A lack-of-fit test in Tobit errors-in-variables regression models

Weixing Song *, Weixin Yao

Kansas State University, 108F Dickens Hall, 66502 Manhattan, KS, United States

ABSTRACT

The problem of fitting a parametric model in Tobit errors-in-variables regression models is discussed in this paper. The proposed test is based on the supremum of the Khmaladze type transformation of a certain partial sum process of calibrated residuals. This framework covers the usual error-free Tobit model as a special case. The asymptotic null distribution of this transformed process is shown to be the same as that of a time transformed standard Brownian motion. Consistency against some fixed alternatives and asymptotic power under some local nonparametric alternatives of this test are also discussed. Simulation studies are conducted to assess the finite sample performance of the proposed test.

Published by Elsevier B.V.

1. Introduction

Tobit regression models are widely used in economics. They were first proposed by Tobin (1958) to describe censored data as an extension of probit analysis. They assume that a latent variable Y^* is related to covariates X, in the fashion of $Y^* = m(X) + \varepsilon$, where $m(x) = E(Y^*|X = x)$ is the regression function and ε is the random error. But instead of observing Y^* completely, one can actually only observe the value $Y = \max\{Y^*, y_0\}$, where y_0 is a known number. This is the so called type I Tobit regression model. See Amemiya (1984) for an extensive introduction to this model. In practice, the predictor X may also not be observed directly. Instead, a surrogate Z, which is related to X in an additive manner, is observed. That is, one has the Tobit errors-in-variables regression model which can be written as

$$Y^* = m(X) + \varepsilon, \qquad Y = \max\{Y^*, y_0\}, \qquad Z = X + u, \tag{1.1}$$

where u is the measurement error. By assuming that $m(x)$ has a parametric form $m(x, \beta)$, where $\beta \in \mathbb{R}^r$ and the integer $r \geq 1$ is known, the existing work on this model mainly focuses on the estimation of the unknown parameter β. If X can be observed directly or $u = 0$, under a normality assumption on ε, many estimation procedures have been proposed for the parameter β. These include the probit maximum likelihood estimator, least squares estimator (including nonlinear least squares, nonlinear weighted least squares), Heckman's two-step estimator, and the Tobit maximum likelihood estimator (see Heckman (1976, 1979) and Amemiya (1973, 1984) for a detailed description of these procedures). A robust estimator of β was proposed by Powell (1984) based on the least absolute deviations. When X is measured with error, Wang (1998) considered the model (1.1) with $m(X) = \beta_0 + \beta_1' X$ and proposed a two-step moment estimation procedure and a maximum likelihood estimation procedure to estimate the unknown parameters.

The predetermined parametric form of the regression function is either based on some empirical evidence or is selected for mathematical convenience. Misspecification of the regression function often results in inefficient or misleading

---

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

conclusions. Therefore, it is necessary to develop some formal numerical tests to check the adequacy of the selected regression functions. For a class of Tobit median regression models without measurement errors in which the median of the random error is assumed to be 0, Wang (2007) proposed a simple nonparametric test for checking the nonlinearity of the regression function. The literature is scant in developing lack-of-fit testing procedures for the Tobit mean regression model with measurement errors in covariates. This paper tries to fill this void by proposing a testing procedure based on the Khmaladze type transformation of a certain marked residual process. We show that the transformed marked residual empirical process converges weakly to a time-transformed Brownian motion in a uniform metric under the null hypothesis. 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 resampling methods.

The rest of the paper is organized as follows. The test statistic and its asymptotic null distribution are discussed under quite broad assumptions in Section 2. Consistency and asymptotic power of the test against $n^{-1/2}$-local nonparametric alternatives are presented in Section 3. In Section 4, simulation studies are conducted to illustrate the finite sample performance of the proposed test, and all proofs are postponed to Section 5.

Throughout this paper, B denotes the standard Brownian motion on [0, 1]. For any random variable V, F_V and f_V denote its cumulative distribution and density functions, respectively. For any vector a, $\|a\|$ will denote its Euclidean norm. For any stochastic process $W_n(x)$, $W_n(x) = u_p(1)$ indicates that $W_n(x)$ converges to 0 in probability uniformly in x over the specified range.

2. Test statistic

In model (1.1), we shall assume that X is one-dimensional, ε, u and X are independent with $E(\varepsilon) = E(u) = 0, E(X) = \mu_x$, $E(\varepsilon^2) = \sigma_\varepsilon^2, E(u^2) = \sigma_u^2$ and $\text{Var}(X) = \sigma_x^2$. The density functions of X, ε, u are also assumed to be known. In some cases the model may not be identifiable even the density functions of ε, u and X are known, and some additional conditions might be needed to ensure identifiability. For example, Wang (1998) assumed the noise-to-signal ratio σ_u^2/σ_X^2 is known in addition to the normalities of u and X.

Let

$$K_j(x) = \int_x^\infty w^j f_\varepsilon(w) dw, \quad j = 0, 1, 2, \tag{2.1}$$

and $g(z) = E(Y|Z = z)$. From (1.1), we obtain

$$g(z) = \int \mu(x) f_{X|Z}(x|z) dx, \tag{2.2}$$

where $\mu(x) = y_0 - [y_0 - m(x)] K_0(y_0 - m(x)) + K_1(y_0 - m(x))$ and $f_{X|Z}$ is the conditional density function of X given Z. Therefore, one can consider a transformed regression model

$$Y = g(Z) + \xi, \tag{2.3}$$

where $\xi = Y - g(Z)$ is uncorrelated with Z and has mean 0 and conditional variance

$$\tau^2(z) = E(\xi^2|Z = z) = \int v(x) f_{X|Z}(x|z) dx - g^2(z) \tag{2.4}$$

with

$$v(x) = m^2(x) K_0(y_0 - m(x)) + 2m(x) K_1(y_0 - m(x)) + K_2(y_0 - m(x)) + y_0^2 F_\varepsilon(y_0 - m(x)).$$

We will denote $\theta = (\beta', \mu_x, \sigma_\varepsilon^2, \sigma_u^2, \sigma_X^2)'$ for the remainder of this paper. In some cases, the identifiability condition may result in different specifications of θ. For example, in Wang (1998)'s set up, we may assume $\theta = (\beta', \mu_x, \sigma_\varepsilon^2, \sigma_u^2)'$, since σ_u^2/σ_X^2 is known. Therefore, to test $H_0 : m(x) = m(x, \beta)$ for some $\beta \in \mathbb{R}^r$, one can test $H_0 : g(z) = g(z, \theta)$ for some $\theta \in \mathbb{R}^{r+1} \times R_+^3$ or $\mathbb{R}^{r+1} \times R_+^2$ in the transformed model (2.3), where $g(z, \theta)$ is the same as $g(z)$ in (2.2) with $m(x)$ replaced with $m(x, \beta)$. For two functions $m_1(x)$ and $m_2(x)$, with corresponding $g(z)$ by $g_1(z)$ and $g_2(z)$, one can easily see $m_1(x) = m_2(x)$ for all x implies $g_1(z) = g_2(z)$ for all z. However if $g_1(z) = g_2(z)$ for all z, one may not have $m_1(x) = m_2(x)$ for all x. If the conditional density function $f_{X|Z}(x|z)$, with z as a parameter, is a complete family, and $F_\varepsilon(x)$ is strictly increasing, then $g_1(z) = g_2(z)$ for all z indeed implies $m_1(x) = m_2(x)$ for all x.

