

$$\frac{1}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i| w_{ij} = O_p(1).$$

Therefore, $nh^{d/2}A_{n1} = o_p(1)$ from Condition (C6). For A_{n2}, we have

$$nh^{d/2}A_{n2} \le nh^{d/2}\|\hat{\theta}_n - \theta_0\|^2 \cdot \frac{4B}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i|\|\dot{m}(X_j, \theta_0)\|^2 w_{ij} = o_p(1)$$

by the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and

$$\frac{1}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i|\|\dot{m}(X_j, \theta_0)\|^2 w_{ij} = O_p(1).$$

Hence

$$nh^{d/2}V_{2n1} = o_p(1). \tag{5.4}$$

Adding and subtracting $(\hat{\theta}_n - \theta_0)'\dot{m}(X_j, \theta_0)$ from $m(X_j, \hat{\theta}_n) - m(X_j, \theta_0)$, V_{2n2} can be written as the sum of the following two terms

$$B_{n1} = -\frac{2}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)\xi_i \delta_{nj} Q_0(-m(X_j, \theta_0)) w_{ij},$$

$$B_{n2} = -\frac{2}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)\xi_i (\hat{\theta}_n - \theta_0)'\dot{m}(X_j, \theta_0) Q_0(-m(X_j, \theta_0)) w_{ij}.$$

By Condition (C4),

$$|B_{n1}| \le \sup_{1 \le i \le n} |\delta_{ni}| \cdot \frac{2}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)|\xi_i| w_{ij} = O_p(1/n),$$

thus, $nh^{d/2}B_{n1} = o_p(1)$. As for B_{n2}, it is easily seen that

$$|B_{n2}| \le \|\hat{\theta}_n - \theta_0\| \cdot \left\| \frac{2}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)\xi_i m(X_j, \theta_0) Q_0(-m(X_j, \theta_0)) \right\| w_{ij}.$$

Using the similar method as in proving Lemma 3.3b in [32], one can show that the second norm in the above inequality is of the order of $O_p(1/\sqrt{n})$, which, together with the $\sqrt{n}$-consistency of $\hat{\theta}_n$, implies $nh^{d/2}B_{n2} = o_p(1)$. Thus,

$$nh^{d/2}V_{2n2} = o_p(1). \tag{5.5}$$

From (5.4) and (5.5), we obtain

$$nh^{d/2}V_{2n} = o_p(1). \tag{5.6}$$

The proof of $nh^{d/2}V_{3n} = o_p(1)$ follows the same thread as above, hence is omitted here for the sake of brevity.

To show that $\hat{\sigma}^2$ defined in (2.6), it is sufficient to show that

$$\frac{2}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)(\hat{\xi}_i^2\hat{\xi}_j^2-\xi_i^2\xi_j^2)w^2(X_i)w^2(X_j) = o_p(1), \tag{5.7}$$

$$\frac{2}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i^2\xi_j^2 w^2(X_i)w^2(X_j) = \sigma^2 + o_p(1). \tag{5.8}$$

Adding and subtracting ξ_i from $\hat{\xi}_i$, ξ_j from $\hat{\xi}_j$, the term on the left hand side of (5.7) can be written as the sum of the following five terms,

$$\frac{2}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)(\hat{\xi}_i-\xi_i)^2(\hat{\xi}_j-\xi_j)^2 w^2(X_i)w^2(X_j),$$

$$\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j(\hat{\xi}_i-\xi_i)(\hat{\xi}_j-\xi_j)w^2(X_i)w^2(X_j),$$

$$\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i(\hat{\xi}_i-\xi_i)(\hat{\xi}_j-\xi_j)^2 w^2(X_i)w^2(X_j),$$

$$\frac{4}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i^2(\hat{\xi}_j-\xi_j)^2 w^2(X_i)w^2(X_j),$$

$$\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2(\hat{\xi}_i-\xi_i)w^2(X_i)w^2(X_j). \tag{5.9}$$

