Journal of Multivariate Analysis 111 (2012) 380–396

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Empirical L_2-distance lack-of-fit tests for Tobit regression models

Weixing Song *, Yi Zhang

Kansas State University, United States

ARTICLE INFO

Article history:
Received 26 May 2011
Available online 2 May 2012

AMS subject classifications:
primary 62G08
secondary 62G10

Keywords:
Tobit regression model
Empirical L_2-distance
Consistency and local power

ABSTRACT

Standard Tobit regression models assume a linear relationship between the partially observed response variable and the predictors, while applications often see some nonlinear connections. This paper proposes an empirical L_2-distance lack-of-fit test to check the adequacy of the presumed parametric form for the regression function in Tobit regression models. The proposed test statistic is shown to be asymptotically normal, consistent against some fixed alternatives, and has nontrivial power for some local nonparametric alternatives. Simulation studies are conducted to assess the finite sample performance of the proposed test.

Published by Elsevier Inc.

1. Introduction

Tobin [24] demonstrated an example of using linear regression to study the relationship between the household expenditures on durable goods and income. The model takes account of the fact that the expenditure cannot be negative. Tobin named his model the model of limited dependent variables. Later on, the term *Tobit Models* was coined by Goldberger [6] to name Tobin's model because of its similarity to the *Probit* models. Nowadays, the Tobit regression model is a frequently used tool for modeling censored or truncated variables in many areas such as econometrics, biometrics, agriculture and engineering etc. For some empirical examples, see [2,4,18,16,1,5] and the references therein.

To be specific, let Y be a household's expenditure on a durable good, y_0 be the price of the cheapest available durable good, Z be all the other expenditure, and X be the income. Tobin [24] considered an utility maximization model in which a household is assumed to maximize utility $U(Y, Z)$ with the budget constraint $Y+Z \le X$, and the boundary constraint $Y \ge y_0$. Suppose Y^* is the solution of the maximization subject to $Y + Z \le X$ but ignoring the other constraint. Then the solution Y to the original problem can be defined by the renowned Tobit regression model: $Y = Y^* I(Y^* > y_0) + y_0 I(Y^* \le y_0)$. Usually, y_0 is assumed to be known. Without loss of generality, we shall assume $y_0 = 0$ throughout this paper. Amemiya [2] defined the standard Tobit model as follows:

$$Y_i^* = X_i'\beta + \varepsilon_i, \qquad Y_i = Y_i^* \ \text{ if } Y_i^* > 0 \text{ or 0 otherwise}, \ i = 1, 2, \ldots, n,$$

where ε_i are assumed to be i.i.d. copies from $N(0, \sigma^2)$, and the error term ε and the design variable X are independent. It is assumed that (X_i, Y_i)'s are observed for $i = 1, 2, \ldots, n$, but Y_i^* are unobserved if $Y_i^* \le 0$. The standard Tobit regression model is one of the five types Tobit regression models defined by Amemiya [2]. In this paper, we will mainly focus on the standard Tobit regression model. After certain modifications, the proposed lack-of-fit tests are also applicable for other four types Tobit regression models.

By assuming that the regression function $m(X)$ is linear, the existing work on the standard Tobit regression model mainly focuses on the estimation of the unknown regression parameters $\theta = (\beta', \sigma^2)'$. Under the normality assumption of the error

* Corresponding author.
E-mail address: weixing@ksu.edu (W. Song).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2012.04.015

term ε, Amemiya [3] and Heckman [10,11] proposed consistent estimators for θ, but these estimators are not consistent if the normality assumption fails. A robust estimator of θ was proposed by Powell [21] based on the least absolute deviations. See [2] for a comprehensive discussion on the estimation issue related to Tobit regression models.

The predetermined parametric form of the regression function is either based on some empirical evidence or simply for the sake of mathematical convenience. Misspecification of the regression function often results in misleading conclusions. For example, it is well known that violation of the linearity assumption can produce inconsistent estimators of the parameters and biased prediction of the survival time. See [12]. Therefore, it is theoretically important and practically significant to develop some formal numeric tests to check the adequacy of the selected regression functions.

Testing the specification of the regression function in Tobit models started in the early 1980's. Among others, Nelson [19]'s test is a general specification test which compares restricted and unrestricted estimates of various moments of the dependent variable, and it is not directed to any specific alternatives. The same is true for Olsen [20]'s test which compares the actual and predicted numbers of 0 observations. Rudd [22]'s suggestion followed the same thread and his test is for checking the significance of the difference between the Tobit and Probit estimates. Lee [15]'s procedure is a test of normality against the alternative of a more general member of the Pearson family. Lin and Schmidt [17] proposed a relatively simple test of the hypothesis that the Tobit model is correctly specified, against the alternative that different sets of parameters determine the probability of a 0 observation and the density of the non-0 observations. Wang [25] proposed a simple nonparametric test for checking the nonlinearity in Tobit median regression model in which the median of the random error is assumed to be 0. Compared with existing methods in the literature, the author claimed that the test has the advantage of allowing the alternative to be any smooth function and does not require any knowledge of the distribution of the random error. However, a problem that was not resolved in [25] is the selection of the window width. Song [23] developed a lack-of-fit testing procedure for a more general null hypothesis, not limited to linear functions, by assuming that the mean of ε is 0. The proposed test is based on the Khamaladze type transformation of a certain marked residual process. The transformed residual process converges weakly to a time-transformed Brownian motion in a uniform metric. Consequently, any test based on a continuous functional of this process is asymptotically distribution free, and can be implemented at least for moderate to large samples without resorting to a resampling method. Different from Wang [25]'s test, we can use some existing objective rules to select the bandwidth, such as the one to minimize the asymptotic integrated mean squares in estimation setup. Also, the proposed procedure can test any parametric regression functions rather than only linear ones, and the computation of the test statistic is very fast. The most restrictive assumption in Song's test is that the predictor variable X must be one-dimensional. Following a few of the significant works such as [7,28,14] in the classic regression models, we will try to develop a lack-of-fit test in this paper, based on an empirical L_2-distance 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 the linear, and the predictor can be multidimensional. To avoid the potential curse of dimensionality and to have a better performance, a larger sample size might be needed for the proposed test when the predictor is multidimensional. In addition, compared to other L_2-distance based tests, the proposed one is very efficient computationally.

The paper is organized as follows. The empirical L_2-distance statistic is proposed in Section 2, together with a list of technical assumptions required for the asymptotic results; the main results are presented in Section 3, including the asymptotic null distribution, the consistency and the local power of the test; simulation studies are conducted in Section 4, and proofs of the main results are deferred to Section 5.

Throughout this paper, we will use f_υ, F_υ to denote the density and the cumulative distribution function (CDF) of a random variable υ, $\Longrightarrow$ to denote the convergence in distribution, and $I(A)$ to denote the indicator function of the set A.

2. Empirical L_2-distance test statistics

Consider the classic regression model $Y = m(X) + \varepsilon$. The problem of interest is to test the following hypothesis

$$H_0 : m(x) = m(x, \theta) \quad \text{for some } \theta \in \Theta, \text{ versus } H_1 : H_0 \text{ is not true} \tag{2.1}$$

where $m(x, \theta)$ has a parametric form with parameter θ. An extensive introduction on the model specification hypothesis above can be found in [9] and the references therein. Koul and Ni [14] used the minimum distance method to construct the lack-of-fit tests for H_0. In a finite sample comparison of these tests with some other existing tests, they noted that a member of this class preserves the asymptotic level and has relatively very high power against some alternatives. The present paper nontrivially extends their method to the standard Tobit regression model.

To be specific, [14] considered the following tests of H_0 in (2.1) where the design is random and observable, and the errors are heteroscedastic. For any kernel density K, let $K_h(x) = K(x/h)/h^d$, $h > 0, x \in \mathbb{R}^d$. Define, $\hat{f}_w(x) = n^{-1}\sum_{j=1}^n K_w(x - X_j)$, $w = w_n \sim (\log n/n)^{1/(d+4)}$,

$$T_n(\theta) = \int_C \left[\frac{1}{n}\sum_{j=1}^n K_h(x - X_j)(Y_j - m(X_j, \theta))\right]^2 \frac{dG(x)}{\hat{f}_w^2(x)},$$

and $\hat{\theta}_n = \text{argmin}_{\theta\in\Theta} T_n(\theta)$, where $h = h_n$ and $w = w_n$ are the bandwidths, Θ is a compact subset of $\mathbb{R}^d$, and G is a σ-finite measure on the compact subset C of $\mathbb{R}^d$. They proved the consistency and the asymptotic normality of $\hat{\theta}_n$, and the asymptotic normality of $D_n = nh_n^{d/2}(T_n(\hat{\theta}_n) - \hat{C}_n)/\hat{\Gamma}_n^{1/2}$ under the null hypothesis, where

$$\hat{C}_n = \frac{1}{n^2}\sum_{i=1}^n \int_C K_h^2(x - X_i)e_i^2 \hat{f}_w^{-2}(x)dG(x), \qquad e_i = Y_i - m(X_i, \hat{\theta}_n)$$

and

$$\hat{\Gamma}_n = \frac{1}{n^2h^{3d}} \sum_{i\neq j=1}^n \left(\int_C K\left(\frac{x - X_i}{h}\right) K\left(\frac{x - X_j}{h}\right) e_i e_j \hat{f}_w^{-2}(x)dG(x)\right)^2.$$

The test based on D_n is preferable over the tests developed in [7,28]. Unlike in other related papers, [14] did not require the null regression function to be twice continuously differentiable in the parameter vector, nor do their proofs need the rate for uniform consistency of nonparametric regression function estimators. A consequence of the asymptotic normality result above is that at least for large samples one does not need to use any resampling method to implement these tests.