There are many lack-of-fit test procedures in the literature that check the adequacy of the parametric regression functions. See Hart (1997) for an extensive discussion of this topic. One of these test procedures based on the following cumulative marked residual process

$$T_n(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - g(Z_i, \theta)}{\tau(Z_i, \theta)} I(Z_i \leq z)$$

where $\{(Z_i, Y_i), i = 1, \ldots, n\}$ is a sample from the model (1.1), has been receiving much attention in recent years. $\tau(z, \theta)$ is the same as $\tau(z)$ in (2.4) with $m(x)$ being replaced with $m(x, \beta)$. If all the parameters in model (1.1) are known under the

null hypothesis, one can show that $T_n(z) \Longrightarrow B \circ F_Z(z)$ in $D[-\infty, \infty]$ and uniform metric. Therefore, if F_Z and all parameters are known, any reasonable continuous functional of $T_n(z)$ might be used to test the hypothesis. For example, one can reject H_0 whenever $\sup_{z \in \mathbb{R}} |T_n(z)/\sqrt{F_Z(z)}|$ exceeds a certain critical value obtained from the distribution of $\sup_{0 \le t \le 1} |B(t)|$. More about this test statistic can be found in Stute et al. (1998) (STZ).

Among the regularity conditions to ensure that $T_n(z) \Longrightarrow B \circ F_Z(z)$ in $D[-\infty, \infty]$ and uniform metric, a commonly adopted one is to assume that the standard deviation function $\tau(z, \theta)$ is uniformly bounded below from zero for all z. This technical condition keeps the exposition relatively transparent. However, this condition is often violated in real applications. As an illustration, let $u = 0, \alpha = 0, \beta = 1, y_0 = 0$, and $\varepsilon \sim N(0, 1)$, then

$$\tau(z) = z^2 \Phi(z) + z\phi(z) + \Phi(z) - [z\Phi(z) + \phi(z)]^2.$$

Although for any $z < \infty, \tau(z) > 0$, it is not bounded below from 0. But if we restrict all z's such that $z \ge c$ for some known constant c, then $\tau(z)$ will be bounded below from 0. Usually $\tau(z)$ vanishes at the boundary points in its support, thus we may consider the following modified version of $T_n(z)$,

$$\widetilde{T}_n(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - g(Z_i, \theta)}{\tau(Z_i, \theta)} I(c \le Z_i \le z).$$

If θ is known, then we can prove the following lemma

Lemma 2.1. *Suppose $E[m^2(X, \theta)] + E(\varepsilon^2) < \infty$ and $F_Z(z)$ is continuous. Under the null hypothesis, $\widetilde{T}_n(z)$ is tight over $[c, \infty]$. Moreover $\widetilde{T}_n(z) \Longrightarrow B \circ (F_Z(z) - F_Z(c))$ over $[c, \infty]$ and in uniform metric.*

If θ is unknown, one may replace it with some consistent estimator $\hat{\theta}_n$. Denote the resulting process by $\hat{T}_n(z)$. One can show that the limiting process of $\hat{T}_n(z)$ will not be distribution free. In fact, if the estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent, then under some regularity conditions, the limiting process is the sum of a Brownian motion and a Gaussian process with mean 0. Therefore, the limiting process is a Gaussian process with mean 0, but the covariance matrix has a complicated form which depends on the CDF of Z, the derivative of g with respect to θ, and the conditional variance $\tau^2(z, \theta)$. As a consequence, it is hard to determine the critical values used for testing. The same phenomena occur in the lack-of-fit test in classical regression models and measurement error models, see STZ, Koul and Song (2008) and the references therein.

To construct a distribution-free test statistic, STZ, Koul and Song (2008) applied the so-called Khmaladze type transformation on the test statistics. This transformation was first considered by Khmaladze (1981, 1988), and soon became a powerful tool for constructing distribution-free test statistics. Suppose that a stochastic process $R(z)$ has the same distribution as the sum of a Brownian motion $B(z)$, and a Gaussian process $U(z)$. The Khmaladze type transformation of $R(z)$ is a linear transformation L of R, such that $LR(z) = L(B(z)) + L(U(z)) = B(z)$ in distribution. For more about the Khmaladze type transformation, see Khmaladze (1981, 1988), STZ, Khmaladze and Koul (2004), Koul (2006) and the references therein. In this paper, we shall extend this method to the current setup. In particular, the Khmaladze type transformation of $\hat{T}_n(z)$ takes the form

$$\widehat{W}_n(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{e}_i I(c \le Z_i \le z) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{e}_i \left[\frac{1}{n} \sum_{j=1}^{n} \hat{l}'(Z_j) \hat{M}_j^{-1} I(c \le Z_j \le Z_i \wedge z) \right] \hat{l}(Z_i),$$

where

$$\hat{e}_i = \frac{Y_i - g(Z_i, \hat{\theta}_n)}{\tau(Z_i, \hat{\theta}_n)}, \qquad \hat{l}(Z_i) = \frac{\dot{g}(Z_i, \hat{\theta}_n)}{\tau(Z_i, \hat{\theta}_n)}, \qquad \hat{M}_j = \frac{1}{n} \sum_{k=1}^{n} \hat{l}(Z_k) \hat{l}'(Z_k) I(Z_k \ge Z_j \ge c)$$

and $\hat{\theta}_n$ is any $\sqrt{n}$-consistent estimator of θ_0. In the following, we shall use $e_i, l(Z_i)$, and M_j to denote $\hat{e}_i, \hat{l}(Z_i)$ and $\hat{M}_j$ with $\hat{\theta}_n$ replaced with the true parameter θ_0. Under the assumption that ε, X and u are independently and normally distributed, Wang (1998) discussed two methods of estimating the unknown parameters, two-step moment estimation and maximum likelihood estimation procedures. Both procedures provide $\sqrt{n}$-consistent estimators for θ. To date, there is no thorough discussion on model (1.1) for the estimation problem when the random components follow other distributions. This deserves an independent study. In the following discussion, we assume that there exist $\sqrt{n}$-consistent estimators for all unknown parameters in the model (1.1).