We only show that (5.9) is the order of $o_p(1)$. Note that

$$\hat{\xi}_i-\xi_i = -d_{ni}-\delta_{ni}Q_0(-m(X_i,\theta_0))-(\hat{\theta}_n-\theta_0)'\dot{m}(X_i,\theta_0)Q_0(-m(X_i,\theta_0)),$$

so it suffices to show the following three terms are all of the order $o_p(1)$,

$$-\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2 d_{ni}w^2(X_i)w^2(X_j), \tag{5.10}$$

$$-\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2\delta_{ni}Q_0(-m(X_i,\theta_0))w^2(X_i)w^2(X_j), \tag{5.11}$$

$$-\frac{8(\hat{\theta}_n-\theta_0)'}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2\dot{m}(X_i,\theta_0)Q_0(-m(X_i,\theta_0))w^2(X_i)w^2(X_j). \tag{5.12}$$

For any continuous function $L_1(x)$, $L_2(x)$ such that $E[L_1^2(X)+L_2^2(X)]<\infty$, we can show that

$$E\left[\frac{1}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)|\xi_i|\xi_j^2L_1(X_i)L_2(X_j)\right] = 2E\frac{1}{h^d}K^2\left(\frac{X_1-X_2}{h}\right)|\xi_1|\xi_2^2L_1(X_1)L_2(X_2)$$
$$\leq \int K^2(u)du\int\tau^3(x)L_1(x)L_2(x)f_X^2(x)dx + o(1).$$

If the leading term in the above upper bound is finite, then by the Markov inequality,

$$\left[\frac{1}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)|\xi_i|\xi_j^2L_1(X_i)L_2(X_j)\right]=O_p(1).$$

Because $\sup_{1\le i\le n}|d_{ni}|$ and $\sup_{1\le i\le n}|\delta_{ni}|$ are both negligible, it is easily seen that (5.10)–(5.12) are all of the order of $o_p(1)$.

The proof of (5.8) is similar to the proof of Lemma 3.3(e) in Zheng [32]. This completes the proof of the theorem. □

Proof of Theorem 3.2. The proof is similar to that of Theorem 3.1. We only outline the main steps here for the sake of brevity.

Substituting $Y_i-g_\gamma(X_i)+g_\gamma(X_i)-g(X_i,\hat\theta_n)=\xi_i+g_\gamma(X_i)-g(X_i,\hat\theta_n)$ for $\hat\xi_i$ and $Y_j-g_\gamma(X_j)+g_\gamma(X_j)-g(X_j,\hat\theta_n)=\xi_j+g_\gamma(X_j)-g(X_j,\hat\theta_n)$ for $\hat\xi_j$ in V_n, V_n can be written as the sum of the following three terms

$$V_{1n}^a=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)\xi_i\xi_jw_{ij},$$

$$V_{2n}^a=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)\xi_i[g_\gamma(X_j)-g(X_j,\hat\theta_n)]w_{ij},$$

$$V_{3n}^a=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)[g_\gamma(X_i)-g(X_i,\hat\theta_n)][g_\gamma(X_j)-g(X_j,\hat\theta_n)]w_{ij}.$$

V_{3n}^a can be further written as the sum $V_{3n1}^a+V_{3n2}^a+V_{3n3}^a$, where

$$V_{3n1}^a=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)[g_\gamma(X_i)-g(X_i,\theta_a)][g_\gamma(X_j)-g(X_j,\theta_a)]w_{ij},$$

$$V_{3n2}^a=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)[g_\gamma(X_i)-g(X_i,\theta_a)][g(X_j,\theta_a)-g(X_j,\hat\theta_n)]w_{ij},$$

$$V_{3n3}^a=\frac{1}{n(n-1)}\sum_{i\neq j}K_h(X_i-X_j)[g(X_i,\theta_a)-g(X_i,\hat\theta_n)][g(X_j,\theta_a)-g(X_j,\hat\theta_n)]w_{ij}.$$