These findings thus motivate us to look for the lack-of-fit tests in the standard Tobit regression models using a similar method as in [14], but some modifications on the test above are needed because the response Y^* in the standard Tobit regression models are not always observable. Also, the implementation of the above procedure requires the calculation of the integrations in T_n, $\hat{C}_n$ and $\hat{\Gamma}_n$. These integrations usually do not have tractable forms, so some numerical integration techniques are needed to approximate the integration. In addition, the kernel estimator $\hat{f}_w(x)$ usually takes small values on the boundary of the domain of x, which makes the computation very unstable. We will propose an empirical L_2 test to address all the issues mentioned above.

To be specific, we shall consider the Tobit regression model $Y^* = m(X) + \varepsilon$, $Y = Y^*\{Y^* > 0\}$, where $E\varepsilon = 0$ and (X, Y) is observable. Since Y_i^*, hence (X_i, Y_i^*), $i = 1, 2, \ldots, n$ are not always observable, we have to construct the test statistic based on the observations (X_i, Y_i), $i = 1, 2, \ldots$. Therefore, certain relationships between Y and X should be found. A natural way of finding such a relationship is to consider the regression of Y against X. Let $Q_j(x) = \int_x^\infty u^j f_\varepsilon(u)du$, $j = 0, 1$. Then we can show that

$$E(Y|X = x) = m(x)Q_0(-m(x)) + Q_1(-m(x)). \tag{2.2}$$

Thus, one can consider the following regression model based on (2.2),

$$Y = m(X)Q_0(-m(X)) + Q_1(-m(X)) + \xi = g(X) + \xi. \tag{2.3}$$

Both ξ are uncorrelated with X.

If the density function f_ε of ε is totally unknown, then (2.2) is not applicable. Throughout this paper, we shall assume that the density function f_ε is known for the sake of simplicity, readability and model identifiability. A more realistic assumption should be that f_ε has a known parametric form with mean 0 and unknown parameter, say, β. In this case, Q_0 and Q_1 are also functions of β. Adding more regularity conditions to the model, it can be shown that the proposed test procedures in this paper are also applicable.

As a functional of $m(x)$, g is strictly monotone provided that F_ε is strictly increasing. This can be easily verified by checking the derivatives of $g(x)$, as functions of $m(x)$. Therefore, to test $H_0 : m(x) = m(x, \theta)$, it is equivalent to test

$$H_0 : g(x) = g(x, \theta) \quad \text{for some } \theta \in \Theta, \text{ versus } H_1 : H_0 \text{ is not true} \tag{2.4}$$

for regression model (2.3), where $g(x, \theta)$ is the same as $g(x)$ with $m(x)$ replaced by $m(x, \theta)$. If f_ε has a parametric form $f_\varepsilon(\cdot, \beta)$, then one can test the following hypothesis

$$H_0 : g(x) = g(x, \theta, \beta) \quad \text{for some } \theta, \beta, \text{ versus } H_1 : H_0 \text{ is not true}, \tag{2.5}$$

where

$$g(x, \theta, \beta) = m(X, \theta)Q_0(-m(X, \theta), \beta) + Q_1(-m(X, \theta), \beta)$$

with $Q_j(x, \beta) = \int_x^\infty u^j f_\varepsilon(u, \beta)du$, $j = 0, 1$.

Let K be a symmetric density function and h be a sequence of positive numbers depending on the sample size n. As before, denote $K_h(x) = h^{-d}K(x/h)$. The Nadaraya–Watson kernel estimator of the regression function g is defined by

$$\hat{g}(x) = \frac{\sum_{i=1}^n K_h(x - X_i)Y_i}{\sum_{i=1}^n K_h(x - X_i)}.$$

For any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ, the parametric estimator of $g(x)$ under the null hypothesis is $g(x, \hat{\theta}_n)$.

Let $W(x)$ be a weight function that may depend on the sample. Then the L_2-distances $D_n = \int \left[\hat{g}(x) - g(x, \hat{\theta}_n)\right]^2 dW(x)$, might be used for testing the hypotheses (2.4). D_n is similar to Haïrdle and Mammen [7]'s test statistic. As pointed out in [14], the nonparametric estimators $\hat{g}(x)$ has nonnegligible bias, the lack-fit-tests based on the quantity above may not have desirable asymptotic null distributions. Therefore, by mimicking Koul and Ni [14]'s procedure, we might use the following modification

$$\int \left[\frac{\sum_{j=1}^{n} K_h(x - X_j)(Y_j - g(X_j, \hat{\theta}_n))}{\sum_{j=1}^{n} K_h(x - X_j)} \right]^2 dW(x), \tag{2.6}$$

to test (2.4). To avoid the possible instability incurred by the small values of the kernel density estimator $\hat{f}_h$ of f_X in the denominator, and the complexity resulted from the potential intractable integration, we shall choose the weight function $W(x)$ such that $dW(x) = \hat{f}_h^2(x)dF_n(x)$, where F_n is the empirical CDF of the sample from X. Accordingly, (2.6) becomes

$$D_n = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{n} \sum_{j=1}^{n} K_h(X_i - X_j)(Y_j - g(X_j, \hat{\theta}_n)) \right]^2.$$

We shall show that the appropriate standardization for D_n is $T_n = nh^{d/2}\hat{\Gamma}_n^{-1/2}(D_n - \hat{C}_n)$, where

$$\hat{C}_n = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h^2(X_i - X_j)e_i^2,$$

$$\hat{\Gamma}_n = \frac{2h^d}{n^4} \sum_{i \neq j} \left[\sum_{k=1}^{n} K_h(X_k - X_i)K_h(X_k - X_j)e_i e_j \right]^2,$$

and $e_i = Y_i - g(X_i, \hat{\theta}_n)$. Larger values of $|T_n|$ will be the evidence to reject H_0 in (2.4). It is easy to see that the computational burden in calculating T_n is much less than those in [7,14], and it is comparable to that in [28]. The thresholds of the tests will be determined by the asymptotic distributions of T_n under the null hypotheses, which will be studied in the next section.

The following is a list of the needed assumptions to derive the asymptotic results of the test statistics.

(C1) The random error ε is such that $E(\varepsilon) = 0$, and $E(\varepsilon^4) < \infty$; ε and X are independent.
(C2) $\tau^2(x) = E[(Y - g(X))^2|X = x]$, $\sigma^4(x) = E[(Y - g(X))^4|X = x]$ 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 $f_X(x)$ of X and its first-order derivatives are uniformly bounded.
(C4) $g(x, \theta)$ is continuously differentiable with respect to θ, and the derivative $\dot{g}(x, \theta)$ satisfies $E\|\dot{g}(X; \theta)\|^4 < \infty$; for any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_0, the true value of θ under the null hypotheses, $\sup_{1\le i\le n} |g(X_i, \hat{\theta}_n) - g(X_i, \theta_0) - (\hat{\theta}_n - \theta_0)'\dot{g}(X_i, \theta_0)| = O_p(1/n)$.
(C5) The kernel density function $K(x)$ is continuous, bounded and symmetric around 0, $\int u^2K(u)du < \infty$, $\int \left[\int K(u)K(u + v)du\right]^4 dv < \infty$.
(C6) $h \to 0$, $nh^{2d} \to \infty$ as $n \to \infty$.

Conditions (C2) and (C3) are the same as the Assumption 1 in [28], and are very typical in nonparametric smoothing literature. (C4) appears simpler but indeed more restrictive than the conditions (m4) in [14]. Certainly we can introduce some more complex conditions to replace (C4) and adjust the proofs accordingly, but we decide not to do so for the sake of neatness of presentation. Roughly speaking, the role played by (C4) in this paper is the same as the conditions (m4) and (m5) in [14]. If f_ε has a parametric form $f_\varepsilon(\cdot, \beta)$, then to develop a test for (2.5), one should replace θ with (θ, β) in (C4).

3. Main results

Without further emphasis, we shall assume that there always exists a $\sqrt{n}$-consistent estimator for the parameter θ in the regression function under the null hypothesis. The theorem below states the asymptotic null distribution of the test statistics T_n.