To derive the asymptotic distribution of $\widehat{W}_n(z)$, the assumptions below are needed. Let θ_0 be the value of the parameter under the null hypothesis, $\dot{g}(z, \theta)$ and $\dot{\tau}(z, \theta)$ be the derivatives of $g(z, \theta)$ and $\tau(z, \theta)$ with respect to θ, respectively.

(g1) For every $z, g(z, \theta)$ is differentiable with respect to θ and $Eg^2(Z, \theta_0) + E\|\dot{g}(Z, \theta_0)\|^2 < \infty$. For any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_0,

$$\max_{1 \le i \le n} |g(Z_i, \hat{\theta}_n) - g(Z_i, \theta_0) - (\hat{\theta}_n - \theta_0)' \dot{g}(Z_i, \theta_0)| = o_p(1/\sqrt{n}).$$

(g2) For every z, there exists a square matrix $\ddot{g}(z; \theta_0)$, having finite expectation and a nonnegative function $k(z; \theta_0)$ with $Ek(Z, \theta_0) < \infty$ satisfying the following: $\forall \delta > 0, \exists \zeta > 0$ such that $\|\theta - \theta_0\| \leq \zeta$ implies

$$\|\dot{g}(z; \theta) - \dot{g}(z; \theta_0) - \ddot{g}(z; \theta_0)(\theta - \theta_0)\| \leq \delta k(z; \theta_0)\|\theta - \theta_0\|, \quad \forall z.$$

(v1) $\tau^2(z, \theta_0)$ is bounded below from 0 over $[c, \infty)$, where c is a known real number.

(v2) For every z, $\tau(z, \theta)$ is differentiable with respect to θ with $E\|\dot{\tau}(Z, \theta_0)\|^2 < \infty$, and for all Z_i's such that $Z_i \geq c$,

$$\max_{1 \leq i \leq n} |\tau^2(Z_i, \hat{\theta}_n) - \tau^2(Z_i, \theta_0) - 2(\hat{\theta}_n - \theta_0)' \tau(Z_i, \theta_0)\dot{\tau}(Z_i, \theta_0)| = o_p(1/\sqrt{n}).$$

(M) For all $c \leq z < \infty$, $M_z = El(Z)l'(Z)I(Z \geq z) > 0$.

Conditions (g1), (g2), (v1) and (v2) require a certain degree of smoothness on g and τ^2 as functions of θ. (v1) is not necessary in the case of known θ, and one does not need to assume the process to be defined on $[c, \infty]$ either. All we need is that the variance function $\tau^2(z, \theta_0)$ is positive everywhere. Note that g and τ^2 are defined through the functions m, K and the density functions of ε, X and u, so it might be more desirable to impose the regularity conditions directly on them. As a result, we would expect a long list of regularity conditions, which seems unnecessary. Condition (M) is a technical assumption to ensure that certain matrices used in the martingale transformation are invertible. It might be possible to remove this condition in some special scenarios, as in Khmaladze and Koul (2004)'s work on goodness-of-fit tests. However, for the general case, we will keep condition (M). The necessity of (M) surely deserves a deep investigation in the future.

The following theorem gives the weak convergence result for the process $\widehat{W}_n(z)$.

Theorem 2.1. *Suppose* (g1), (g2), (v1), (v2) *and* (M) *hold. Then under* H_0, *for every* $z_0 < \infty$, $\widehat{W}_n(z) \Rightarrow B \circ (F_Z(z) - F_Z(c))$ *in* $D([c, z_0])$ *and uniform metric.*

Consequently, the test that rejects H_0 whenever $\sup_{c \leq z \leq z_0} |\widehat{W}_n(z)/\sqrt{\hat{F}_Z(z_0) - \hat{F}_Z(c)}| > b_\alpha$ will be of asymptotic size α, where b_α is such that $P(\sup_{0 \leq u \leq 1} |B(u)| > b_\alpha) = \alpha$. As we mentioned before, the restriction of the weak convergence of $\widehat{W}_n(z)$ over $[c, z_0]$ is a technical one. In some cases, one can always choose a sufficiently small value c such that there is no observation eliminated from the left side. The choice of z_0 introduces some subjectiveness into our test. In practice, our recommendation is to choose a large z_0 to cover the majority of Z range. For example, one can choose z_0 to be the 95th or 99th percentile of the empirical CDF $\hat{F}_Z$.

3. Consistency and local power

The ability to detect any deviations from the null hypothesis is referred as consistency. In this section, we show that under some regularity conditions, the test based on $\widehat{W}_n(z)$ is consistent for certain fixed alternatives, and has nontrivial asymptotic power against a large class of $n^{-1/2}$-local nonparametric alternatives.

3.1. Consistency

Let $h(x)$ be a known real-valued function and $h(x) \notin \{m(x, \beta) : \beta \in \mathbb{R}^r\}$. Consider the alternative $H_a : m(x) = h(x)$, for all $x \in \mathbb{R}$. In the null case, we assume that the estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent. Would this estimator still be $\sqrt{n}$ consistent under the alternative hypothesis H_a? This question is of interest in its own right, and deserves further study. In the classical regression setup, Jennrich (1969) and White (1981, 1982) 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 this section, we assume that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$ under the alternative H_a for some $\theta_a \in \mathbb{R}^{r+1} \times R_+^3$. The arguments in Jennrich (1969) and White (1981, 1982) shed light on the justification of this assumption. In fact, if we use the two-step moment estimation procedure, the asymptotic properties of $\hat{\mu}_x$ and $\hat{\sigma}_u^2$ stay the same, because these two estimators are not affected by the actual form of the regression function.

Now define new random variables

$$Y_i^{a*} = m(X_i, \beta_a) + \varepsilon_i, \qquad Y_i^a = \max\{Y_i^{a*}, y_0\} \tag{3.1}$$

$$e_i^a = \frac{Y_i^a - g(Z_i, \theta_a)}{\tau(Z_i, \theta_a)}, \qquad \hat{e}_i^a = \frac{Y_i^a - g(Z_i, \hat{\theta}_n)}{\tau(Z_i, \hat{\theta}_n)}, \quad i = 1, 2, \ldots, n.$$

Then $\widehat{W}_n(z)$ can be written as the sum of the following two terms

$$\hat{W}_n^a(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{e}_i^a \left[I(c \leq Z_i \leq z) - \int_{c \leq x \leq z} \hat{l}'(x)\widehat{M}_x^{-1} I(Z_i \geq x) d\hat{F}_Z(x)\hat{l}(Z_i) \right],$$

$$\hat{R}_n^a(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - Y_i^a}{\tau(Z_i, \hat{\theta}_n)} \left[I(c \leq Z_i \leq z) - \int_{c \leq x \leq z} \hat{l}'(x)\widehat{M}_x^{-1} I(Z_i \geq x) d\hat{F}_Z(x)\hat{l}(Z_i) \right].$$