One can show that $V_{3n1}^a=\int K^2(u)du\cdot\int[g_\gamma(x)-g(x,\theta_a)]^2w^2(x)f_X^2(x)dx+o_p(1)$, and $V_{3n2}^a=o_p(1)$, $V_{3n3}^a=o_p(1)$, $V_{2n}^a=o_p(1)$. Eventually, one can show that

$$nh^{d/2}V_n=nh^{d/2}V_{1n}^a+nh^{d/2}\int K^2(u)du\cdot\int[g_\gamma(x)-g(x,\theta_a)]^2w^2(x)f_X^2(x)dx+o_p(nh^{d/2}).$$

Finally, we can show that

$$\hat\sigma^2=2\int K^2(u)du\cdot\int[\tau^2(x)+[g_\gamma(x)-g(x,\theta_a)]^2]^2w^4(x)f_X^2(x)dx+o_p(1).$$

This completes the proof. □

Proof of Theorem 3.3. Now, define $Y_i^{*L}=m(X_i,\theta_0)+\varepsilon_i$, $Y_i^L=\max\{Y_i^{*L},0\}$, and $W_i=Y_i-Y_i^L$. The elementary inequality $\max\{a,0\}=(a+|a|)/2$ implies $W_i=[\delta(X_i)+\Delta_n(X_i)]/2\sqrt{nh^{d/2}}$ with

$$\Delta_n(X_i)=\left|\sqrt{nh^{d/2}}m(X_i,\theta_0)+\delta(X_i)+\sqrt{nh^{d/2}}\varepsilon_i\right|-\left|\sqrt{nh^{d/2}}m(X_i,\theta_0)+\sqrt{nh^{d/2}}\varepsilon_i\right|.$$

Define $\hat\xi_i^L=Y_i^L-g(X_i,\hat\theta_n)$. Then $\hat\xi_i=\hat\xi_i^L+W_i$ and V_n can be written as a sum of the following terms

$$V_{1n}^L=\frac{1}{n(n-1)h^d}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h}\right)\hat\xi_i^L\hat\xi_j^Lw_{ij},$$

$$V_{2n}^L=\frac{2}{n(n-1)h^d}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h}\right)\hat\xi_i^LW_jw_{ij},$$

$$V_{3n}^L=\frac{1}{n(n-1)h^d}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h}\right)W_iW_jw_{ij}.$$

Similar to the proof of Theorem 3.1, $nh^{d/2}V_{1n}^L\Rightarrow N(0,\sigma^2)$, where σ^2 is defined in (2.5).

It is easily seen that $\delta(x) + \Delta_n(x) = \delta(x)I(m(x,\theta_0) + \varepsilon > 0)$. By the independence of ε and X_1, X_2, EV_{3n}^L equals

$$\begin{aligned}
&\frac{1}{h^d} EK\left(\frac{X_1 - X_2}{h}\right) W_1 W_2 w(X_1) w(X_2) \\
&= \frac{1}{nh^{3d/2}} \iint K\left(\frac{x_1 - x_2}{h}\right) Q_0(-m(x_1,\theta_0)) Q_0(-m(x_2,\theta_0)) \delta(x_1)\delta(x_2) w(x_1) w(x_2) f_X(x_1) f_X(x_2) dx_1 dx_2 \\
&= \frac{1}{nh^{d/2}} \iint K(u) Q_0(-m(x+hu,\theta_0)) Q_0(-m(x,\theta_0)) \delta(x+hu)\delta(x) w(x+hu) w(x) f_X(x+hu) f_X(x) dx du \\
&= \frac{1}{nh^{d/2}} \int Q_0^2(-m(x,\theta_0)) \delta^2(x) w^2(x) f_X^2(x) dx + o\left(\frac{1}{nh^{d/2}}\right).
\end{aligned}$$

The last integral is exactly the μ defined in Theorem 3.3. Therefore,

$$nh^{d/2} EV_{3n}^L \to \mu. \tag{5.13}$$

Now consider $\mathrm{Var}(V_{3n}^L)$. For convenience, let

$$S_{ij} = K\left(\frac{X_i - X_j}{h}\right) W_i W_j w_{ij} - EK\left(\frac{X_i - X_j}{h}\right) W_i W_j w_{ij}.$$