Theorem 3.1. *Suppose* (C1)–(C6) *hold. Then under the null hypotheses* H_0 *in* (2.4), $T_n = nh^{d/2}\hat{\Gamma}_n^{-1/2}(D_n - \hat{C}_n) \Longrightarrow N(0, 1)$.

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

A basic and reasonable requirement for any tests is the consistency. That is, for fixed alternatives, a test should have power approaching to 1 as the sample size goes to ∞.

Consider a class of fixed alternative hypotheses:

$$H_a : E(Y^*|X = x) = m(x) \tag{3.1}$$

such that $Em^2(X) < \infty$ and $m(x) \not\equiv m(x, \theta)$ for any θ.

Under the null hypothesis, we have assumed that estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent for the true parameter θ_0. Would this estimator still have the similar property under the alternative hypothesis H_a? The question is of interest in its own right. In the classic regression setup, [13,26,27] 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. In the following, we simply assume that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$ under the alternative H_a for some $\theta_a \in \mathbb{R}^p$. We will not justify this assumption rigorously here.

Theorem 3.2. *Suppose all the conditions in Theorem 3.1 hold with θ_0 replaced by θ_a, $\int [g(x) - g(x; \theta_a)]^2 f_X^3(x)dx > 0$. Then for any $0 < \alpha < 1$, the test that rejects H_0 in (2.4) whenever $|T_n| > z_{1-\alpha/2}$ is consistent against the alternatives H_a in (3.1).*

Sometimes it is desirable to investigate the performance of a test statistic at local alternatives. For this purpose, let $\delta(x)$ be a continuous function such that $E\delta^2(X) < \infty$. Consider the following sequence of local alternatives

$$H_{\text{Loc}} : m(x) = m(x, \theta_0) + \delta(x)/\sqrt{nh^{d/2}}. \tag{3.2}$$

We continue to assume that the estimators $\hat{\theta}_n$ used in the test statistic are satisfying $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. Define $\mu = E\delta^2(X)Q_0^2(-m(X, \theta_0))f_X^2(X)$, then we have

Theorem 3.3. *Suppose all the conditions in Theorem 3.1 hold, Then under H_{Loc} in (3.2), $T_n \Longrightarrow N(\mu, 1)$.*

4. Simulation studies

Two Monte Carlo simulations are conducted in this section to assess the finite sample performance of the proposed test. We choose linear regression functions ($d = 1$ and $d = 2$) in 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 for all simulations. For each scenario and sample sizes $n = 100, 200, 300, 500, 800$ and 1000, we repeat the tests 1000 times, the empirical level and power are calculated by $\#\{|T_n| \geq 1.96\}/1000$. The `vglm` function in the `R` package `VGAM` is used to calculate the estimates of all unknown parameters.

Simulation 1: The data are generated from the models

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

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 choose 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 at 0 when $\gamma = 0$.

The simulation results are presented in the left part 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 with sample sizes getting larger.

In general, the bootstrap provides a more accurate approximation to the distribution of the test statistic than the asymptotic normal distribution. Under the null hypothesis, the test statistic T_n has an asymptotic standard normal distribution. Therefore, T_n is asymptotically pivotal, which enables us to conduct a parametric bootstrap. To find the parametric bootstrap critical values, for each sample size, we repeat the simulation under the null hypothesis 1000 times, the critical values are then obtained by finding out the upper 97.5th% and lower 2.5th% of these 1000 test statistics. Using the bootstrap critical values, we conduct the simulation again, and the empirical levels and powers are taken as the relative frequencies of how many times the test statistics being lower than the 2.5th% and bigger than the 97.5th%. The right part of Table 1 reports the simulation results. The same random seeds are used to obtain the bootstrap critical values and conduct the simulations, therefore, all the empirical levels are exactly 0.05 for all cases. It is easily seen that the powers are much larger than the ones reported in the left part of Table 1.

Simulation 2: To see the performance of the proposed test when $d > 1$, we generate the data from the models

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

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 regression 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

Table 1
$d = 1$. Empirical powers based on critical values from Theorem 3.1 and bootstrap.

γ	100	300	500	800	1000	100	300	500	800	1000
0	0.002	0.001	0.002	0.005	0.003	0.050	0.050	0.050	0.050	0.050
0.1	0.009	0.041	0.102	0.196	0.280	0.088	0.216	0.253	0.399	0.501
0.2	0.098	0.420	0.775	0.961	0.991	0.333	0.760	0.895	0.989	0.997
0.3	0.346	0.915	0.998	1.000	1.000	0.624	0.991	1.000	1.000	1.000
0.5	0.882	1.000	1.000	1.000	1.000	0.979	1.000	1.000	1.000	1.000

Table 2
$d = 2$. Empirical powers based on critical values from Theorem 3.1 and bootstrap.

γ	100	300	500	800	1000	100	300	500	800	1000
0	0.002	0.004	0.006	0.005	0.005	0.050	0.050	0.050	0.050	0.050
0.1	0.022	0.185	0.358	0.621	0.792	0.190	0.382	0.619	0.776	0.872
0.2	0.289	0.896	0.995	1.000	1.000	0.608	0.968	0.999	1.000	1.000
0.3	0.750	0.999	1.000	1.000	1.000	0.936	1.000	1.000	1.000	1.000
0.5	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3
Empirical powers for the simple hypotheses.

γ	Model (4.1): $d = 1$					Model (4.2): $d = 2$				
	100	300	500	800	1000	100	300	500	800	1000
0	0.036	0.044	0.062	0.056	0.046	0.046	0.043	0.053	0.038	0.041
0.1	0.054	0.075	0.126	0.193	0.195	0.086	0.189	0.292	0.435	0.533
0.2	0.114	0.280	0.470	0.738	0.852	0.332	0.761	0.938	0.998	1.000
0.3	0.244	0.673	0.921	0.996	0.999	0.677	0.991	1.000	1.000	1.000
0.5	0.689	0.997	0.999	1.000	1.000	0.990	1.000	1.000	1.000	1.000

$\gamma = 0$ 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 $P(\varepsilon \le -1 - X_1 - X_2) \approx 28\%$ observations of Y^* are truncated below at 0 when $\gamma = 0$.

The simulation results are presented in the left part of Table 2. The test again appears conservative, the power increases with increasing sample sizes. We also did some simulation studies when X_1 and X_2 are weakly and moderately correlated. The results are not reported here because of their similarity to the left part of Table 2. Similar to the one-dimensional case, we also conduct a parametric bootstrap simulation in which the same random seeds are used to obtain the bootstrap critical values and conduct the simulations. The results are shown in the right part of Table 2. 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 part of Table 2.

Remark 4.1. For comparison purposes, a simulation study for simple null hypotheses is also conducted. Using the same setups as in Simulations 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 3. Similar patterns as in the previous tables are also appeared in Table 3, but the empirical level is much closer to the nominal level 0.05 in all cases.

5. Proofs of the main results

Let $\xi_i = Y_i - g(X_i, \theta_0)$ and

$$\tilde{D}_n = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)\xi_j\right]^2, \qquad \tilde{C}_n = \frac{1}{n^3}\sum_{i,j=1}^n K_h^2(X_i - X_j)\xi_j^2,$$

$$\tilde{\Gamma}_n = \frac{2h^d}{n^4}\sum_{i\neq j}\left[\sum_{k=1}^n K_h(X_k - X_i)K_h(X_k - X_j)\xi_i\xi_j\right]^2,$$

$$\Gamma = 2\int\left[\int K(u)K(u+v)\right]^2 dv \cdot \int \tau^4(x)f^4(x)dx,$$

where $\tau^2(x)$ is defined in (C2).

The proof of Theorem 3.1 is facilitated by a series of lemmas stated below. Lemma 5.2 can be proved using Theorem 1 of [8] which is reproduced here for the sake of completeness.

Lemma 5.1. *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

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

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 $n^2EH_n^2(Z_1, Z_2)/2$.

Lemma 5.2. *Suppose the conditions* (C1)–(C3), (C5) *and* (C6) *hold. Then under the null hypotheses H_0 in* (2.4), $nh^{d/2}(\tilde{D}_n - \tilde{C}_n) \Longrightarrow N(0, \Gamma)$.

Proof. Expanding the square term in $\tilde{D}_n$, $\tilde{D}_n - \tilde{C}_n$ can be written as the sum of the following two terms:

$$A_{n1} = \frac{1}{n^2}\sum_{j\ne k}\left[\frac{1}{n}\sum_{i\ne j,k} K_h(X_i - X_j)K_h(X_i - X_k)\right]\xi_j\xi_k$$

$$A_{n2} = \frac{2K_h(0)}{n^3}\sum_{j\ne k} K_h(X_j - X_k)\xi_j\xi_k.$$

Note that $EA_{n2} = 0$, and

$$\mathrm{Var}(A_{n2}) = \frac{8K^2(0)n(n-1)}{n^6h^{4d}}EK^2\left(\frac{X_1 - X_2}{h}\right)\tau^2(X_1)\tau^2(X_2).$$

By (C2) and (C5), one can show that $\mathrm{Var}(A_{n2}) = O(1/n^4h^{3d})$. Therefore $nh^{d/2}A_{n2} = o_p(1)$.