Let

$$Q(x, \varepsilon) = h(x) - m(x, \theta_a) + |h(x) + \varepsilon - y_0| - |m(x, \theta_a) + \varepsilon - y_0|,$$

$$d_1(z) = E \frac{Q(X, \varepsilon)}{2\tau(Z, \theta_a)} I(c \leq Z \leq z), \tag{3.2}$$

$$\rho(z) = E \frac{Q(X, \varepsilon)}{2\tau(Z, \theta_a)} l(Z, \theta_a) I(c \leq Z \leq z), \tag{3.3}$$

$$d_2(z) = E l'(Z, \theta_a) M_Z^{-1} \rho(Z) I(c \leq Z \leq z). \tag{3.4}$$

Then we have the following result:

Theorem 3.1. *Suppose all the conditions in Theorem 2.1 hold with θ_0 replaced by θ_a and $d(z_0) = \sup_{z \leq z_0} |d_1(z) - d_2(z)| > 0$. Then for any $0 < \alpha < 1$, the test that rejects H_0 whenever $\sup_{c \leq z \leq z_0} |\widehat{W}_n(z)/\sqrt{\hat{F}_Z(z_0) - \hat{F}_z(c)}| > b_\alpha$ is consistent.*

One can replace the Q function in (3.2) and (3.3) with $E(Q(X, \varepsilon)|Z)$ which has the complicated form $E(Q(X, \varepsilon)|Z) = \int q(x) f(x|z) dx$ with

$$q(x) = [h(x) - m(x, \theta_a)] K_0(y_0 - h(x) \wedge m(x, \theta_a)) + [2y_0 - h(x) - m(x, \beta_a)][K_0(y_0 - m(x, \beta_a)) - K_0(y_0 - h(x))]$$
$$+ 2[K_1(y_0 - h(x)) - K_1(y_0 - m(x, \beta_a))],$$

where $K_0(x), K_1(x)$ are defined in (2.1).

3.2. Local power

Sometimes it is desirable to investigate the performance of a test statistic at local alternatives, since consistency tells nothing about the power when the sample size is relatively small. Let $\delta(x)$ be a measurable function such that $E(\delta^2(X)) < \infty$. Consider the following sequence of local alternatives $H_{\text{Loc}} : m(x) = m(x, \beta_0) + \delta(x)/\sqrt{n}$. We shall assume that the estimator $\hat{\theta}_n$ satisfies $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. As in the fixed alternative case, we will not justify this assumption here.

By introducing the notation

$$Y_i^{L*} = m(X_i, \beta_0) + \varepsilon_i, \qquad Y_i^L = \max\{Y_i^{L*}, y_0\} \tag{3.5}$$

$$e_i^L = \frac{Y_i^L - g(Z_i, \theta_0)}{\tau(Z_i, \theta_0)}, \qquad \hat{e}_i^L = \frac{Y_i^L - g(Z_i, \hat{\theta}_n)}{\tau(Z_i, \hat{\theta}_n)}, \quad i = 1, 2, \ldots, n.$$

$\widehat{W}_n(z)$ can be written as the sum of the following two terms

$$\hat{W}_n^L(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{e}_i^L \left[I(c \leq Z_i \leq z) - \int_{c \leq x \leq z} \hat{l}'(x) \widehat{M}_x^{-1} I(Z_i \geq x) d\hat{F}_Z(x) \hat{l}(Z_i) \right],$$

$$\hat{R}_n^L(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - Y_i^L}{\tau(Z_i, \hat{\theta}_n)} \left[I(c \leq Z_i \leq z) - \int_{c \leq x \leq z} \hat{l}'(x) \widehat{M}_x^{-1} I(Z_i \geq x) d\hat{F}_Z(x) \hat{l}(Z_i) \right].$$

Define

$$d_1^L(z) = E \frac{\delta(X) K_0(y_0 - m(X, \theta_0))}{\tau(Z, \theta_0)} I(c \leq Z \leq z), \tag{3.6}$$

$$\rho(z) = E \frac{\delta(X) K_0(y_0 - m(X, \theta_0)) l(Z, \theta_0)}{\tau(Z, \theta_0)} I(Z \geq z \geq c), \tag{3.7}$$

$$d_2^L(z) = E l'(Z, \theta_0) M_Z^{-1} \rho(Z) I(c \leq Z \leq z). \tag{3.8}$$

The power of the test against H_{Loc} can be readily obtained from the following theorem:

Theorem 3.2. *Suppose all the conditions in Theorem 2.1 hold. Then under H_{Loc}, $\widehat{W}_n(z) \Longrightarrow B \circ F_Z(z) + d_1^L(z) - d_2^L(z)$ weakly in $D[c, z_0]$ and in uniform metric.*

4. Simulation study

To explore the finite sample performance of the proposed test, we carried out a simulation study. We shall assume all the random components follow normal distributions. In this case, the regression function $g(z, \theta)$ and the variance function $\tau^2(z, \theta)$ have explicit expressions.

The data are generated from the regression function $m(x) = \beta_0 + \beta_1 x + \delta x^2$ with $\delta = 0, 0.1, 0.3$ and $y_0 = 0$. The regression function under the null hypothesis corresponds to $\delta = 0$, which will be used to check the finite sample levels of the proposed test. The regression functions with $\delta = 0.1, 0.3$ will serve as alternative models from which the finite sample powers will be investigated. $\varepsilon \sim N(0, \sigma_\varepsilon^2), X \sim N(\mu_x, \sigma_x^2), u \sim N(0, \sigma_u^2)$. By the independence of ε, X and u, one can show that

$$\varepsilon - \beta_1 u | Z = z \sim N \left(-q\beta_1(z - \mu_x), \sigma_\varepsilon^2 + p\beta_1^2\sigma_u^2 \right),$$

where $p = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$. For the sake of identifiability, we shall assume p is known. If we denote $\tilde{\beta}_0 = \beta_0 + q\beta_1\mu_x$, $\tilde{\beta}_1 = p\beta_1$, and $\xi = \varepsilon - \beta_1 u + q\beta_1(Z - \mu_x)$, then the model (1.1) reduces to the following Tobit regression model without measurement error

$$Y_i^* = \tilde{\beta}_0 + \tilde{\beta}_1 Z_i + \xi_i, \qquad Y_i = \max\{Y_i^*, 0\},$$

and $\xi_i \sim N(0, \sigma_\varepsilon^2 + p\beta_1^2\sigma_u^2)$. Let $\sigma^2 = \sigma_\varepsilon^2 + p\beta_1^2\sigma_u^2, \alpha = \tilde{\beta}_0/\sigma^2, \beta = \tilde{\beta}_1/\sigma^2$, and $\theta = (\alpha, \beta, \sigma)$. One can show that $g(z, \theta) = \sigma(\alpha + \beta z)\Phi(\alpha + \beta z) + \sigma\phi(\alpha + \beta z)$ and

$$\tau^2(z, \theta) = \sigma^2(\alpha + \beta z)^2\Phi(\alpha + \beta z) + \sigma^2(\alpha + \beta z)\phi(\alpha + \beta z) + \sigma^2\Phi(\alpha + \beta z) - g^2(z, \theta).$$