A tedious but simple derivation leads to

$$\mathrm{Var}(V_{3n}^L) = \frac{4n(n-1)}{[n(n-1)h^d]^2} ES_{12}^2 + \frac{8n(n-1)(n-2)}{3![n(n-1)h^d]^2} ES_{12}S_{13}.$$

By the Cauchy–Schwarz inequality, $|ES_{12}S_{13}| \le ES_{12}^2$, and hence

$$\mathrm{Var}(V_{3n}^L) \le \left[\frac{4n(n-1)}{[n(n-1)h^d]^2} + \frac{8n(n-1)(n-2)}{3![n(n-1)h^d]^2}\right] ES_{12}^2.$$

One can show that

$$\begin{aligned}
ES_{12}^2 &= E\left[K\left(\frac{X_1 - X_2}{h}\right) W_1 W_2 w(X_1) w(X_2) - EK\left(\frac{X_1 - X_2}{h}\right) W_1 W_2 w(X_1) w(X_2)\right]^2 \\
&\le EK^2\left(\frac{X_1 - X_2}{h}\right) W_1^2 W_2^2 w^2(X_1) w^2(X_2).
\end{aligned}$$

Since $|W_i| \le |\delta(X_i)|/\sqrt{nh^{d/2}}$, for each i, we have

$$ES_{12}^2 \le \frac{1}{n^2 h^d} EK^2\left(\frac{X_1 - X_2}{h}\right) \delta^2(X_1)\delta^2(X_2) w^2(X_1) w^2(X_2) = O\left(\frac{1}{n^2}\right).$$

Therefore,

$$\mathrm{Var}(V_{3n}^L) = \left[\frac{4n(n-1)}{[n(n-1)h^d]^2} + \frac{8n(n-1)(n-2)}{3![n(n-1)h^d]^2}\right] \cdot \frac{1}{n^2} = O\left(\frac{1}{n^3 h^{2d}}\right),$$

which implies

$$V_{3n}^L - EV_{3n}^L = O_p\left(\frac{1}{\sqrt{n^3 h^{2d}}}\right). \tag{5.14}$$

From (5.13) and (5.14), one now readily obtains

$$nh^{d/2} V_{3n}^L = nh^{d/2}[V_{3n}^L - EV_{3n}^L] + nh^{d/2} EV_{3n}^L \to_p \mu.$$

Similarly, one can show that $nh^{d/2} V_{2n}^L = o_p(1)$, and $\hat{\sigma}^2 \to_p \sigma^2$. The details are omitted for the sake of brevity. Summarizing the above arguments, we can finish the proof of Theorem 3.3. □

Acknowledgments

The authors would like to thank the Editor and the referees for their critical comments that helped to substantially improve the presentation of the paper.