Let $H(X_j, X_k, h) = E[K_h(X_i - X_j)K_h(X_i - X_k)|X_j, X_k]$ which is a symmetric function of X_j, X_k. Then A_{n1} can be written as the sum of the following two terms:

$$A_{n11} = \frac{1}{n^3}\sum_{j\ne k}\left[\sum_{i\ne j,k} K_h(X_i - X_j)K_h(X_i - X_k) - h^{-d}H(X_j, X_k, h)\right]\xi_j\xi_k,$$

$$A_{n12} = \frac{1}{n^2h^d}\sum_{j\ne k} H(X_j, X_k, h)\xi_j\xi_k.$$

Let

$$G(X_j, X_k, h) = n^{-1}\sum_{i\ne j,k} K_h(X_i - X_j)K_h(X_i - X_k) - h^{-d}H(X_j, X_k, h).$$

Note that $EA_{n11} = 0$,

$$\begin{aligned}
EA_{n11}^2 &= E\left[\frac{1}{n^2}\sum_{j\ne k} G(X_j, X_k, h)\xi_j\xi_k\right]^2 \\
&\le 2E\left[\frac{1}{n^2}\sum_{j<k} G(X_j, X_k, h)\xi_j\xi_k\right]^2 + 2E\left[\frac{1}{n^2}\sum_{j>k} G(X_j, X_k, h)\xi_j\xi_k\right]^2 \\
&= \frac{2(n-1)}{n^3}EG^2(X_1, X_2, h)\tau^2(X_1)\tau^2(X_2).
\end{aligned}$$

While $EG^2(X_1, X_2, h)\tau^2(X_1)\tau^2(X_2)$ equals

$$\begin{aligned}
&E\left[\frac{1}{n}\sum_{i=3}^{n} K_h(X_i - X_1)K_h(X_i - X_2) - h^{-d}H(X_1, X_2, h)\right]^2\tau^2(X_1)\tau^2(X_2) \\
&\le 2\left(\frac{n-2}{n}\right)^2 E\left[\frac{1}{n-2}\sum_{i=3}^{n}[K_h(X_i - X_1)K_h(X_i - X_2) - h^{-d}H(X_1, X_2, h)]\right]^2\tau^2(X_1)\tau^2(X_2) \\
&\quad + \frac{8}{n^2h^{2d}}EH^2(X_1, X_2, h)\tau^2(X_1)\tau^2(X_2).
\end{aligned} \tag{5.1}$$

Conditioning on (X_1, X_2), and by (C2), (C3),

$$
\begin{aligned}
&E\left[\frac{1}{n-2}\sum_{i=3}^{n}[K_h(X_i-X_1)K_h(X_i-X_2)-h^{-d}H(X_1,X_2,h)]\right]^2\tau^2(X_1)\tau^2(X_2)\\
&\le \frac{1}{n-2}EK_h^2(X_3-X_1)K_h^2(X_3-X_2)\tau^2(X_1)\tau^2(X_2)\\
&= \frac{1}{h^{2d}}\iiint K^2(u)K^2(v)\tau^2(x_3-uh)\tau^2(x_3-vh)f(x_3-uh)f(x_3-vh)f(x_3)dudvdx_3\\
&= O(1/((n-2)h^{2d})).
\end{aligned}
$$

Therefore, the first term in (5.1) has the order of $O_p(1/(nh^{2d}))$. Similarly, by (C2) and (C3), we have

$$
\begin{aligned}
&EH^2(X_1,X_2,h)\tau^2(X_1)\tau^2(X_2)\\
&= \iint\left(\int K(y)K(y+(x_1-x_2)/h)f(x_1+hy)dy\right)^2\tau^2(x_1)\tau^2(x_2)f(x_1)f(x_2)dx_1dx_2\\
&= h^d\iint\left(\int K(y)K(y+u)f(x_1+hy)dy\right)^2\tau^2(x_1)\tau^2(x_1-uh)f(x_1)f(x_1-uh)dx_1du\\
&= O(h^d).
\end{aligned}
$$

Therefore, the second term in (5.1) is the order of $O_p(n^{-2}h^{-d})$. Hence

$$
EA_{n1}^2 = O_p\left(\frac{1}{n^3h^{2d}}\right) + O\left(\frac{1}{n^4h^d}\right)
$$

which implies $nh^{d/2}A_{n1} = o_p(1)$, and eventually

$$
nh^{d/2}(\tilde{D}_n-\tilde{C}_n) = nh^{d/2}A_{n12} + o_p(1) = \frac{1}{nh^{d/2}}\sum_{j\neq k}H(X_j,X_k,h)\xi_j\xi_k + o_p(1).
$$

Denote $Z_j = (X_j', \xi_j)'$, and $H_n(Z_j,Z_k) = n^{-1}h^{-d/2}H(X_j,X_k,h)\xi_j\xi_k$, we have

$$
nh^{d/2}(\tilde{D}_n-\tilde{C}_n) = 2\sum_{1\le j<k\le n}H_n(Z_j,Z_k) + o_p(1).
$$

Note that $H_n(x,y)$ is symmetric, and $E[H_n(Z_1,Z_2)|Z_1] = 0$. Also, for each n, by (C2) and (C3),

$$
\begin{aligned}
EH_n^2(Z_j,Z_k) &= \frac{1}{n^2h^d}\iint H^2(x,y,h)\tau^2(x)\tau^2(y)f(x)f(y)dxdy\\
&= \frac{1}{n^2h^d}\iint\left[\int K(u)K(u+(x-y)/h)f(x+hu)du\right]^2\tau^2(x)\tau^2(y)f(x)f(y)dxdy\\
&= \frac{1}{n^2}\iint\left[\int K(u)K(u+v)f(x+hu)du\right]^2\tau^2(x)\tau^2(x-vh)f(x)f(x-vh)dxdv\\
&= \frac{1}{n^2}\int\left[\int K(u)K(u+v)du\right]^2dv\cdot\int\tau^4(x)f^4(x)dx + o\left(\frac{1}{n^2}\right) < \infty.
\end{aligned}
\tag{5.2}
$$

Hence, in view of Lemma 5.1, it suffices to verify that

$$
\frac{EM_n^2(Z_1,Z_2)}{[EH_n^2(Z_1,Z_2)]^2}\to 0, \qquad \frac{H_n^4(Z_1,Z_2)}{n[EH_n^2(Z_1,Z_2)]^2}\to 0.
\tag{5.3}
$$

For this purpose, write $t\in\mathbb{R}^{d+1}$ as $t' = (t_1', t_2)$ with $t_1\in\mathbb{R}^d$. Then for $t, s\in\mathbb{R}^{d+1}$,

$$
M_n(t,s) = EH_n(Z,t)H_n(Z,s) = \frac{1}{n^2h^d}EH(X,t_1,h)H(X,s_1,h)\xi^2t_2s_2.
$$

Note that $EH(X,t_1,h)H(X,s_1,h)\xi^2 = EH(X,t_1,h)H(X,s_1,h)\tau^2(X)$, and it further equals

$$
\iiint K(x)K(y)K(x+(u-t_1)/h)K(y+(u-s_1)/h)f(u+hx)f(u+hy)\tau^2(u)f(u)dxdydu.
$$

Changing variable, $(u-t_1)/h = v$, the above integration can be written as $h^d B_h(t_1, s_1)$ with $B_h(t_1, s_1)$ equals

$$\iiint K(x)K(y)K(x+v)K(y+v+(t_1-s_1)/h)f(t_1+vh+xh)f(t_1+vh+hy)\cdot f(t_1+vh)\tau^2(t_1+vh)dxdydv.$$

Therefore,

$$\begin{aligned} EM_n^2(Z_1, Z_2) &= \frac{1}{n^4} EB_h^2(X_1, X_2)\xi_1^2\xi_2^2 = \frac{1}{n^4} EB_h^2(X_1, X_2)\tau^2(X_1)\tau^2(X_2) \\ &= \frac{1}{n^2}\int B_h^2(t_1, s_1)\tau^2(t_1)\tau^2(s_1)f(t_1)f(s_1)dt_1 ds_1 \\ &= \frac{h^d}{n^2}\int B_h^2(t_1, t_1+wh)\tau^2(t_1)\tau^2(t_1+wh)f(t_1)f(t_1+wh)dt_1 dw. \end{aligned}$$