Differentiating $g(z, \theta)$ with respect to α, β, and σ, we obtain

$$\frac{\partial g(z, \theta)}{\partial \alpha} = \sigma\Phi(\alpha + \beta z), \qquad \frac{\partial g(z, \theta)}{\partial \beta} = \sigma z\Phi(\alpha + \beta z),$$

and

$$\frac{\partial g(z, \theta)}{\partial \sigma} = (\alpha + \beta z)\Phi(\alpha + \beta z) + \phi(\alpha + \beta z).$$

We used the two-step moment estimation procedure in Wang (1998) to estimate all the unknown parameters. Since all the estimators are $\sqrt{n}$-consistent and asymptotically normal, α, β and σ are continuous functions of $\alpha, \beta, \mu_x, \sigma_\varepsilon^2$ and σ_u^2, so the estimators $\hat{\alpha}, \hat{\beta}$ and $\hat{\sigma}$ are also $\sqrt{n}$-consistent and asymptotically normal.

To generate the sample from the model, we used the same setup as in Wang (1998) in which $\beta_0 = -6, \beta_1 = 0.6$, $\sigma_\varepsilon = \sigma_u = 18, \mu_x = 20$ and $\sigma_x = 180$. Thus the true noise-to-signal ratio is $\sigma_u/\sigma_x = 0.1$. The sample sizes are chosen to be $n = 100, 200, 300$ and 500. The cutoff point x_0 is chosen to be the 95th percentile of the sample. The nominal level was chosen to be 0.05. In each scenario, the test procedure was repeated 1000 times, and the empirical levels and powers were evaluated by

$$\# \left\{ \sup_{c \leq z \leq z_0} |\widehat{W}_n(z)|/\sqrt{0.95 - \hat{F}_Z(c)} > 2.24241 \right\} \Big/ 1000,$$

where 2.24241 is such that $P(\sup_{0 \leq u \leq 1} |B(u)| > 2.24241) = 0.05$. We selected c small enough so that $\hat{F}_Z(c) = 0$, therefore, the restriction of $z \geq c$ is same as the restriction $z > -\infty$ on the test statistic. The following table shows the simulation results.

Model	100	200	300	500
$\delta = 0.0$	0.060	0.059	0.051	0.053
$\delta = 0.1$	0.825	0.975	0.996	1.000
$\delta = 0.3$	0.928	0.996	1.000	1.000

When the sample size is small or moderate, the empirical level is slightly larger than the nominal level 0.05, but becomes stable around 0.05 when the sample size is larger. The empirical power is very satisfying.

5. Proofs

To prove Lemma 2.1, we shall need the following result.

Lemma 5.1. *Suppose U and V are random variables with $E(U|V) = 0, E(U^2) < \infty$. Let $\sigma^2(v) = E(U^2|V = v)$, and (U_i, V_i), $1 \leq i \leq n$ be i.i.d. copies of (U, V), c is a real constant, and*

$$S_n(v) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{U_i}{\sigma(V_i)} I(c \leq V_i \leq v), \quad v \in [c, \infty].$$

Assume $F_V(v)$ to be continuous. Then,

$$S_n(v) \Rightarrow B \circ (F_V(v) - F_V(c)), \text{ in } D([c, \infty]) \text{ and uniform metric.}$$

Proof. We can show that all finite dimensional distributions converge weakly to the right limit by using the CLT and the Cramér–Wald device.

To show the tightness of the S_n-process over $[c, \infty]$, note that for any $c \leq v_1 \leq v \leq v_2$,

$$E[S_n(v_2) - S_n(v)]^2[S_n(v) - S_n(v_1)]^2 = [F_V(v_2) - F_V(v)][F_V(v) - F_V(v_1)] \leq [F_V(v_2) - F_V(v_1)]^2.$$

This bound, combined with Theorem 15.6 of Billingsley (1968), implies that $S_n(v)$ is tight in uniform metric on $[c, \infty]$. □

Proof of Lemma 2.1. The result is a simple consequence of Lemma 5.1 applied to $U = (Y - g(Z, \theta))/\tau(Z, \theta)$, $\sigma(V) = 1$ and $V = Z$. □

To state the next lemma, let U be a continuous r.v. with d.f. G. Let $\ell(u)$ be a vector of q functions with $E\|\ell(U)\|^2 < \infty$. Assume that the matrix $C_u := E\ell(U)\ell'(U)I(U \geq u)$ is positive definite for all $u \geq c$. For a real valued function $\gamma \in L_2([c, \infty), G)$ define the transforms

$$\mathcal{T}_\gamma(u) := \int_{c \leq y \leq u} \gamma(y)\ell'(y)C_y^{-1}dG(y)\,\ell(u), \qquad \mathcal{K}_\gamma(u) := \gamma(u) - \mathcal{T}_\gamma(u).$$

The following lemma, together its proof, is similar to the one from Proposition 4.1 of Khmaladze and Koul (2004) and Lemma 9.1 of Koul (2006), which in turn has its origin in Khmaladze (1988). For the sake of brevity, the proof is omitted here.

Lemma 5.2. *Under the above set up,*

$$E\mathcal{K}_\gamma(U)\ell'(U) = 0, \quad \forall\,\gamma \in L_2([c, \infty), G)$$
$$E\mathcal{K}_{\gamma_1}(U)\mathcal{K}_{\gamma_2}(U) = E\gamma_1(U)\gamma_2(U), \quad \forall\,\gamma_1, \gamma_2 \in L_2([c, \infty), G).$$

Proof of Theorem 2.1. Denote the first term in $\widehat{W}_n(z)$ by $\widehat{W}_{n1}(z)$, and the second term by $\widehat{W}_{n2}(z)$. Let $\Delta_n(Z_i) = \tau(Z_i, \theta_0)/\tau(Z_i, \hat{\theta}_n) - 1$, and $e_i = (Y_i - g(Z_i, \theta_0))/\tau(Z_i, \theta_0)$. Then $\widehat{W}_{n1}(z)$ can be written as the sum of four terms:

$$\widehat{W}_{n11}(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i I(c \leq Z_i \leq z),$$

$$\widehat{W}_{n12}(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \Delta_n(Z_i)I(c \leq Z_i \leq z),$$

$$\widehat{W}_{n13}(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{g(Z_i, \theta_0) - g(Z_i, \hat{\theta}_n)}{\tau(Z_i, \theta_0)} I(c \leq Z_i \leq z),$$

$$\widehat{W}_{n14}(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{g(Z_i, \theta_0) - g(Z_i, \hat{\theta}_n)}{\tau(Z_i, \theta_0)} \Delta_n(Z_i)I(c \leq Z_i \leq z).$$