References

[1] A.A. Adesina, M.M. Zinnah, Technology characteristics, farmers' perceptions and adoption decisions: a Tobit model application in Sierra Leone, Agric. Econ. 9 (4) (1993) 297–311.
[2] T. Amemiya, Regression analysis when the dependent variable is truncated normal, Econometrica 41 (1973) 997–1016.
[3] T. Amemiya, Tobit models: a survey, J. Econometrics 24 (1–2) (1984) 3–61.
[4] R. Beran, Simulated power functions, Ann. Statist. 14 (1986) 151–173.
[5] J.R. Blaylock, W.N. Blisard, Women and the demand for alcohol: estimating participation and consumption, J. Consum. Aff. 27 (2) (1993) 319–334.
[6] H. Dette, A consistent test for the functional form of a regression based on a difference of variance estimators, Ann. Statist. 27 (1999) 1012–1040.
[7] H. Dette, von Lieres, C. und Wilkau, Testing additivity by kernel-based methods—what is a reasonable test? Bernoulli 7 (2001) 669–697.
[8] D.M. Drukker, Bootstrapping a conditional moments test for normality after Tobit estimation, Stat J. 2 (2002) 125–139.
[9] C. Ekstrand, T.E. Carpenter, Using a Tobit regression model to analyse risk factors for foot-pad dermatitis in commercially grown broilers, Preventive Veterinary Med. 37 (1–4) (1998) 219–228.
[10] P.J. Hall, Central limit theorem for integrated square error of multivariate nonparametric density estimators, J. Multivariate Anal. 14 (1) (1984) 1–16.
[11] W. Haïdle, E. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Statist. 21 (4) (1993) 1926–1947.
[12] J.J. Heckman, The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models, Ann. Econom. Social Meas. 5 (4) (1976) 475–492.
[13] J.J. Heckman, Sample selection bias as a specification error, Econometrica 47 (1) (1979) 153–162.
[14] J.L. Horowitz, G.R. Neumann, Specification testing in censored regression models: parametric and semiparametric methods, J. Appl. Econometrics 4 (1989) 61–86. supplement issue.
[15] R.I. Jennrich, Asymptotic properties of non-linear least squares estimators, Ann. Math. Statist. 40 (2) (1969) 633–643.
[16] Hira L. Koul, P. Ni, Minimum distance regression model checking, J. Statist. Plann. Inference 119 (1) (2004) 109–141.
[17] A. Lewbel, O.B. Linton, Nonparametric censored and truncated regression, Econometrica 70 (2002) 765–779.
[18] E. Lichtenberg, L.K. Shapiro, Agriculture and nitrate concentrations in Maryland community water system wells, J. Environ. Qual. 26 (1) (1997) 145–153.
[19] O. Lopez, V. Patilea, Nonparametric lack-of-fit tests for parametric mean-regression models with censored data, J. Multivariate Anal. 100 (1) (2009) 210–230.
[20] C.E. McConnel, M.R. Zetzman, Urban/rural differences in health service utilization by elderly persons in the United States, J. Rural Health 9 (4) (1993) 270–280.
[21] J.L. Powell, Least absolute deviations estimation for the censored regression model, J. Econometrics 25 (3) (1984) 303–325.
[22] J. Shao, D. Tu, The Jackknife and Bootstrap, Springer-Verlag, New York, Inc., 1995.
[23] W.X. Song, Distribution-free test in Tobit mean regression model, J. Statist. Plann. Inference 141 (8) (2011) 2891–2901.
[24] W.X. Song, Y. Zhang, Empirical L_2-distance lack-of-fit tests for Tobit regression models, J. Multivariate Anal. 111 (2012) 380–396.
[25] J. Tobin, Estimation of relationships for limited dependent variables, Econometrica 26 (1) (1958) 24–36.
[26] L. Wang, A simple nonparametric test for diagnosing nonlinearity in Tobit median regression model, Statist. Probab. Lett. 77 (10) (2007) 1034–1042.
[27] L. Wang, X. Zhou, Assessing the adequacy of variance function in heteroscedastic regression models, Biometrics 63 (2007) 1218–1225.
[28] H. White, Consequences and detection of misspecified nonlinear regression models, J. Amer. Statist. Assoc. 76 (374) (1981) 419–433.
[29] H. White, Maximum likelihood estimation of misspecified models, Econometrics 50 (1) (1982) 1–25.
[30] Chien-Fu Wu, Asymptotic theory of nonlinear least squares estimation, Ann. Statist. 9 (1981) 501–513.
[31] C. Zhang, H. Dette, A power comparison between nonparametric regression tests, Statist. Probab. Lett. 66 (2004) 289–301.
[32] J.X. Zheng, A consistent test of functional form via nonparametric estimation techniques, J. Econometrics 75 (2) (1996) 263–289.
[33] X.B. Zhou, Semiparametric and nonparametric estimation of Tobit models, Ph.D. Thesis, Department of Economics, Hong Kong University of Science and Technology, 2007.