By (C2) and (C3), and the fact that

$$B_h(t_1, t_1+wh) = \iiint K(x)K(y)K(x+v)K(y+v-w)f^3(t_1)\tau^2(t_1)dxdydv + o(1),$$

we obtain $EM_n^2(Z_1, Z_2) = O(h^d/n^4)$, which implies

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

Similarly, one obtains

$$\begin{aligned} EH_n^4(Z_j, Z_k) &= \frac{1}{n^4h^{2d}}\iint H^4(x, y, h)\sigma^4(x)\sigma^4(y)f(x)f(y)dxdy \\ &= \frac{1}{n^4h^{2d}}\iint \left[\int K(u)K(u+(x-y)/h)f(x+hu)du\right]^4 \sigma^4(x)\sigma^4(y)f(x)f(y)dxdy \\ &= \frac{1}{n^4h^{d}}\iint \left[\int K(u)K(u+v)f(x+hu)du\right]^4 \sigma^4(x)\sigma^4(x-vh)f(x)f(x-vh)dxdv \\ &= \frac{1}{n^4h^{d}}\int \left[\int K(u)K(u+v)du\right]^4 dv\cdot\int \sigma^8(x)f^6(x)dx + o\left(\frac{1}{n^4h^d}\right). \end{aligned}$$

Therefore,

$$\frac{H_n^4(Z_1, Z_2)}{n[EH_n^2(Z_1, Z_2)]^2} = \frac{O(1/(n^4h^d))}{n\cdot O(1/n^4)} = O\left(\frac{1}{nh^d}\right) \to 0.$$

This completes the proof of (5.3). By (5.2), we have

$$\frac{1}{2}n^2 EH_n(Z_1, Z_2) = \frac{1}{2}\int\left[\int K(u)K(u+v)du\right]^2 dv\cdot\int \tau^4(x)f^4(x)dx + o(1).$$

The theorem is then proved by using Lemma 5.1. □

Lemma 5.3. *Suppose* (C1)–(C3), (C5) *and* (C6) *hold. A function $L(x)$ is continuously differentiable, $EL^2(X) < \infty$, and its derivative is bounded above by a measurable function $b(x)$ such that $Eb^2(X) < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)L(X_j)\right]^2 = O_p(1), \tag{5.4}$$

$$\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\right]^2 = O_p(1/nh^d), \tag{5.5}$$

$$\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j L(X_i)\right] = O_p(1/\sqrt{n}). \tag{5.6}$$

Proof. Note that

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)L(X_j)\right]^2\right)=E\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_1-X_j)L(X_j)\right]^2,$$

the right hand side is bounded above by

$$\frac{2K_h^2(0)}{n^2}EL^2(X_1)+\frac{2(n-1)^2}{n^2}E\left[\frac{1}{n-1}\sum_{j=2}^{n}K_h(X_1-X_j)L(X_j)\right]^2.$$

The first term is the order of $O(1/(nh^d)^2)=o_p(1)$ by (C5). For any $j\neq 1$, denote

$$J(X_1,h)=E[K_h(X_1-X_j)L(X_j)|X_1]=\int K(u)L(X_1-uh)f(X_1-uh)du.$$

Then we have $E\left[(n-1)^{-1}\sum_{j=2}^{n}K_h(X_1-X_j)L(X_j)\right]^2$ to be bounded above by

$$2E\left[\frac{1}{n-1}\sum_{j=2}^{n}K_h(X_1-X_j)L(X_j)-J(X_1,h)\right]^2+2EJ^2(X_1,h).$$

The continuity of $L(x)$ and (C3) imply $EJ^2(X_1,h)=O(1)$. While the first term above is further bounded above by

$$\frac{2}{n-1}EK_h^2(X_1-X_2)L^2(X_2)=\frac{2}{(n-1)h^d}\iint K^2(u)L^2(x_2)f(x_2+uh)f(x_2)dudx_2=o_p(1).$$

This proves (5.4).

A similar argument implies that the left-hand side of (5.5) is bounded above by

$$\frac{2K_h^2(0)}{n^2}E\xi_1^2+\frac{2(n-1)^2}{n^2}\cdot\frac{1}{n-1}EK_h^2(X_1-X_2)\tau^2(X_2).$$

Thus, (5.5) can be obtained by (C2) and (C3).

To show (5.6), note that the left-hand side of (5.6) can be written as

$$\frac{K(0)}{n^2h^d}\sum_{i=1}^{n}\xi_iL(X_i)+\frac{1}{n^2}\sum_{i\neq j}K_h(X_i-X_j)\xi_jL(X_j).$$

A simple expectation-variance argument, together with the finiteness of $EL^2(X)$, implies the first term above is $O_p(1/n\sqrt{n}h^d)$. It is also easy to see that the expectation of the second term is 0, and the second moment can be written as

$$\frac{n(n-1)}{2n^4}EK_h(X_1-X_2)\tau^2(X_2)L^2(X_1)+\frac{n(n-1)(n-2)}{n^4}EK_h(X_1-X_2)K_h(X_3-X_2)\tau^2(X_2)L(X_1)L(X_2).$$

(C2), (C3) and the continuity of $L(x)$ imply that the first term is the order of $O(1/n^2h^d)$, and the second term is $O(1/n)$. In summary, the left-hand side of (5.6) has the order of

$$O_p\left(\frac{1}{n\sqrt{n}h^d}\right)+O_p\left(\frac{1}{nh^{d/2}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right)$$

which is $O_p(1/\sqrt{n})$ based on our assumption $nh^d\to\infty$. □

Lemma 5.4. *Suppose* (C1)–(C6) *hold. Under the null hypotheses H_0 in* (2.4), *$nh^{d/2}(D_n-\tilde{D}_n)\to 0$ in probability.*

Proof. Subtracting and adding $g(X_j,\theta_0)$ from $Y_j-g(X_j,\hat{\theta}_n)$, D_n can be written as the sum $\tilde{D}_n+B_{n1}-2B_{n2}$, where

$$B_{n1}=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)[g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)]\right]^2,$$

$$B_{n2}=\frac{1}{n}\sum_{i=1}^{n}\left(\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)[g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)]\right]\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)\xi_j\right]\right).$$

Denote

$$\delta_{jn}=g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)-(\hat{\theta}_n-\theta_0)'\dot{g}(X_j,\theta_0),\tag{5.7}$$

then one can show that B_{n1} is bounded above by

$$\sup_{1\le i\le n} \delta_{in}^2 \cdot \frac{2}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\right]^2 + \frac{2\|\hat\theta_n-\theta_0\|^2}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\|\dot g(X_j,\theta_0)\|\right]^2.$$

From (C4), the $\sqrt{n}$-consistency of $\hat\theta_n$, and Lemma 5.3, we obtain $B_{n1}=O_p(1/n)$, which implies $nh^{d/2}B_{n1}=o_p(1)$.

Now, let us consider B_{n2}. Adding and subtracting $(\hat\theta_n-\theta_0)'\dot g(X_j,\theta_0)$ from $g(X_j,\hat\theta_n)-g(X_j,\theta_0)$, B_{n2} can be written as the sum of the following two terms:

$$B_{n21}=\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\delta_{in}\right]\cdot\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\right],$$

$$B_{n22}=\frac{(\hat\theta_n-\theta_0)'}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\dot g(X_j,\theta_0)\right]\cdot\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\right].$$

By the Cauchy–Schwarz inequality, B_{n21} is bounded above by

$$\sup_{1\le i\le n}|\delta_{in}|\cdot\left(\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\right]^2\right)^{1/2}\cdot\left(\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\right]\right)^{1/2}.$$

The first factor is $o_p(1/\sqrt{n})$ according to assumption (C4). From Lemma 5.3, the second term is $O_p(1)$, and the third is $O_p(1/\sqrt{nh^d})$. Thus, $nh^{d/2}B_{n21}=o_p(1)$.

Without loss of generality, let us assume $d=1$. For $d>1$, one can argue elementwise. Note that

$$\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\dot g(X_j,\theta_0)=\frac{1}{n}K_h(0)\dot g(X_i,\theta_0)+\frac{1}{n}\sum_{j\ne i}^n K_h(X_i-X_j)\dot g(X_j,\theta_0),$$

so B_{n22} can be written as the sum of the following two terms:

$$B'_{n22}=\frac{(\hat\theta_n-\theta_0)K_h(0)}{n}\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\dot g(X_i,\theta_0)\right].$$

$$B''_{n22}=\frac{(\hat\theta_n-\theta_0)}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j\ne i}^n K_h(X_i-X_j)\dot g(X_j,\theta_0)\right]\cdot\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\right].$$

By (5.6), $nh^{d/2}B'_{n22}=O_p(1/n\sqrt{n}h^{d/2})=o_p(1)$. Define $P_h(x)=EK_h(x-X)\dot g(X,\theta_0)$. B''_{n22} can be written as the sum of $O_p\left(1/\sqrt{n}\right)[R_{n1}+R_{n2}]$ with

$$R_{n1}=\frac{1}{n}\sum_{i=1}^n\left(\left[\frac{1}{n-1}\sum_{j\ne i}^n K_h(X_i-X_j)\dot g(X_j,\theta_0)-P_h(X_j)\right]\cdot\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_j\right]\right)$$

and

$$R_{n2}=\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)\xi_jP_h(X_i)\right].$$