By Lemma 5.1, $\widehat{W}_{n11}(z) \Longrightarrow B \circ (F_Z(z) - F_Z(c))$ on $D[c, \infty]$ in uniform metric. After some algebra, we can rewrite $\widehat{W}_{n12}$ as the sum of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{e_i[\tau^2(Z_i, \theta_0) - \tau^2(Z_i, \hat{\theta}_n)]}{2\tau^2(Z_i, \theta_0)} I(c \leq Z_i \leq z), \tag{5.1}$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \frac{\tau^2(Z_i, \theta_0) - \tau^2(Z_i, \hat{\theta}_n)}{\tau(Z_i, \theta_0)} D_n(Z_i)I(c \leq Z_i \leq z), \tag{5.2}$$

where

$$D_n(Z_i) = \frac{\tau(Z_i, \theta_0)}{\tau(Z_i, \hat{\theta}_n)(\tau(Z_i, \theta_0) + \tau(Z_i, \hat{\theta}_n))} - \frac{1}{2\tau(Z_i, \theta_0)}.$$

By condition (v2), (5.1) can be further written as

$$-\sqrt{n}(\hat{\theta}_n - \theta_0)'\frac{1}{n} \sum_{i=1}^n \frac{e_i\dot{\tau}(Z_i, \theta_0)}{\tau(Z_i, \theta_0)} I(c \leq Z_i \leq z) + u_p(1).$$

In view of (v1), (v2), a Glivenko–Cantelli argument implies that

$$\frac{1}{n} \sum_{i=1}^n \frac{e_i\dot{\tau}(Z_i, \theta_0)}{\tau(Z_i, \theta_0)} I(c \leq Z_i \leq z) = u_p(1).$$

Then, by the $\sqrt{n}$-consistency of $\hat{\theta}_n$, we can show that (5.1) has the order of $u_p(1)$. Note that (5.2), by (v2), is bounded above by

$$\sup_{1 \le i \le n} |D_n(Z_i)| \cdot \sqrt{n}\|\hat{\theta}_n - \theta_0\| \cdot \frac{2}{n}\sum_{i=1}^n |e_i|\|\dot{\tau}(Z_i, \theta_0)\|.$$

From conditions (v1) and (v2), one can show that $\sup_{1 \le i \le n} |D_n(Z_i)| = o_p(1)$. This, together with $E\|\dot{\tau}(Z, \theta_0)\|^2 < \infty$ by (v2), and $\sqrt{n}$-consistency of $\hat{\theta}_n$, implies that (5.2) is $u_p(1)$. Therefore, $\widehat{W}_{n12}(z) = u_p(1)$. It follows from (g2) that

$$\widehat{W}_{n13}(z) = -\sqrt{n}(\hat{\theta}_n - \theta_0)'El(Z)I(c \le Z \le z) + u_p(1).$$

Note that (v1), (v2) also imply $\sup_{1 \le i \le n} |\Delta_n(Z_i)| = o_p(1)$, then from (v1), one can also show that $\widehat{W}_{n14}(z) = u_p(1)$. Hence, we obtain

$$\widehat{W}_{n1}(z) = \widehat{W}_{n11}(z) - \sqrt{n}(\hat{\theta}_n - \theta_0)'El(Z)I(c \le Z \le z) + u_p(1). \tag{5.3}$$

Now, let's consider the term $\widehat{W}_{n2}(z)$. Let

$$\widehat{U}_n(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{e}_i\hat{l}(Z_i)I(Z_i \ge z \ge c), \qquad U_n(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i l(Z_i)I(Z_i \ge z \ge c),$$

then $\widehat{W}_{n2}(z) = \int_{c \le x \le z} \hat{l}'(x)\hat{M}_x^{-1}\widehat{U}_n(x)d\hat{F}_Z(x)$. Recall the notation $\Delta_n(Z_i)$, one can rewrite $\widehat{U}_n(z)$ as the sum of the following eight terms

$$\widehat{U}_{n1}(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\dot{g}(Z_i, \hat{\theta}_n) - \dot{g}(Z_i, \theta_0)}{\tau(Z_i, \theta_0)}e_i\Delta_n(Z_i)I(Z_i \ge z \ge c),$$

$$\widehat{U}_{n2}(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n l(Z_i)e_i\Delta_n(Z_i)I(Z_i \ge z \ge c),$$

$$\widehat{U}_{n3}(z) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{(g(Z_i, \hat{\theta}_n) - g(Z_i, \theta_0))(\dot{g}(Z_i, \hat{\theta}_n) - \dot{g}(Z_i, \theta_0))}{\tau^2(Z_i, \theta_0)}\Delta_n(Z_i)I(Z_i \ge z \ge c),$$

$$\widehat{U}_{n4}(z) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{(g(Z_i, \hat{\theta}_n) - g(Z_i, \theta_0))\dot{g}(Z_i, \theta_0)}{\tau^2(Z_i, \theta_0)}\Delta_n(Z_i)I(Z_i \ge z \ge c),$$

$$\widehat{U}_{n5}(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\dot{g}(Z_i, \hat{\theta}_n) - \dot{g}(Z_i, \theta_0)}{\tau(Z_i, \theta_0)}e_iI(Z_i \ge z \ge c),$$

$$\widehat{U}_{n6}(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^n l(Z_i)e_iI(Z_i \ge z \ge c),$$

$$\widehat{U}_{n7}(z) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{(g(Z_i, \hat{\theta}_n) - g(Z_i, \theta_0))(\dot{g}(Z_i, \hat{\theta}_n) - \dot{g}(Z_i, \theta_0))}{\tau^2(Z_i, \theta_0)}I(Z_i \ge z \ge c),$$

$$\widehat{U}_{n8}(z) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{(g(Z_i, \hat{\theta}_n) - g(Z_i, \theta_0))\dot{g}(Z_i, \theta_0)}{\tau^2(Z_i, \theta_0)}I(Z_i \ge z \ge c).$$

It can be shown that all $\widehat{U}_{nj}(z) = u_p(1)$ except for $j = 6, 8$ by conditions (g1), (g2), (v1) and (v2). Note that $\widehat{U}_{n6}(z)$ is $U_n(z)$, from condition (g2), we can show that

$$\widehat{U}_{n8}(z) = -\frac{1}{n}\sum_{i=1}^n l(Z_i)l'(Z_i)I(Z_i \ge z \ge c)\sqrt{n}(\hat{\theta}_n - \theta_0) + u_p(1) = -M_z\sqrt{n}(\hat{\theta}_n - \theta_0) + u_p(1).$$

In fact, the law of large numbers implies the pointwise convergence, and a Glivenko–Cantelli type argument gives the uniformity. Hence

$$\sup_{z \in [c, \infty]} \|\widehat{U}_n(z) - U_n(z) + M_z\sqrt{n}(\hat{\theta}_n - \theta_0)\| = o_p(1). \tag{5.4}$$

Similarly, one can show that $\sup_{z \in [c, \infty]} \|\hat{M}_z - M_z\| = o_p(1)$. Since $M_z > 0$ for all $z \ge c$, so for any $z_0 < \infty$, we have $\sup_{z \in [c, z_0]} |\hat{M}_z^{-1} - M_z^{-1}| = o_p(1)$. Therefore, for any $z \in [c, z_0]$, $\widehat{W}_{n2}(z)$ equals

$$\int_{c \leq x \leq z} l'(x) M_x^{-1} U_n(x) d\widehat{F}_Z(x) + \int_{c \leq x \leq z} l'(x) M_x^{-1} [\widehat{U}_n(x) - U_n(x)] d\widehat{F}_Z(x) + u_p(1).$$

By Lemma 6.6.4 in Koul (2002), the first term is $\int_{c \leq x \leq z} l'(x) M_x^{-1} U_n(x) dF_Z(x) + u_p(1)$. The second term, from (5.4), can be written as $-El'(Z)I(c \leq Z \leq z)\sqrt{n}(\hat{\theta}_n - \theta_0) + u_p(1)$. Therefore, for all z such that $c \leq z \leq z_0$,

$$\left| W_{n2}(z) - \int_{c \leq x \leq z} l'(x) M_x^{-1} U_n(x) dF_Z(x) + El'(Z) I(c \leq Z \leq z)\sqrt{n}(\hat{\theta}_n - \theta_0) \right| = u_p(1),$$

which, together with (5.3), implies $\widehat{W}_n(z) = \widehat{W}_{n11}(z) - \int_{c \leq x \leq z} l'(x) M_x^{-1} U_n(x) dF_Z(x) + u_p(1)$. The desired result in Theorem 2.1 can be obtained by applying Lemma 5.2 with $\gamma(u) = I(c \leq u \leq z), U = Z, G = F_Z, l(Z) = \dot{g}(Z, \theta_0)/\tau(Z, \theta_0)$ and $C = M$. □

Proof of Theorem 3.1. From the assumption that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$ under the alternative hypothesis, one can obtain $\widehat{W}_n^a(z) \Longrightarrow B \circ (F_Z(z_0) - F_Z(c))$ using the similar arguments as in showing $\widehat{W}_n(z) \Longrightarrow B \circ (F_Z(z_0) - F_Z(c))$ in the proof of Theorem 2.1.

Write $R_n^a(z)$ as $R_{n1}^a(z) - R_{n2}^a(z)$, where

$$R_{n1}^a(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - Y_i^a}{\tau(Z_i, \hat{\theta}_n)} I(c \leq Z_i \leq z),$$

$$R_{n2}^a(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - Y_i^a}{\tau(Z_i, \hat{\theta}_n)} \int_{c \leq x \leq z} \hat{l}'(x) \widehat{M}_x^{-1} I(Z_i \geq x) d\widehat{F}_Z(x) \hat{l}(Z_i).$$

Based on the elementary equality $\max\{a, b\} = (a + b + |a - b|)/2$,

$$n^{-1/2} R_{n1}^a(z) = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i^* - Y_i^{*a}) + (|Y_i^* - y_0| - |Y_i^{*a} - y_0|)}{2\tau(Z_i, \hat{\theta}_n)} I(c \leq Z_i \leq z).$$

Note that $Y_i^* - Y_i^{*a} = h(X_i) - m(X_i, \theta_a)$, $|Y_i^*| - |Y_i^{*a}| = |h(X_i) + \varepsilon_i - y_0| - |m(X_i, \theta_a) + \varepsilon_i - y_0|$, we have $n^{-1/2} R_{n1}^a(z) = d_1(z) + u_p(1)$, where $d_1(z)$ is defined in (3.2).

Now define

$$\widehat{V}_n(z) = \frac{1}{n} \sum_{i=1}^n \frac{Y_i - Y_i^a}{\tau(Z_i, \hat{\theta}_n)} \hat{l}(Z_i) I(Z_i \geq z),$$

then $n^{-1/2} R_{n2}^a(z) = \int_{c \leq x \leq z} \hat{l}(x) \widehat{M}_x^{-1} \widehat{V}_n(x) d\widehat{F}_Z(x)$. From the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and a Glivenko–Cantelli type argument, one can show that $\widehat{V}_n(z) = \rho(z) + u_p(1)$, where $\rho(z)$ is given in (3.3). Then a routine argument leads to $n^{-1/2} R_{n2}^a(z) = d_2(z) + u_p(1)$, where $d_2(z)$ is given in (3.4). Thus

$$\sup_{c \leq z \leq z_0} |n^{-1/2} R_n^a(z) - [d_1(z) - d_2(z)]| = o_p(1). \tag{5.5}$$

Finally, the consistency of our test is derived by combining (5.5), the inequality

$$\sup_{c \leq z \leq z_0} |\widehat{W}_n(z)| \geq \sqrt{n} \sup_{c \leq z \leq z_0} |n^{-1/2} \widehat{R}_n^a(z)| - \sup_{c \leq z \leq z_0} |\widehat{W}_n^a(z)|,$$

and the condition $d(z_0) = \sup_{c \leq z \leq z_0} |d_1(z) - d_2(z)| > 0$. □

Proof of Theorem 3.2. Using the similar method as in the null case, one can show that $W_n^L(z) \Longrightarrow B \circ (F_Z(z_0) - F_Z(c))$. To deal with $R_n^L(z)$, we can rewrite it as $R_{n1}^L(z) - R_{n2}^L(z)$, where

$$R_{n1}^L(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - Y_i^L}{\tau(Z_i, \hat{\theta}_n)} I(c \leq Z_i \leq z),$$

$$R_{n2}^L(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{Y_i - Y_i^L}{\tau(Z_i, \hat{\theta}_n)} \int_{c \leq x \leq z} \hat{l}'(x) \widehat{M}_x^{-1} I(Z_i \geq x) d\widehat{F}_Z(x) \hat{l}(Z_i).$$