Similar to the argument in the proof of (5.5), one can show that

$$\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n-1}\sum_{j\ne i}^n K_h(X_i-X_j)\dot g(X_j,\theta_0)-Q_h(X_j)\right]^2=O_p\left(\frac{1}{nh^d}\right).$$

Therefore, by the Cauchy–Schwarz inequality and (5.5), one has $R_{n1}=O_p(1/(nh^d))$, $R_{n2}=O_p(1/\sqrt{n})$. Hence $nh^{d/2}B_{n22''}=o_p(1)$ and $nh^{d/2}B_{n22}=o_p(1)$. The lemma is proved. □

Lemma 5.5. *Suppose all the conditions in Lemma 5.4 hold. Then under the null hypotheses H_0 in (2.4), $nh^{d/2}(\hat C_n-\tilde C_n)\to 0$ in probability.*

Proof. First we claim that, for any nonnegative continuous function $L(x)$ such that $EL(X) < \infty$,

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)L(X_i) = O_p\left(\frac{1}{nh^d}\right). \tag{5.8}$$

In fact, the claim follows from the expectation of the left-hand side, which equals

$$\frac{1}{n^2}K_h^2(0)EL(X) + \frac{n(n-1)}{n^3}EK_h^2(X_1 - X_2)L(X_1) = O\left(\frac{1}{n^2h^{2d}}\right) + O\left(\frac{1}{nh^d}\right),$$

and the assumption that $nh^d \to \infty$.

Adding and subtracting $g(X_j; \theta_0)$ from $Y_j - g(X_j, \hat{\theta}_n)$, $\hat{C}_n - \tilde{C}_n$ can be written as the sum of the following two terms:

$$C_{n1} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)[g(X_j; \hat{\theta}_n) - g(X_j; \theta_0)]^2,$$

$$C_{n2} = -\frac{2}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)\xi_j[g(X_j; \hat{\theta}_n) - g(X_j; \theta_0)].$$

Recall the notation δ_{jn} in (5.7), one can show that C_{n1} is bounded above by

$$\sup_{1\le j\le n}|\delta_{jn}|^2 \cdot \frac{2}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j) + \frac{2\|\hat{\theta}_n - \theta_0\|^2}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)\|\dot{g}(X_j, \theta_0)\|^2.$$

Applying (5.8) with $L(x) = 1$ and $\|\dot{g}(x; \theta_0)\|^2$, together with the condition (C4) on δ_{in} and the $\sqrt{n}$-consistency of $\hat{\theta}_n$, one can show that this upper bound is the order of $O_p(1/n^2h^d)$. Hence, $nh^{d/2}C_{n1} = o_p(1)$.

Similarly, one can rewrite $-C_{n2}$ as the following sum

$$\frac{2}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)\xi_j\delta_{jn} + \frac{2(\hat{\theta}_n - \theta_0)'}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)\xi_j\dot{g}(X_j; \theta_0).$$

By the Cauchy–Schwarz inequality, the first term is bounded above by

$$\sup_{1\le j\le n}|\delta_{jn}| \cdot \frac{1}{n} \cdot \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}K_h^2(X_i - X_j)\right)^2} \cdot \sqrt{\frac{1}{n}\sum_{j=1}^{n}\xi_j^2} = O_p(1/n^2).$$

Similar to the proof of (5.4), one can show that

$$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{i=1}^{n}K_h^2(X_i - X_j)\right)^2 = O_p(h^{-2d}).$$

Therefore, by the assumption on δ_{jn}, the finiteness of $E\xi^2$, the first term is the order of $O_p(1/n^2h^d)$. Finally, similar to the proof of (5.6), one can show that

$$\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h^2(X_i - X_j)\xi_j\dot{g}(X_j; \theta_0) = O_p\left(\frac{1}{n^3h^{2d}}\right) + O_p\left(\frac{1}{n^{3/2}}\right).$$

Therefore, by the $\sqrt{n}$-consistency of $\hat{\theta}_n$, we have $nh^{d/2}C_{n2} = o_p(1)$. This completes the proof of the lemma. □

Lemma 5.6. *Suppose all the conditions in Lemma 5.4 hold. Then under the null hypotheses H_0 in (2.4), $\hat{\Gamma}_n - \tilde{\Gamma}_n \to 0$, $\tilde{\Gamma}_n \to \Gamma$ in probability.*

Proof. Let $t_i = g(X_i; \hat{\theta}_n) - g(X_i; \theta_0)$. Then $\hat{\Gamma}_n - \tilde{\Gamma}_n$ can be written as the sum of the following two terms:

$$\Gamma_{n1} = \frac{2h^d}{n^4}\sum_{i\neq j}\left[\sum_{k=1}^{n}K_h(X_k - X_i)K_h(X_k - X_j)[\xi_i t_j + \xi_j t_i - t_i t_j]\right]^2,$$

$$\Gamma_{n2} = \frac{2h^d}{n^4}\sum_{i\neq j}\left(\left[\sum_{k=1}^{n}K_h(X_k - X_i)K_h(X_k - X_j)\xi_i\xi_j\right]\left[\sum_{k=1}^{n}K_h(X_k - X_i)K_h(X_k - X_j)[\xi_i t_j + \xi_j t_i - t_i t_j]\right]\right).$$

At this point, we would state the following claims:

$$\frac{h^d}{n^4}\sum_{i\neq j}\left[\sum_{k=1}^{n}K_h(X_k-X_i)K_h(X_k-X_j)|\xi_i\xi_j|\right]^2=O_p(1), \tag{5.9}$$

$$\frac{h^d}{n^4}\sum_{i\neq j}\left[\sum_{k=1}^{n}K_h(X_k-X_i)K_h(X_k-X_j)|\xi_i|\right]^2=O_p(1), \tag{5.10}$$

$$\frac{h^d}{n^4}\sum_{i\neq j}\left[\sum_{k=1}^{n}K_h(X_k-X_i)K_h(X_k-X_j)\right]^2=O_p(1). \tag{5.11}$$

For the sake of brevity, we only present the proof for (5.9)–(5.11) can be similarly argued. By taking expectation,

$$\begin{aligned}
&E\sum_{i\neq j}\left[\sum_{k=1}^{n}K_h(X_k-X_i)K_h(X_k-X_j)|\xi_i|\right]^2\\
&\quad=n(n-1)E\left[2K_h(0)K_h(X_1-X_2)|\xi_1|+\sum_{k=3}^{n}K_h(X_k-X_1)K_h(X_k-X_2)|\xi_1|\right]^2\\
&\quad\leq 8n(n-1)K_h^2(0)EK_h^2(X_1-X_2)\tau^2(X_1)+2n(n-1)E\left[\sum_{k=3}^{n}K_h(X_k-X_1)K_h(X_k-X_2)|\xi_1|\right]^2.
\end{aligned}$$

By (C2) and (C3), the first term on the right-hand side is $O(n^2/h^{3d})$. While the second term equals

$$\begin{aligned}
&2n(n-1)E\left[\sum_{k=3}^{n}K_h^2(X_k-X_1)K_h^2(X_k-X_2)|\xi_1|^2\right]\\
&\quad+2n(n-1)E\left[\sum_{j\neq k}^{n}K_h(X_j-X_1)K_h(X_j-X_2)K_h(X_k-X_1)K_h(X_k-X_2)|\xi_1|^2\right]\\
&=2n(n-1)(n-2)EK_h^2(X_3-X_1)K_h^2(X_3-X_2)\tau^2(X_1)\\
&\quad+2n(n-1)(n-2)(n-3)E\left[K_h(X_3-X_1)K_h(X_3-X_2)K_h(X_4-X_1)K_h(X_4-X_2)\tau^2(X_1)\right]\\
&=O(n^3/h^{2d})+O(n^4/h^d).
\end{aligned}$$

Therefore, the left-hand side of (5.10) has the order of

$$\frac{h^d}{n^4}\left[O_p\left(\frac{n^2}{h^{3d}}\right)+O_p\left(\frac{n^3}{h^{2d}}\right)+O_p\left(\frac{n^4}{h^d}\right)\right]=O_p(1)$$

which is the desired result. Note that $E\|\dot g(X;\theta_0)\|^2<\infty$ which implies

$$\|\hat\theta_n-\theta_0\|\cdot\max_{1\leq i\leq n}\|\dot g(X_i;\theta_0)\|=o_p(1).$$

Combining with the fact $\sup_{1\leq i\leq n}\|\delta_{in}\|=o_p(1)$, we have $\sup_{1\leq i\leq n}|t_i|=o_p(1)$. Therefore, $\Gamma_{n1}=o_p(1)$ by (5.10) and (5.11), and the fact that t_i, $1\leq i\leq n$ are free from x. It further implies that $\Gamma_{n2}=o_p(1)$ by (5.9) and applying the Cauchy–Schwarz inequality to the double sum. Hence $\hat\Gamma_n-\tilde\Gamma_n=o_p(1)$. To show $\tilde\Gamma_n\to\Gamma$ in probability, first note that

$$\tilde\Gamma_n=\frac{2h^d}{n^4}\sum_{i\neq j}\left[2K_h(0)K_h(X_i-X_j)\xi_i\xi_j+\sum_{k\neq i,j}K_h(X_k-X_i)K_h(X_k-X_j)\xi_i\xi_j\right]^2.$$

Expanding the quadratic term, $\tilde\Gamma_n$ can be written as the sum of the following three terms:

$$\tilde\Gamma_{n1}=\frac{8h^d}{n^4}\sum_{i\neq j}K_h^2(X_i-X_j)\xi_i^2\xi_j^2,$$

$$\tilde\Gamma_{n2}=\frac{2h^d}{n^4}\sum_{i\neq j}\left[\sum_{k\neq i,j}K_h(X_k-X_i)K_h(X_k-X_j)\xi_i\xi_j\right]^2,$$

$$\tilde{\Gamma}_{n3} = \frac{8h^d}{n^4} \sum_{i \neq j} \left[(K_h(X_i - X_j)\xi_i\xi_j) \sum_{k \neq i,j} K_h(X_k - X_i)K_h(X_k - X_j)\xi_i\xi_j \right].$$

Taking expectation on $\tilde{\Gamma}_{n1}$ gives $\tilde{\Gamma}_{n1} = O_p(1/n^2h^{2d}) = o_p(1)$. Recall the notation $G(x, y)$, $H(x, y, h)$ in the proof of Lemma 5.2, $\tilde{\Gamma}_{n2}$ can be written as the sum

$$\frac{2h^d}{n^2} \sum_{i \neq j} G^2(X_iX_j)\xi_i^2\xi_j^2 + \frac{4}{n^2} \sum_{i \neq j} G(X_i, X_j)H(X_i, X_j, h)\xi_i^2\xi_j^2 + \frac{2}{n^2h^d} \sum_{i \neq j} H^2(X_i, X_j, h)\xi_i^2\xi_j^2$$
$$= \tilde{\Gamma}_{n21} + \tilde{\Gamma}_{n22} + \tilde{\Gamma}_{n23}.$$

From the proof of Lemma 5.2,

$$E\tilde{\Gamma}_{n21} = \frac{n(n-1)h^d}{n^2} EG^2(X_1, X_2)\tau^2(X_1)\tau^2(X_2) = o(1)$$

which implies $\tilde{\Gamma}_{n21} = o_p(1)$. Recall the notation $H_n(Z_i, Z_j)$ in the proof of Lemma 5.2, where $Z_i = (X_i\,\xi_i)$, $1 \le i \le n$, $\tilde{\Gamma}_{n23}$ is simply $2\sum_{i \neq j} H_n^2(Z_i, Z_j)$. By the Cauchy–Schwarz inequality, and the fact that the variance is bounded above by the second moment, one has

$$E\left[\sum_{i \neq j} H_n^2(Z_i, Z_j) - n(n-1)EH_n^2(Z_1, Z_2)\right]^2 = E\left[\sum_{i \neq j}[H_n^2(Z_i, Z_j) - EH_n^2(Z_1, Z_2)]\right]^2$$
$$= \sum_{i \neq j} E[H_n^2(Z_i, Z_j) - EH_n^2(Z_1, Z_2)]^2 + \sum_{i \neq j \neq k} E[H_n^2(Z_i, Z_j) - EH_n^2(Z_1, Z_2)][H_n^2(Z_j, Z_k) - EH_n^2(Z_1, Z_2)]$$
$$\le (n^2 + cn^3)EH_n^4(Z_i, Z_j).$$

The proof in Lemma 5.2 shows that the upper bound is $O((nh^d)^{-1})$. Also from the proof in Lemma 5.2, we know that, as $n \to \infty$,

$$n(n-1)EH_n^2(Z_1, Z_2) \to \int \left[\int K(u)K(u+v)du\right]^2 dv \cdot \int \tau^4(x)f^4(x)dx,$$

so $\tilde{\Gamma}_{n23} = 2\sum_{i \neq j} H_n^2(Z_i, Z_j) \to \Gamma$. Using the Cauchy–Schwarz inequality and the results for $\tilde{\Gamma}_{n21}$, $\tilde{\Gamma}_{n23}$, one can show that $\tilde{\Gamma}_{n22} = o_p(1)$. Finally, using the Cauchy–Schwarz inequality again and the results for $\tilde{\Gamma}_{n1}$, $\tilde{\Gamma}_{n2}$, one can show that $\tilde{\Gamma}_{n3} = o_p(1)$. This completes the proof of the lemma. □

Proof of Theorem 3.2. The proof can be proceed in a similar way as the proof of Theorem 3.1. For the sake of brevity, we only outline the main steps.

Adding and subtracting $g(X_j)$ from $Y_j - g(X_j, \hat{\theta}_n)$, D_n can be written as the sum of the following three terms:

$$D_{n1} = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(Y_j - g(X_j))\right]^2,$$
$$D_{n2} = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(g(X_j) - g(X_j, \hat{\theta}_n))\right]^2,$$
$$D_{n3} = \frac{2}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(Y_j - g(X_j)) \cdot \frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(g(X_j) - g(X_j, \hat{\theta}_n))\right].$$

D_{n2} can be further written as the sum $D_{n21} + D_{n22} + D_{n23}$, where

$$D_{n21} = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(g(X_j) - g(X_j, \theta_a))\right]^2,$$
$$D_{n22} = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(g(X_j, \theta_a) - g(X_j, \hat{\theta}_n))\right]^2,$$
$$D_{n23} = \frac{2}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(g(X_j) - g(X_j, \theta_a)) \cdot \frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)(g(X_j, \theta_a) - g(X_j, \hat{\theta}_n))\right].$$

One can show that

$$D_{n21}=E\left[\int K(u)L(x-uh)f(x-uh)du\right]^2+o_p(1)=\int L^2(v)f^3(v)dv+o_p(1),$$

$nh^{d/2}D_{n22}=o_p(1)$, $nh^{d/2}D_{n23}=o_p(\sqrt{nh^{d/2}})$, and $nh^{d/2}D_{n3}=o_p(nh^{d/2})$.

Adding and subtracting $g(X_j)$ in e_j, $\hat{C}_n$ can be written as the sum of the following three terms:

$$\hat{C}_{n1}=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n K_h^2(X_i-X_j)[Y_j-g(X_j)]^2,$$

$$\hat{C}_{n2}=\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n K_h^2(X_i-X_j)[g(X_j)-g(X_j,\hat{\theta}_n)]^2,$$

$$\hat{C}_{n3}=\frac{2}{n^3}\sum_{i=1}^n\sum_{j=1}^n K_h^2(X_i-X_j)[Y_j-g(X_j)][g(X_j)-g(X_j,\hat{\theta}_n)].$$

Eventually, one can show that $nh^{d/2}\hat{C}_{n2}=o_p(nh^{d/2})$ and $nh^{d/2}C_{n3}=o_p(nh^{d/2})$. Therefore

$$nh^{d/2}(\hat{D}_n-\hat{C}_n)=nh^{d/2}(D_{n1}-\hat{C}_{n1})+nh^{d/2}\int L^2(v)f^3(v)dv+o_p(nh^{d/2}).$$

Finally, we have $\hat{\Gamma}_n=\Gamma+O_p(1)$. This completes the proofs. □

Proof of Theorem 3.3. 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

$$W_i=\frac{\delta(X_i)}{2\sqrt{nh^{d/2}}}+\frac{\Delta_n(X_i)}{2\sqrt{nh^{d/2}}},$$

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 $e_i^L=Y_i-g(X_i,\hat{\theta}_n)$. Then $e_i=e_i^L+W_i$ and D_n can be written a sum of the following three terms:

$$D_{n1}=\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)e_j^L\right]^2,$$

$$D_{n2}=\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)W_j\right]^2,$$

$$D_{n3}=\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)e_j^L\right]\cdot\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)W_j\right].$$

According to $j=i$ or not, one can write $D_{n2}=D_{n21}+D_{n22}+D_{n23}$, where

$$D_{n21}=\frac{K^2(0)}{n^3h^{2d}}\sum_{i=1}^n W_i^2,\qquad D_{n22}=\frac{1}{n}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j\neq i}^n K_h(X_i-X_j)W_j\right]^2,$$

and

$$D_{n23}=\frac{2K(0)}{n^2h^d}\sum_{i=1}^n\left[\frac{1}{n}\sum_{j=1}^n K_h(X_i-X_j)W_jW_i\right].$$