Note that $Y_i - Y_i^L$ is equal to

$$\frac{1}{2\sqrt{n}} \left[\delta(X_i) + |\sqrt{n} m(X_i, \theta_0) + \delta(X_i) + \varepsilon_i \sqrt{n} - y_0 \sqrt{n}| - |\sqrt{n} m(X_i, \theta_0) + \varepsilon_i \sqrt{n} - y_0 \sqrt{n}| \right],$$

so $R_{n1}^L(z)$ is the sum of the following two terms:

$$R_{n11}^L(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta(X_i)}{2\tau(Z_i, \hat{\theta}_n)} I(c \leq Z_i \leq z),$$

$$R_{n12}^L(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{|\sqrt{n}[m(X_i, \theta_0) + \varepsilon_i - y_0] + \delta(X_i)| - |\sqrt{n}[m(X_i, \theta_0) + \varepsilon_i - y_0]|}{2\tau(Z_i, \hat{\theta}_n)} I(c \leq Z_i \leq z).$$

It is easy to see that

$$R_{n11}^L(z) = E \frac{\delta(X)}{2\tau(Z, \theta_0)} I(c \leq Z \leq z) + u_p(1). \tag{5.6}$$

One can show that $R_{n12}^L(z)$ equals

$$\frac{1}{n} \sum_{i=1}^{n} \frac{|\sqrt{n}[m(X_i, \theta_0) + \varepsilon_i - y_0] + \delta(X_i)| - |\sqrt{n}[m(X_i, \theta_0) + \varepsilon_i - y_0]|}{2\tau(Z_i, \theta_0)} I(c \leq Z_i \leq z) + u_p(1).$$

For convenience, denote the first term on the right hand side by $\widetilde{R}_{n12}^L(z)$. Note that for all i,

$$\left\| \sqrt{n}m(X_i, \theta_0) + \delta(X_i) + \varepsilon_i\sqrt{n} - y_0\sqrt{n}| - |\sqrt{n}m(X_i, \theta_0) + \varepsilon_i\sqrt{n} - y_0\sqrt{n} \right\| \leq |\delta(X_i)|. \tag{5.7}$$

We can see that $|R_{n12}^L(z) - \widetilde{R}_{n12}^L(z)|$ is bounded above by

$$\sup_{1 \leq i \leq n} \left| \frac{1}{\tau(Z_i, \hat{\theta}_n)} - \frac{1}{\tau(Z_i, \theta_0)} \right| \cdot \frac{1}{n} \sum_{i=1}^{n} |\delta(X_i)|,$$

which is $u_p(1)$ from the fact $\sup_{1 \leq i \leq n} |\Delta_n(Z_i)| = o_p(1)$. By (5.7), one can show that $E[\widetilde{R}_{n12}^L(z) - E\widetilde{R}_{n12}^L(z)]^2 = u(1)$. Hence $\widetilde{R}_{n12}^L(z) = E\widetilde{R}_{n12}^L(z) + u_p(1)$. By the Lebesgue dominated convergence theorem, we can show that

$$E \frac{\left[|\sqrt{n}m(X, \theta_0) + \delta(X) + \varepsilon\sqrt{n} - y_0\sqrt{n}| - |\sqrt{n}m(X, \theta_0) + \varepsilon\sqrt{n} - y_0\sqrt{n}| \right] I(c \leq Z \leq z)}{2\tau(Z, \theta_0)}$$

$$\rightarrow E \frac{\delta(X)}{2\tau(Z, \theta_0)} I(c \leq Z \leq z)[I(m(X, \theta_0) + \varepsilon > y_0) - I(m(X, \theta_0) + \varepsilon < y_0)]$$

as $n \rightarrow \infty$. This, together with (5.6), implies $R_{n1}^L(x) = d_1^L(x) + u_p(1)$, where $d_1^L(x)$ is given by (3.6).

Let

$$\widehat{V}_n(z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - Y_i^L}{\tau(Z_i, \hat{\theta}_n)} \hat{l}(Z_i) I(Z_i \geq z).$$

Then $R_{n2}^L(z) = \int_{c \leq x \leq z} \hat{l}(x) \widehat{M}_x^{-1} \widehat{V}_n(x) d\widehat{F}_Z(x)$. From the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and a Glivenko–Cantelli type argument, it follows that $\widehat{V}_n(z) = \rho(z) + u_p(1)$, where $\rho(z)$ is given in (3.7). Then a routine argument leads to $n^{-1/2}R_{n2}^L(z) = d_2^L(z) + u_p(1)$, where $d_2^L(z)$ is given in (3.8). Therefore, we obtain $\widehat{W}_n(z) = B \circ (F_Z(z) - F_Z(c)) + d_1^L(z) - d_2^L(z) + u_p(1)$ which implies the desired result in Theorem 3.2.   □

References

Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. Econometrica 41, 997–1016.
Amemiya, T., 1984. Tobit models: A survey. J. Econom. 24, 3–61.
Billingsley, P., 1968. Convergence of Probability Measure. Wiley, New York, London, Sydney.
Hart, J., 1997. Nonparametric Smoothing and Lack-of-fit Test. Springer, New York.
Heckman, J., 1976. 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, 475–492.
Heckman, J., 1979. Sample bias as a specification error. Econometrica 47, 153–162.
Jennrich, R.J., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory Probab. Appl. 26, 240–257.
Khmaladze, E.V., 1988. An innovation approach to goodness-of-fit tests in $\mathbb{R}^m$. Ann. Statist. 16, 1503–1516.
Khmaladze, E.V., Koul, H.L., 2004. Martingale transforms goodness-of-fit tests in regression models. Ann. Statist. 32, 995–1034.
Koul, H.L., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models, 2nd edition. Springer.
Koul, H.L., 2006. Model diagnostics via martingale transforms: A brief review. In: Frontiers in Statistics. Imp. Coll. Press, London, pp. 183–206.
Koul, H.L., Song, W., 2008. Model checking with Berkson measurement error. J. Statist. Plann. Inference 138, 1615–1628.
Powell, J.L., 1984. Least absolute deviations estimation for the censored regression model. J. Econom. 25, 303–325.
Stute, W., Thies, S., Zhu, L.X., 1998. Model checks for regression: An innovation process approach. Ann. Statist. 26, 1916–1934.
Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26, 24–36.
Wang, L., 2007. A simple nonparametric test for diagnosing nonlinearity in Tobit median regression model. Statist. Probab. Lett. 77, 1034–1042.
Wang, L.Q., 1998. Estimation of censored linear errors-in-variables models. J. Econom. 84, 383–400.
White, H., 1981. Consequences and detection of misspecified nonlinear regression models. J. Amer. Statist. Assoc. 76, 419–433.
White, H., 1982. Maximum likelihood estimation of misspecified models. Econom. 50, 1–25.