Note that $|\Delta_n(X_i)|\le|\delta(X_i)|$ for any i, hence

$$EW_i^2=E\left[\frac{\delta(X_i)}{2\sqrt{nh^{d/2}}}+\frac{\Delta_n(X_i)}{2\sqrt{nh^{d/2}}}\right]^2\le\frac{E\delta^2(X)}{nh^{d/2}},$$

and $nh^{d/2}D_{n21}=o_p(1)$. Define $\overline{W}_{Ki}=E[K_h(X_i-X)W|X_i]$. Then D_{n22} can be written as

$$D_{n22}=\left(\frac{n}{n-1}\right)^2(D_{n221}+D_{n222}+D_{n223}) \tag{5.12}$$

where

$$D_{n221}=\frac{1}{n}\sum_{i=1}^{n}\overline{W}_{Ki}^{2},\qquad D_{n222}=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n-1}\sum_{j\neq i}[K_h(X_i-X_j)W_j-\overline{W}_{Ki}]\right]^2,$$

$$D_{n223}=\frac{2}{n}\sum_{i=1}^{n}\left[\frac{1}{n-1}\sum_{j\neq i}[K_h(X_i-X_j)W_j-\overline{W}_{Ki}]\overline{W}_{Ki}\right].$$

Consider D_{n222} first. A conditional argument leads to

$$\begin{aligned}ED_{n222}&=E\left[E\left(\left[\frac{1}{n-1}\sum_{j=2}^{n}[K_h(X_1-X_j)W_j-\overline{W}_1]\right]^2\Bigg|X_1\right)\right]\\&\leq\frac{1}{n-1}EK_h^2(X_1-X_2)W_2^2.\end{aligned}$$

Note that $|W_i|\leq\delta(X_i)/\sqrt{nh^{d/2}}$, so

$$EK_h^2(X_1-X_2)W_2^2\leq\frac{1}{nh^{5d/2}}\int K^2\left(\frac{x_1-x_2}{h}\right)\delta^2(x_1)f(x_1)f(x_2)dx_1dx_2=O\left(\frac{1}{nh^{3d/2}}\right)$$

and this implies

$$nh^{d/2}D_{n222}=O_p\left(\frac{1}{nh^d}\right)=o_p(1). \tag{5.13}$$

Now, let us consider D_{n221}. Note that $\mathrm{Var}(D_{n221})\leq n^{-1}E\overline{W}_{K1}^4$. Again using the fact that $|W_i|\leq\delta(X_i)/\sqrt{nh^{d/2}}$, we have

$$E\overline{W}_{K1}^4\leq\frac{1}{n^2h^d}\int\left[\int\frac{1}{h^d}K_h(u-x)|\delta(x)|f_X(x)dx\right]^4f_X(u)du=O\left(\frac{1}{n^2h^d}\right)$$

assuming that $E\delta^4(X)f_X^4(X)<\infty$. Therefore,

$$D_{n221}-ED_{n221}=O_p\left(\frac{1}{\sqrt{n^2h^{d/2}}}\right)\quad\text{or}\quad nh^{d/2}D_{221}=nh^{d/2}E\overline{W}_{K1}+o_p(1). \tag{5.14}$$

A routine calculation shows that $nh^{d/2}E\overline{W}_{K1}^2=\int\delta^2(u)Q_0^2(-m(u;\theta_0))f^3(u)du+o_p(1)$. By the Cauchy–Schwarz inequality, one can show that $nh^{d/2}D_{n223}=o_p(1)$, $nh^{d/2}D_{n23}=o_p(1)$. Hence for D_{n2}, we get $nh^{d/2}D_{n2}=\int\delta^2(u)Q_0^2(-m(u;\theta_0))f^3(u)du+o_p(1)$.

Finally, let us consider D_{n3}. Adding and subtracting $g(X_j;\theta_0)$ from e_i^L, $\overline{W}_{Ki}$ from $K_h(X_i-X_j)W_j$, one can write

$$D_{n3}=2(D_{n31}+D_{n32}+D_{n33}+D_{n34}) \tag{5.15}$$

with

$$D_{n31}=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)\xi_j^L\cdot\frac{1}{n}\sum_{j=1}^{n}[K_h(X_i-X_j)W_j-\overline{W}_{Ki}]\right],$$

$$D_{n32}=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)\xi_j^L\overline{W}_{Ki}\right],$$

$$D_{n33}=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)[g(X_j;\theta_0)-g(X_j;\hat{\theta}_n)]\cdot\frac{1}{n}\sum_{j=1}^{n}[K_h(X_i-X_j)W_j-\overline{W}_{Ki}]\right],$$

$$D_{n34}=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)[g(X_j;\theta_0)-g(X_j;\hat{\theta}_n)]\overline{W}_{Ki}\right].$$

Applying the Cauchy–Schwarz inequality to D_{n31}, we have

$$|D_{n31}|\leq\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)\xi_j^L\right]^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}[K_h(X_i-X_j)W_j-\overline{W}_{Ki}]\right]^2}.$$

Note that the first term on the right-hand side is the order of $O_p(1/\sqrt{nh^d})$, and the second term is the order of $O_p(1/nh^{3d/4})$, hence $nh^{d/2}D_{n31} = O_p(1/\sqrt{nh^{3d/2}}) = o_p(1)$.

Similarly, using the $\sqrt{n}$-consistency of $\hat{\theta}_n$, (C1), (5.6), and the Cauchy–Schwarz inequality, one can show that $nh^{d/2}D_{n3j} = o_p(1)$ for $j = 2, 3, 4$. Also, using a similar argument for the null case, one can show that $\Gamma_n \to \Gamma$ in probability, and $nh^{d/2}(C_n - C_n^L) = o_p(1)$, where C_n^L is the same as C_n with ξ_i replaced with ξ^L. This completes the proof of Theorem 3.3. □

Acknowledgments

The author would like to thank the Editor and two referees for their helpful and critical comments which lead substantially to improving 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, Tobit models: a survey, J. Econometrics 24 (1–2) (1984) 3–61.
[3] T. Amemiya, Regression analysis when the dependent variable is truncated normal, Econometrica 41 (1973) 997–1016.
[4] J.R. Blaylock, W.N. Blisard, Women and the demand for alcohol: estimating participation and consumption, J. Consum. Aff. 27 (2) (1993) 319–334.
[5] C. Ekstrand, T.E. Carpenter, Using a Tobit regression model to analyse risk factors for foot-pad dermatitis in commercially grown broilers, Preventive Vet. Med. 37 (1–4) (1998) 219–228.
[6] A.S. Goldberger, Econometric Theory, Wiley, New York, 1964.
[7] W. Haïrdle, E. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Statist. 21 (4) (1993) 1926–1947.
[8] P.J. Hall, Central limit theorem for integrated square error of multivariate nonparametric density esitmators, J. Multivariate Anal. 14 (1) (1984) 1–16.
[9] J. Hart, Nonparametric Smoothing and Lack-of-Fit Test, Springer, New York, 1997.
[10] 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. Soc. Meas. 5 (4) (1976) 475–492.
[11] J.J. Heckman, Sample selection bias as a specification error, Econometrica 47 (1) (1979) 153–162.
[12] 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.
[13] R.I. Jennrich, Asymptotic properties of non-linear least squares estimators, Ann. Math. Statist. 40 (2) (1969) 633–643.
[14] Hira L. Koul, P. Ni, Minimum distance regression model checking, J. Statist. Plann. Inference 119 (1) (2004) 109–141.
[15] L.F. Lee, 1981, A specification test for normality assumption for the truncated and censored Tobit models, Discussion paper 44, Center for Econometrics and Decision Sciences, University of Florida.
[16] E. Lichtenberg, L.K. Shapiro, Agriculture and nitrate concentrations in Maryland community water system wells, J. Environ. Qual. 26 (1) (1997) 145–153.
[17] T.F. Lin, P. Schmidt, A test of the Tobit specification against an alternative suggested by Cragg, Rev. Econ. Statist. 66 (1) (1984) 174–177.
[18] 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.
[19] F.D. Nelson, A test for misspecification in the censored normal model, Econometrica 49 (5) (1981) 1317–1330.
[20] R.J. Olsen, Approximating a truncated normal regression with the method of moments, Econometrica 48 (5) (1980) 1099–1106.
[21] J.L. Powell, Least absolute deviations estimation for the censored regression model, J. Econometrics 25 (3) (1984) 303–325.
[22] P.A. Rudd, 1982, A score test of consistency, unpublished manuscript, University of California, Berkley.
[23] W.X. Song, Distribution-free test in Tobit mean regression model, J. Stat. Plann. Inference 141 (8) (2011) 2891–2901.
[24] J. Tobin, Estimation of relationships for limited dependent variables, Econometrica 26 (1) (1958) 24–36.
[25] L. Wang, A simple nonparametric test for diagnosing nonlinearity in Tobit median regression model, Statist. Probab. Lett. 77 (10) (2007) 1034–1042.
[26] H. White, Consequences and detection of misspecified nonlinear regression models, JASA 76 (374) (1981) 419–433.
[27] H. White, Maximum likelihood estimation of misspecified models, Econometrics 50 (1) (1982) 1–25.
[28] J.X. Zheng, A consistent test of functional form via nonparametric estimation techniques, J. Econometrics 75 (2) (1996) 263–289.